AEROSPACE RESEARCH IN BULGARIA

Founded in 1978

Volume 31, Sofia, 2019
Space Research and Technology Institute
Bulgarian Academy of Sciences

Address

AEROSPACE RESEARCH IN BULGARIA
Space Research and Technology Institute
Acad. Georgi Bonchev St., block 1, 1113 Sofia, Bulgaria

web-site: http://journal.space.bas.bg
e-mail: journal@space.bas.bg

p-ISSN 1313 – 0927 e-ISSN 2367 – 9522

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

Aerospace Research in Bulgaria

31

Sofia, 2019

C o n t e n t s

С Ъ Д Ъ Р Ж А Н И Е

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e01

LATITUDINAL DEPENDENCE OF THE STRATOSPHERIC OZONE AND TEMPERATURE RESPONSE TO SOLAR PARTICLES' FORCING ON 20 JANUARY 2005

Natalya Kilifarska

National Institute of Geophysics, Geodesy and Geography – Bulgarian Academy of Sciences
e-mail: nkilifarska@geophys.bas.bg

Keywords: *Solar cosmic rays, Solar proton events, Stratospheric temperature and ozone.*

Abstract

This study examines the latitudinal-altitudinal variations of the midday O_3 and temperature response to the forcing of the enhanced flux of energetic particles, during January 2005 Solar Proton Event (SPE). We show that short-term response of the stratospheric O_3 depends strongly on the latitude and the energy of precipitating particles. At polar latitudes, where the relativistic electrons and "soft" protons are able to penetrate deeper into the atmosphere, we found a reduction of the peak ozone density in periods of enhanced particles' fluxes. Such a response is widely explained by the activation of HO_x and NO_x ozone destructive cycles. At mid-latitudes, however, the stratospheric part of the O_3 profile remains insensitive to these lower energy particles, because they affect only the thermospheric and mesospheric O_3. On the other hand, the "hard" protons, emitted during the third solar flare on 20 January, are able to propagate much deeper, affecting even the stratospheric ozone and reducing its density. As a consequence of the thinning of the ozone optical depth, the solar UV penetrates deeper into the atmosphere, activating the Slanger's mechanism for ozone production at lower levels – known also as ozone self-restoration. This could be an explanation for the obtained raise of the mid-latitude peak O_3 density in the period of atmospheric restoration after the SPE'2005.

The earlier raise of the polar ozone maximal density – i.e. between 18 and 21 January – could be related to the fact that at the moment of SPE'2005 it has been already diminished by the relativistic electrons and "soft" protons, getting ahead of the strongest proton flare. So the further ozone destruction (by particles with mixed energies) triggered the activation of its restoration several days earlier. Consequently, the latitudinal differences in the ozone response – found in ERA Interim data – could be attributed to the different energetic spectrum of solar flares, the depth of the particles' penetration into the atmosphere and the zenith angle of stratosphere illumination by the solar UV radiation. Enhancement of the lower and middle stratospheric temperature during the SPE'2005 has to be attributed to the increased ozone density and the more solar UV radiation absorbed.

Introduction

The middle atmosphere response to the influence of energetic particle (emitted by the solar proton events) is investigated by many authors [1–12], with most of them focusing their attention on the *ozone depletion* at polar latitudes. The possibility for *ozone enhancement* has been predicted by Jackman et al., [13] using the 2D chemical-transport model of NASA Goddard Space Flight Centre. While the mechanism for ozone depletion is pretty clear (i.e. due to the activation of the ozone destructive cycles by the increased amount of NO_x and HO_x families), the mechanism of O_3 enhancement was unclear. Jackman et al., [13] attributed it to the downward transportation of the NO_x, which could transform the active chlorine and bromine families into their reservoir species, reducing in such a way the ozone destruction in the lower polar stratosphere.

This explanation is however, non applicable to the raise of the mid-latitudes total ozone density, which has been reported by Krivolutsky [2]. The latter related the ozone raise to the enhanced solar electromagnetic radiation (accompanying the corpuscular one) and corresponding acceleration of the photolysis of the molecular oxygen.

Being one of the most absorbing gases in the stratosphere, the ozone plays a key role in the thermo-dynamical regime of the stratosphere-troposphere system. For this reason the determination of factors altering its variability is of great importance. This article investigates altitude-latitude distribution of O_3 and temperature anomalies (i.e. their deviation from the climatology) before, during and after the January 2005 SPE at the longitude of Greenwich meridian. We focus our attention on the short term response of the stratospheric O_3 and temperature (T) to different spectrum of impacting particles.

Data and methods

Ozone and T profiles at the Greenwich meridian have been derived from ERA Interim reanalysis. The intensity of solar proton fluxes – measured on board the geostationary spacecrafts GOES 10 and 11 – has been used in statistical analyses to estimate the efficiency of solar particles' influence on the stratospheric O_3 and T meridional profiles. The effect of the variable solar UV radiation is estimated using data for the solar radio emission at 10.7 cm ($F_{10.7}$), taken from http://spidr.ngdc.noaa.gov. Data for cosmic ray intensity (i.e. highly energetic particles with galactic origin) are taken from the Climax neutron monitor.

Before applying any statistics, we calculate the "anomalies" of all atmospheric parameters (i.e. their deviations from the climatological means for January, calculated from the whole data records, i.e. 1979–2009 of ERA Interim reanalysis). To identify the short term response of O_3 and T to different forcing factors we have used the Partial Least Square regression technique (PLS). The PLS regression generalizes and combines features from principal components analysis and multiple regression. It is particularly useful when we search for relations between a set of dependent variables (in our case O_3 or T at all 21 levels from 925 to 1 hPa, for given latitude) and a set of independent variables (*predictors*) during a short period of time. PLS regression analysis can be used even when the number of observations is small compared to the number of predictors. PLS searches for a set of *components* that performs a simultaneous decomposition of matrixes of dependent variables Y and predictors X. The main constraint is that these components must explain as much as possible of the *covariance* between X and Y. When a simultaneous impact of several predictors is estimated, each PLS *component* is a weighted function of the impact of different predictors. We have performed the PLS analysis for each of the forcing factors separately, but simultaneously at all levels, which ultimately gives us the maximum impact of each examined factor in the observed T and O_3 profiles.

Atmospheric short term response to proton forcing

January 2005 is characterized by a sequence of 3 intensive solar flares on 15, 17 and 20 January, followed by the largest *ground level event*[1] measured by neutron monitors since 1956 [14–16]. The anomalies in temperature and ozone profiles, derived from ERA Interim reanalysis, are presented in Fig. 1. The figure gives a first impression for the Northern hemisphere stratospheric response to precipitating solar energetic particles during January 2005 SPE. It is well seen that since 17^{-th} January the lower to middle stratosphere is warmer by up to $12\div16^0$ K while the O_3 mixing ratio is higher by $1\div3$ ppmv (part per million by volume). These anomalies persist until the end of January, and appear again on 7 February, lasting for another week or so. Complimentary to the results of Fedulina [1], showing a well pronounced depletion in the O_3 concentration and T below 20 km, we found out that the mid-latitude stratospheric response to the January 2005 SPE, manifest itself with an increased ozone mixing ratio and temperature raise at $20\div30$ km altitude. Both anomalies appear around $40\div60^o$N latitude and expand slowly with time toward the equator.

The reasonable question raised from Fig. 1 is whether these T and O_3 anomalies are related to the solar proton event or appear coincidently? As independent variables, used in the Partial Least Square regression analysis (PLS), have been selected: solar radio emission at 10.7 cm ($F_{10.7}$), integral electron flux with energies $E > 2$ MeV, integral soft proton flux with energies $E > 1$ MeV and the flux of hard protons with $E > 60$ MeV, as well as the galactic cosmic rays (CRs). To analyse the time evolution of the forcing factors and corresponding T and O_3 response, we have examined three periods (characterised by a different spectrum of precipitating energetic particles) i.e. 9–19 January, 17–27 January and 28 Jan–07 February.

Examination of the solar protons with energies $E > 1$ Mev and $E > 60$ MeV (shown in Fig. 2) reveals that during the second solar flare (on 17 January) the

[1]A sharp increase in the ground-level count of cosmic rays by neutron monitors (at least by 5% above background) associated with solar protons of energies > 500 MeV.

"soft" protons' intensity is much higher than that of the "hard" protons. The most powerful flare is observed on 20 January 2005, characterised by a sharp increase of the "hard" protons' spectra. On the next day – 21 January, there is another peak in the "soft" protons' intensity, but it is more than twice weaker than the first one.

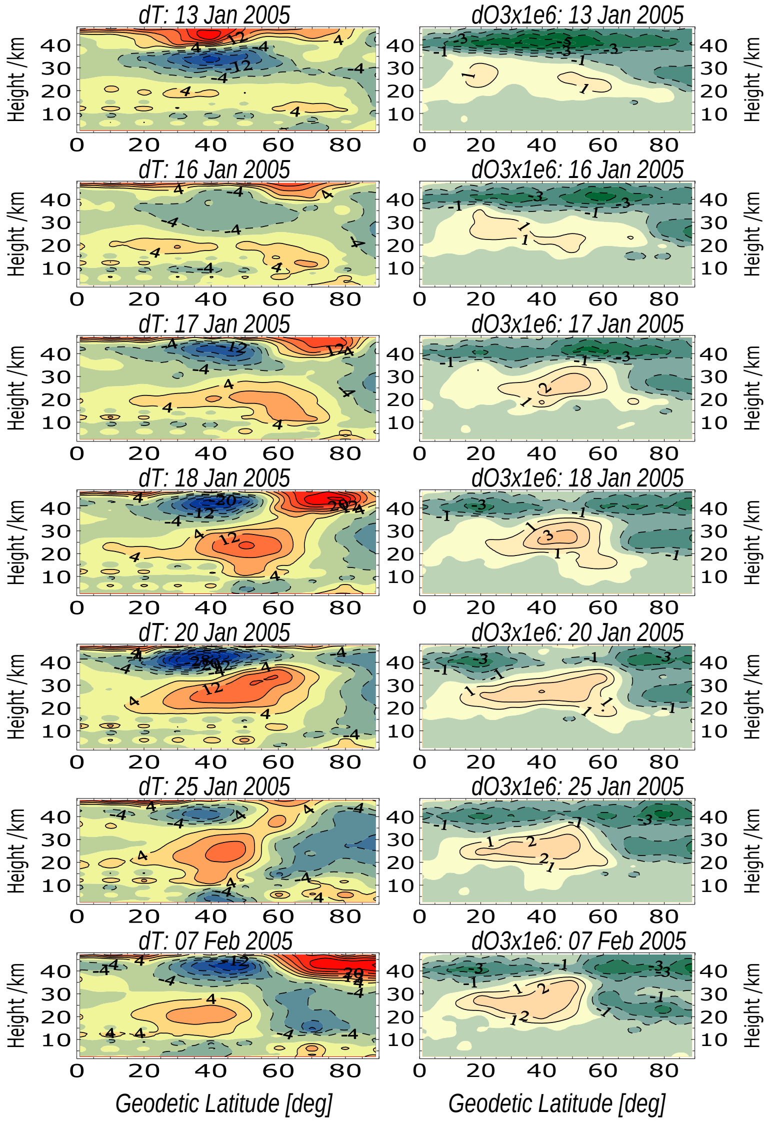

Fig. 1. Meridional cross-section of T and O_3 anomalies before, during and after January 2005 Solar Proton Events. The O_3 contour labels [ppmv] have to be multiplied by 10^{-6}. Dash contours indicate negative anomalies.

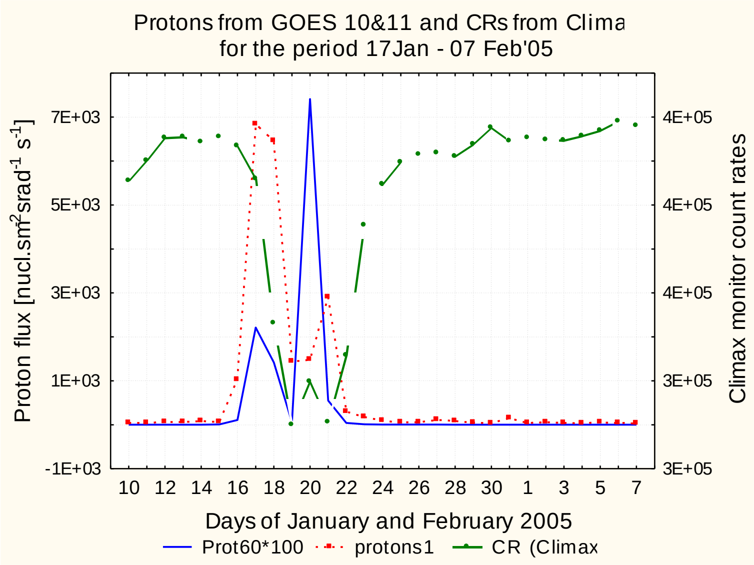

Fig. 2. Time series of protons with energy E > 1 MeV (protons1) and with E > 60 MeV (prot60) compared with the Forbush decrease of galactic cosmic rays

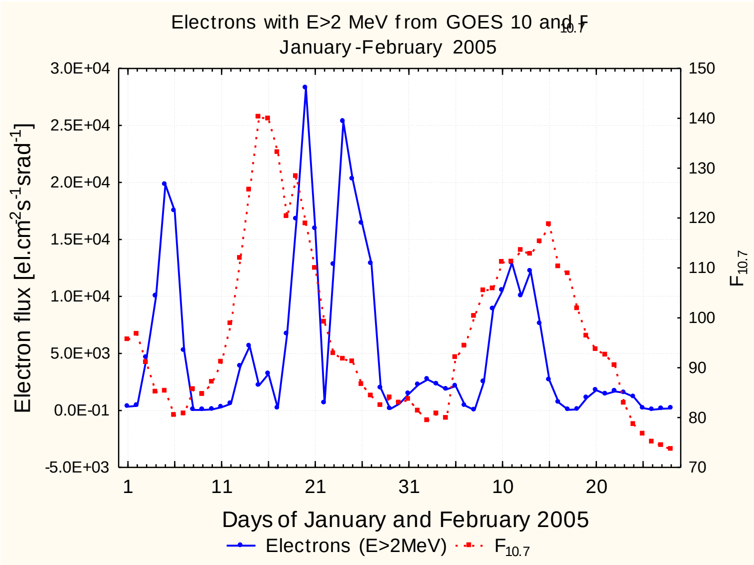

Fig. 3. Time series of the integral electron flux with energy E > 2 MeV, measured on board the GOES 10 satellite (continuous line with dots), and solar radio emission at 10.7 cm ($F_{10.7}$) – a proxy of solar UV radiation (dashed line with squares)

Figure 3 illustrates, in addition, the temporal variability of the *relativistic electrons*, measured on board the GOES 11 satellite. It is worth noting that their temporal variability is determined by the the high speed solar wind streams (emanated from the solar coronal holes), which enhance the population of energetic electrons (with E > 4 KeV) in the magnetosphere and modulate their precipitation into the lower thermosphere and mesosphere [17]. The examination of the time series of relativistic electrons with energies >2 MeV, measured on GOES 10 satellite, for the period 2000–2009, shows that their intensity is an order of magnitude higher in the period 2003–2006. This is obviously related to the fact that solar wind streams become more intensive and recurrent, when the Sun approaches the minimum of its 11 year cycle.

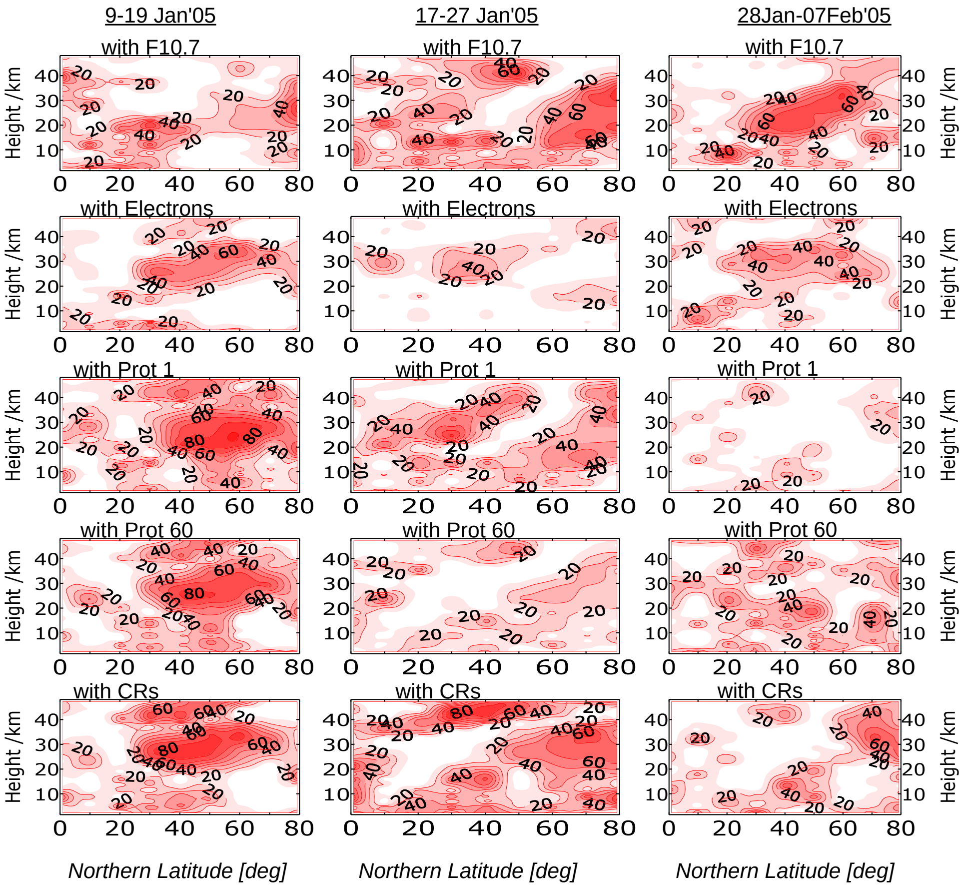

Fig. 4 (a). Percentage impact of the solar UV radiation (1^{-sr} row), electrons with E >2 MeV (2^{-nd} row), protons with E > 1 MeV (3^{-rd} row), protons with E > 60 MeV (4^{-th} row) and cosmic ray flux (from Climax neutron monitor) in temperature variability for the period 10 January – 7 February 2005

Within the investigated period, the first raise of the relativistic electrons' intensity has appeared between 4-th and 7-th January – i.e. 3 day before the first and 10 days before the second solar flare. Furthermore, there are two more picks – on 20-th and 23-th January – obviously related to the SPE'2005, and another raise between 10-th and 14-th February. In addition, the 27-day periodicity of the solar UV radiation is also showed in Fig. 3.

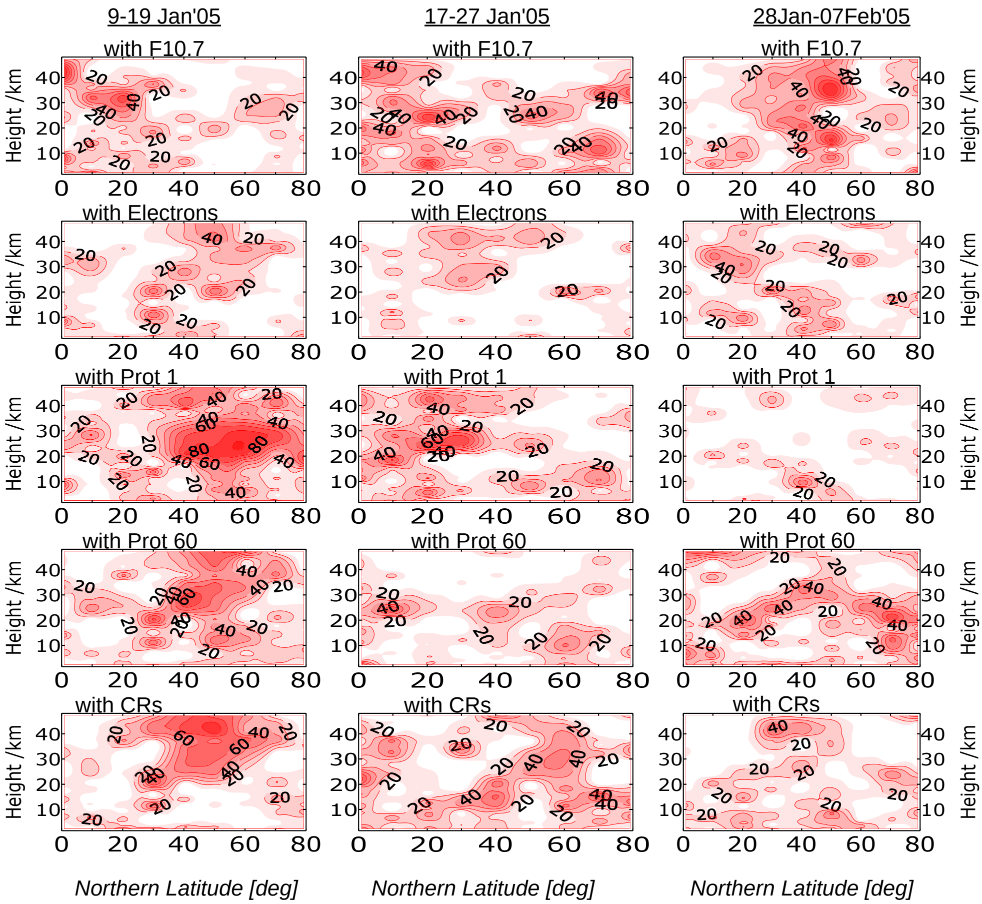

Fig. 4(b). Percentage impact of the solar UV radiation and energetic particles in ozone variability during the period 10 January – 7 February 2005

Figure 4 presents the calculated coefficient of determination R^2 multiplied by 100, which gives the percentage impact of each factor in T and O_3 variability. The analysis of Fig. 4(a) shows that middle-stratospheric warming during the *first* analysed period (refer to Fig. 1) could be attributed mainly to three of the examined factors: (i) increased density of "soft" protons, having a maximum at 17–18-th

January, which describe up to 80% of T variability, (ii) the enhancement of “hard” protons and (iii) the Forbush decrease of GCR. The impact of the last two factors is about 70÷80% of the T variability. The particles’ impact, however, seems to be short lasting and during the *second* and *third* time intervals is strongly weakened and dispersed (see the middle and right columns of Fig. 4(a)). In the second half of January two other factors become particularly important – i.e. the solar UV radiation and the relativistic electrons (with E > 2 MeV). Each of them explains up to 60% of the T variability (Fig.4(a), middle and right columns).

The ozone’s response to the analysed forcing agents is shown in Fig. 4(b). It is easily noticeable that before the solar proton event, the mid- and high latitude O_3 variability is closely related to the particles intensity – soft and hard solar protons and cosmic rays (Fig. 4(b), left column). During the main phase of the SPE’2005, as well as during the atmospheric recovery phase, the particles effect on the O_3 is substantially diluted. As should be expected, the “soft” protons impact is minimised during the recovery phase, due to the severe decrease of their flux intensity (refer to Fig. 2). The “hard” protons’ effect, however, does not disappear for the same reason, which is a hint for existing mechanism ensuring delayed O_3 response to the high speed protons. The impact of relativistic electrons remains noticeable during the recovery phase, due to subsequent spike in their intensity (refer to Fig. 3). The gradual increase of CR intensity, after the Forbush decrease (see Fig. 2), is also well traceable in O_3 variability (the right column in Fig. 4(b). The most important at this period seems the solar UV radiation – especially at middle latitudes – possibly due to the raise of its intensity.

Analysis of ozone profiles’ response to particles’ forcing

The current section is aimed to examine the variability of the ozone’s vertical profiles and if possible to attribute some specificity in its behaviour to the energetic particles’ fluxes. Figures 5 and 6 provide a direct view on the vertical profiles of O_3 anomalies – i.e. its deviations from the O_3 climatology for January, calculated over the entire data record (1979–2009) – at all examined 21 levels. Due to the spectral difference of energetic particles’ temporal variations, we have selected three periods in their temporal variability: (i) non-SPE conditions 3–8 January, characterised by raised integral flux of *relativistic electrons* between 4^{-th} and 7^{-th} January; (ii) the main phase of SPE’2005 (18–21 January) – characterised by a sharp increase of the “hard” solar protons (on 20^{-th} January) and a spike in the flux of *relativistic electrons* (within 19–21 January); (iii) the recovery phase after the SPE’2005 (23–27 January). The variability of O_3 profiles at 60^{o}N and 40^{o}N latitude, at Greenwich meridian, have been examined and compared.

The top panels of Fig. 5 reveal that the enhanced flux of *relativistic electrons* is accompanied by an increased variability of the high latitude ozone profile beneath 20 km, while their effect at mid-latitudes is substantially suppressed. The ozone's response to the mixed forcing (i.e. "soft" and "hard" protons, and *relativistic electrons*) during the SPE'2005 is more complicated – especially at high latitudes. The bottom panels of Fig. 5 shows a sudden enhancement of the peak O_3 density at 60 °N latitude – immediately after the raise of the "soft" protons' flux (on 17–18 January), as well as during the peak of relativistic electrons (on 20 January), followed by the second spike in "soft" protons (on 21 January). The mid-latitude O_3 is practically insensible to these lower energetic particles – with an exception of 19^{-th} January, when the ozone density at 18 km was enhanced by ~ 40 % (Fig. 5, bottom, right column).

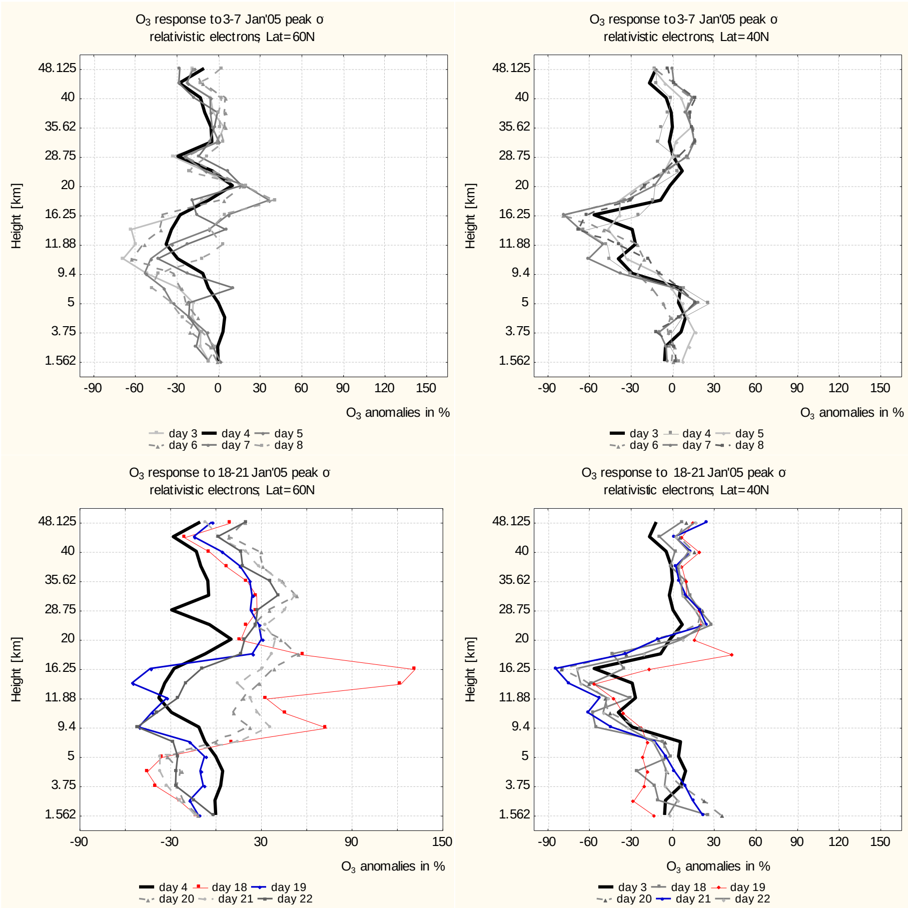

Fig. 5. Ozone profiles' variations at 60^0N (left colun) and 40^0N (right) column, found prior to and during the solar proton event in January 2005. Ozone profiles from 3 and 4 January (thick black lines) are choosen as undisturbed ones.

The recovery phase after the SPE’2005 is characterised by a stable *depletion* of the polar ozone – up to 70% compared to the January climatology (left panel of Fig. 6) – probably related to the downward transportation of NO_x family from the mesospheric levels, where they are produced. Much more surprising, during the recovery phase, is the mid-latitude O_3 behaviour. Thus, with an exception of 23^{-th} January, all other days (i.e. 24^{-th} – 27^{-th} January) are characterised by a dramatically *increased* O_3 density – more than 150% (the right panel of Fig. 6). This ozone enhancement is hardly understandable, despite the coincidence with the 3^{-rd} sharp peak in the *relativistic electrons flux.* The lower energy of these particles does not allow them to penetrate deeper into the atmosphere, due to the geomagnetic shielding, and could not directly influence the exceptional O_3 behaviour. Fig. 3 shows, in addition, that the raise of the peak ozone density should not be attributed to the solar UV radiation, because it was in the minimum of its 27-day periodicity.

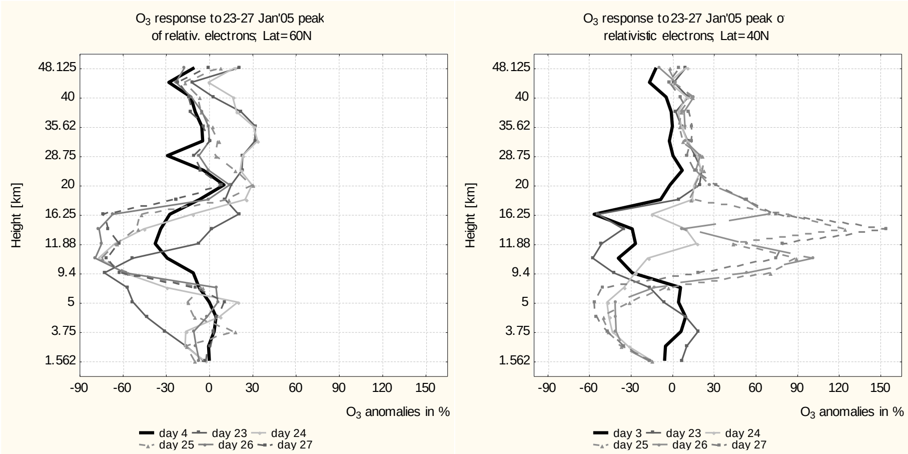

Fig. 6. Ozone profiles’variations at 60^{o}N (left colun) and 40^o N (right) column, found during the atmospheric recovery after the solar proton event in January 2005.

An attempt for explanations of this latitudinal variability of O_3 response to the particles’ forsing, during the solar proton event in January 2005, will be given in the next section.

Mechanism of O_3 enhancement during and after January’2005 SPE

It is broadly accepted that the main effect of precipitating energetic particles in the Earth’s atmosphere is the O_3 destruction, due to the activated HO_x (i.e. H, OH, HO_2) and NO_x (NO, NO_2) ozone destructive cycles [18, 19]. The satellite measurements and modeling results show that enhanced HO_x and NO_x densities can significantly impact the concentration of the mesospheric ozone.

Their influence on the stratospheric O_3, however, goes mainly through a modulation of the ozone's optical depth (i.e. the ozone column aloft a given stratospheric level).

The thinning or thickening of the O_3 optical depth increases or decreases the amount and spectral characteristics of the penetrating solar UV radiation. In normal conditions, the solar UV radiation – capable of reaching the middle stratosphere – could not dissociate molecular oxygen – O_2 [20], and consequently it could not produce ozone at these levels. However, [21] noticed that the large UV continuum, known as Hartley band (200÷350 nm), is able to dissociate ozone creating vibrationally excited molecular oxygen O_2^*. The latter is easily issociated by the freely penetrating at these levels longer UV radiation, creating atomic oxygen. The latter immediately reacts with the oxygen molecules, creating ozone, i.e.

$$O_3 + h\nu(248\ \text{nm}) \rightarrow O_2^* + O$$
$$O_2^* + h\nu(>300\ \text{nm}) \rightarrow 2\,O$$
$$3O + O_2 \rightarrow 3O_3$$

Net: $1\,O_3 \rightarrow 3\,O_3$

In resume, dissociation of a one ozone molecule by solar UV radiation leads to the formation of three new ozone molecules. This effect is known as ozone "*self-restoration*" and has been explained for the first time by Slanger [20].

This mechanism could be activated, when occasionally the thermosphere-mesospheric O_3 is reduced, which allows more UV radiation to reach the stratosphere. The efficiency of the ozone *self-restoration* has been estimated by Kilifarska et al. [21], which created a chemical model of this effect. Using their formula (6) we have estimated the changes in the O_3 profile resulting from a uniform reduction of its optical depth (τ_3) by 30% above 35 km. Calculations have been made at two latitudes – 40° N and 60° N – using the ERA Interim data for 13 January (up to the stratopause) as non-disturbed O_3 and T profiles. The mesospheric T and O_3 concentration, as well as the whole profile of the molecular O_2 have been taken from the US standard atmosphere (1976). The concentrations of the OH radical for examined days have been taken from the MLS instrument on board the AURA satellite.

Results presented in Fig. 7 show that the reduction of the ozone optical depth aloft 35 km is really followed by an O_3 increase at lower levels. This process, however, depends on the solar zenith angle (χ) and the latitude. Thus at mid-latitudes the enhancement of the O_3 concentration is maximal near the peak of the ozone layer during sunlight hours. At sunset, however, (calculations are made for χ = 89° and 105°) a distortion of the O_3 profile is found above 35 km, while the enhancement of the peak O_3 density near 25 km is strongly reduced (see the left side of Fig. 7).

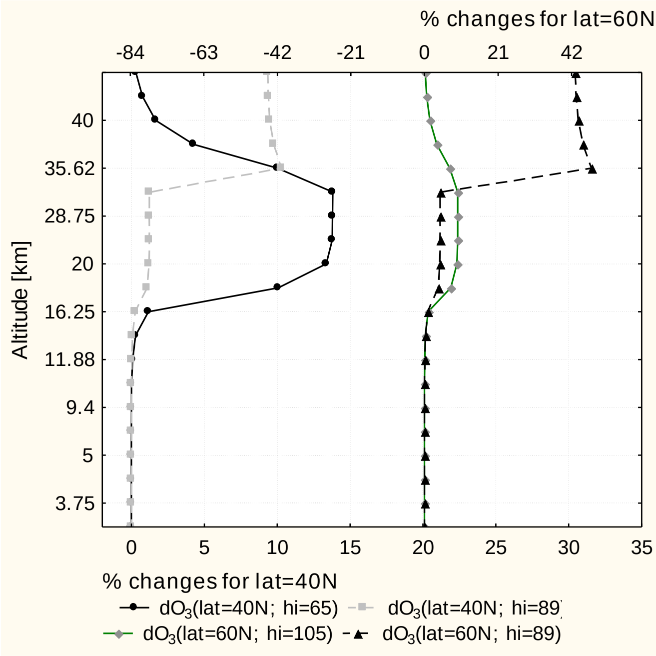

Fig. 7. Ozone response at the middle (left) and the polar latitudes (right) to a 30% reduction of its optical depth above 35 km, calculated by the [14] model. The continuous line with dots gives the O_3 profile at 40°N calculated with a solar zenith angle χ = 65°, while dashed line with squares corresponds to χ = 89° (left side of the figure); continuous line with diamonds corresponds to O_3 profile at 60°N calculated for χ = 105°, and long dashes with triangles – corresponds to χ = 89° (right part of the figure).

In January the mid-day value of the solar zenith angle at 60° N latitude is 89^0 and the calculated changes in the O_3 profile are found out above 35 km, reaching the values of 42÷44% increase of ozone density. Unlike the mid-latitudes, the model predicts a slight enhancement of the middle stratospheric ozone density for higher zenit angles, i.e. $\chi > 90°$ (presented are calculations for $\chi = 105°$; see the right side of Fig. 7). It is worth noting that the ozone *self-restoration* effect depends strongly on the lowest boundary of the mesospheric O_3 depletion. The effect is stronger, when the negative O_3 anomalies reach the upper stratosphere [21]. The results shown in Fig. 7 are a rough estimation of the *self-restoration* effect, because only the O_3 absorption at 250 nm is taken into account. More realistic results can be derived when the entire Harley and Schuman-Runge bands are included in the calculations.

The model's estimations shown in Fig. 7 could help us to understand the irregular response of O_3 profile at polar and mid-latitudes, due to the energetic particles' impact. For example, the puzzling enhancement of the polar peak O_3

density during the SPE'2005 (bottom, left panel in Fig. 5) could be attributed to the reduced O_3 optical depth, preconditioned by the increased flux of *relativistic electrons* since the beginning of January 2005 (refer to Fig. 3). The raise of "soft" protons in 17–18^{-th} January makes the polar O_3 column even thinner. At these circumstances, the higher zenith angle of the solar UV radiation illuminating the middle stratosphere, activates the self-restoration mechanism near the peak of the O_3 layer (refer to the right grey O_3 profile in Fig. 7).

At mid-latitudes, the ozone *self-restoration* is activated after the SPE'2005. This behaviour could be attributed to the less sensitivity of the mid-latitude stratospheric O_3 to the lower energetic protons and electrons, penetrating the upper atmosphere before and during the SPE'2005. However, the "hard" protons striking the atmosphere on 20^{-th} January, reduces the O_3 density deeper into the stratosphere (see the right, bottom panel of Fig. 5). In accordance with the model's simulations [21], the O_3 reduction at stratospheric levels serves as a trigger for activation of the *self-restoration* mechanism.

Conclusions

Thorough analysis of the atmospheric response to energetic particles' forcing (during Solar Proton Event SPE'2005 on 20 January 2005), shows that the stratospheric O_3 is very sensitive to the energy spectrum of penetrating energetic particles, as well as on the latitude. The evaluation and modeling of the cosmic ray interaction with the substance of the stratosphere is done by means of full Monte Carlo simulations and appropriate hadron and atmospheric models [22–25].

Evidence for an enhanced maximal O_3 density at polar latitudes and a decreased one at mid-latitudes is shown during the strongest, and with "hard" particles' spectrum, solar flare. During the recovery phase has been found the opposite response – i.e. raised mid-latitude and reduced polar ozone density.

These irregularities have been attributed do the modulation of ozone's optical depth, due to the activation of the HO_x and NO_x ozone destructive cycles at thermospheric and mesospheric levels. Thus thinning of the O_3 optical depth activates the Slanger's mechanism for ozone formation at lower atmospheric levels – an effect known also as O_3 *self-restoration*. We demonstrate that efficiency of ozone's *self-restoration* depends on the zenith angle of stratospheric illumination by the solar UV radiation – especially important at high latitudes.

The observed warming of mid-latitude stratosphere, during the SPE'2005, has been attributed to the ozone enhancement and the more solar UV radiation absorbed. The influence of high energy particles on the stratosphere continues further down to the troposphere and results in various meteorological and climatic effects [26].

References

1. Fedulina, I. N. Changes of ozone contents at middle latitudes during Forbush decreases of cosmic rays, Studia Geoph. Geod., 1998, 42, 521–32.
2. Krivolutsky, A. A. Global structure of ozone response to solar and galactic cosmic ray influence (Ground based and satellite data analysis), Adv. Space Res., 1999, 24, 5, 641–48.
3. Krivolutsky, A. A., Klyuchnikova, A. V., Zakharov, G. R., Vyushkova, T. Yu., and Kuminov, A. A. Dynamical response of the middle atmosphere to solar proton event of July 2000: Three-dimensional model simulations, Adv. Space Res., 2006, 37, 1602–13.
4. Seppälä, A., Verronen, P., Sofieva, V., Tamminen J., and Kyrola, E. Destruction of the tertiary ozone maximum during a solar proton event, Geophys. Res. Lett., 2006, 33, L07804, doi:10.1029/2005GL025571.
5. Tassev, Y., Velinov, P. I. Y., and Tomova, D. Increase of stratospheric ozone in Pfotzer maximum due to solar energetic particles during Ground Level Enhancement of cosmic rays on 20 January 2005, C. R. Acad. Bulg. Sci., 2006, 59, 11, 1153–59.
6. Tassev, Y., Velinov P. I. Y., Mateev L., and Tomova D. Comparison Between Effects of Solar Proton Events and Geomagnetic Storms on the Ozone Profiles. Adv. Space Res., 2003, 31, 9, 2163–68.
7. Tassev, Y.K., Yanev T., Velinov P. I. Y., Mateev L. Variations in the ozone profiles during the solar proton events from October 19-31, 1989. Adv. Space Res., 1999, 24, 5, 649–55.
8. Tassev, Y.K., Yanev T., Velinov P. I. Y., and Mateev L. Ozone variations in the middle atmosphere due to solar proton event from 19 October 1989.. C. R. Acad. Bulg. Sci., 1997, 50, 3, 35–38.
9. Tassev, Y., Spassov C., and Velinov P. I. Y. On the relationships between vertical ozone distribution and middle atmosphere dynamics during stratospheric warming at solar minimum, Adv. Space Res., 1993, 13, 1, 321–24.
10. Velinov, P. I. Y., Nestorov G., Spassov C., Dachev T., and Tassev Y. Ionospheric and stratospheric effects of proton flare during unusual solar activity on 22 November 1977. Adv. Space Res., 1984, 4, 4, 163–66.
11. Tassev, Y., Velinov P. I. Y., Eroshenko E., Mishev A., Mateev L., Tomova D. Numerical modeling of ozone density in the atmosphere after ground level enhancement of cosmic rays on 20 January 2005. C. R. Acad. Bulg. Sci., (Suppl.) Fund. Space Res., 2010, 63, 137–41.
12. Verronen, P. T., Rodger, C. J., Clilverd, M. A., Pickett, H. M., and Turunen, E. Latitudinal extent of the January 2005 solar proton event in the Northern Hemisphere from satellite observations of hydroxyl. Ann. Geophys., 2007, 25, 2203–15.
13. Jackman, C. H., Fleming, E. L., Vitt, F. M. Influence of extremely large solar proton events in a changing stratosphere, J. Geophys. Res., 2000, 105, D5, 11659–70.
14. Bieber, J., Clem, J., Evenson, P., Pyle R., Duldig M., Humble J., Ruffolo D., Rujiwarodom M., and Sáiz A. Largest GLE in half a century: Neutron monitor observations of the January 20, 2005 event, In: Acharya, B.S., Gupta, S,

Jagadeesan, P. (Eds.), Proceedings of the 29th International Cosmic Ray Conference, Tata Institute of Fundamental Research, Mumbai, India, 2005, 1, 237–40.
15. Mishev, A., Velinov P. I. Y., Mateev L., and Tassev Y. Ionization effect of solar protons in the Earth atmosphere – Case study of the 20 January 2005 SEP event. Adv. Space Res. 2011, 48, 7, 1232–37.
16. Mishev, A., Velinov P. I. Y., Mateev L., and Tassev Y. Ionization effect of nuclei with solar and galactic origin in the Earth atmosphere during GLE 69 on 20 January 2005. J. Atmos. Solar-Terr. Phys., 2012, 89, 1–7.
17. Thorne, R. M. The importance of energetic particle precipitation on the chemical composition of the middle atmosphere, Pure Appl. Geophys., 1980, 118, 128–51.
18. Rusch, D. W., Gerard J. C., Solomon S., Crutzen P. J., and Reid G. C. The effect of particle precipitation events on the neutral and ion chemistry of the middle atmosphere – I. Odd nitrogen, Planet, Space Sci., 1981, 29, 7, 767–74.
19. Banks, P. M., Kockarts G. Aeronomy - part B, Acad. Press, New York and London, 1973.
20. Slanger, T. G., Jusinski L. E., Black G., and Gadd G. E. A new laboratory source of ozone and its potential atmospheric implications, Science, 1988, 241, 945–50.
21. Kilifarska, N. A., Bakhmutov V. G., and Melnyk G. V. Energetic particles influence on the Southern Hemisphere ozone variability, C. R. Acad. Bulg. Sci., 2013, 66, 11, 1613–22.
22. Mishev, A. and Velinov P. I. Y. The effect of model assumptions on computations of cosmic ray induced ionization in the atmosphere. J. Atmos. Solar-Terr. Phys., 2010, 72, 476–81.
23. Mishev, A. and Velinov P. I. Y. Normalized ionization yield function for various nuclei obtained with full Monte Carlo simulations. Adv. Space Res., 2011, 48, 19–24.
24. Mishev, A. and Velinov P. I. Y. Influence of hadron and atmospheric models on computation of cosmic ray ionization in the atmosphere - Extension to heavy nuclei. J. Atmos. Solar-Terr. Phys., 2014, 120, 12, 111–20.
25. Mishev, A., Velinov, P. I. Y. Ionization effect in the middle stratosphere due to cosmic rays during strong GLE events. C. R. Acad. Bulg. Sci., 2018, 71, 4, 523–28.
26. Eroshenko, E., Velinov P. I. Y., Belov A., Yanke V., Pletnikov E., Tassev Y., Mishev A., and Mateev L. Relationships between neutron fluxes and rain flows. Adv. Space Res., 2010, 46, 637–41.

ШИРОЧИННА ЗАВИСИМОСТ НА ВАРИАЦИИТЕ В СТРАТОСФЕРНИЯ ОЗОН И ТЕМПЕРАТУРА ПО ВРЕМЕ НА СЛЪНЧЕВОТО ПРОТОННО СЪБИТИЕ ОТ 20 ЯНУАРИ 2005

Н. Килифарска

Резюме

Статията представя анализ на измененията, наблюдавани във вертикалните профили на озона и температурата на Гринуичкия меридиан, по време на протонното събитие от януари 2005 г. Изследвана е зависимостта от енергетичния спектър на частиците, измерени на геостационарния спътник GOES 11. Показано е, че озонният профил реагира по различен начин на средни и на полярни ширини. Представеното обяснение на тези особености е базирано на измененията в оптичната плътност на озона, вследствие на увеличената продукция на озоно-разрушаващите HO_x и NO_x семейства. Така намаляването на оптичната плътност на озона улеснява проникването на слънчевия ултравиолет и активира производството на озон в стратосферата по механизма на Слангер (известен още като самовъзстановяване на озона). Преставени са моделни разчети на произведения по този механизъм озон. Отбелязано е, че чувствителността на стратосферния озон към измененията в оптичната му плътност зависи, както от енергетичния спектър на частиците, така и от дълбочината на проникването им в атмосферата (контролирана от геомагнитното поле). Процесът на самовъзстановяването на озона зависи още от зенитния ъгъл на огряване на стратосферата от слънчевия ултравиолет. Ръстът на температура в средната стратосфера по време на протонното събитие е обяснен с увеличената озонна плътност и по-голямото количество адсорбирана ултравиолетова радиация.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e02

TRANSFORMATION OF THE CHARACTERISTICS OF QUASI-BIENNIAL OSCILLATION IN A NEW VERSION OF THE SERIES OF WOLF (RELATIVE SUNSPOT) NUMBERS

Igor Shibaev

IZMIRAN, Troitsk, Russia
e-mail: ishib@izmiran.ru

Keywords: *Wolf number, Quasi-biennial oscillations, Earth ionosphere.*

Abstract

With the introduction from June 2015 of a new methodology for estimation of Wolf numbers W (или WSN — Wolf sunspot number), this series was corrected from January 1749 to May 2015, i.e. a new version of the series WSN was proposed. The greatest transformation affected the cycles of a statistically reliable part of the series (since, 1849), which was clearly reflected in their amplitude correction and, accordingly, in the long-period component of the series, determining the epoch of maximum/minimum solar activity.

The quasi-biennial oscillations available in the solar magnetic field and in the total flux of its radiation also manifest themselves in a number of parameters of the Earth ionosphere and evaluation of their transformation degree is of high significance. This article compares the characteristics of the frequency interval of the quasi-biennial oscillations of both versions of a series.

Introduction

The influence of the Sun, through solar-terrestrial relations, on the climate and a human determines the traditional interest in the solar activity indices. The Zurich series of average monthly Wolf numbers (W) is the most representative one and is widely used in various applications. Since June 2015 with the introduction of a new methodology for estimation of Wolf numbers, this series was corrected from January 1749 to May 2015 (http://sidc.oma.be), i.e. a new version of the Wolf sunspot number (WSN) series was proposed. It is to be recalled that the series of average monthly Wolf numbers includes a series of regular instrumental observations from 1849 to the present day is a reliable series, and a series of restored values from 1749 to 1849. The greatest transformation affected the cycles of a statistically reliable part of the series (since 1849), which was clearly reflected in their amplitude correction and, accordingly, in the long-period component of the series, determining the epoch of maximum/minimum solar activity [1].

The quasi-biennial oscillations available in the solar magnetic field and in the total flux of its radiation [2] also manifest themselves in a number of parameters of the Earth ionosphere and evaluation of their transformation degree is of high significance. This paper compares the frequency interval of the quasi-biennial oscillations of both versions of a series. Officially, "quasi-biennial oscillations" enrich the spectra of the series and provide the cycles with an individual look. Their parameters can also perform diagnostic functions, i.e., in the representation of the "envelope curve-instantaneous frequency" signal («A(t) – F(t)»), one can judge the nature of a process by the degree of smoothness of these variables [3]. This is well illustrated by the example of WSN series in the transition to the Wolfer system in 1894.

Initial data

An overview of both versions of a series of Wolf (W_{new} – the new version, W_{old} – the old version) numbers and their relationship W_{new}/W_{old} is presented in Fig. 1. It is apparent that the reliable part of the series since 1849 has been transformed to the most extent. This will affect the long-period component of the W_{new} series with which the manifestation of the epoch of maximum/minimum solar activity is usually associated. Comparison of cycles in the old and new versions of the Wolf number series and analysis of their long-period components are conducted in the work [1]. Note that the periods of T_{new} = 131 years and T_{old} = 149 years were obtained with the sinus approximation of the long-period components.

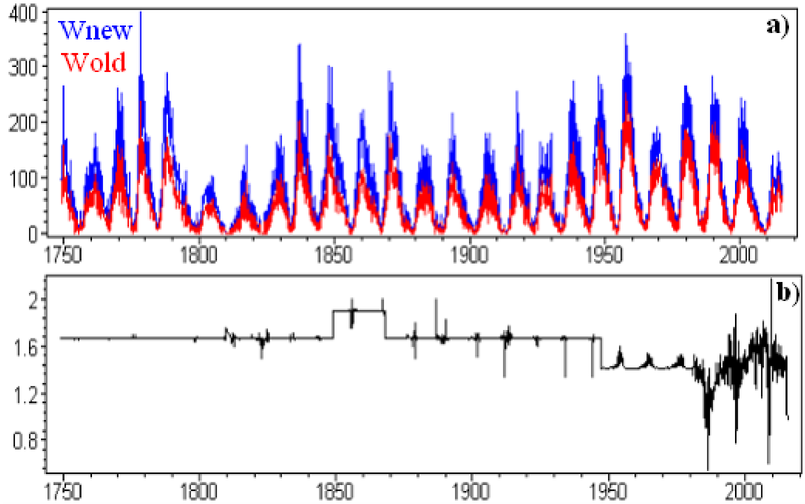

Fig. 1. (a) – an overview of W_{new}, W_{old}; (b) – relationship W_{new}/W_{old}. Axis OX – date

In paper [3], based on the spectrum nature, row $W(=W_{old})$ is divided into five spectral intervals with the following time periods in years: P1 [24 < T], P2 [6.8 < T < 24], P3 [4.26 < T < 6.8], P4 [1.66 < T < 4.26], and P5 [T<1.66]. Fig. 2 demonstrates an overview of the spectrum with assigned intervals. Recall the role of P1 – P5 components of W series. The sum of the long-period component of P1 and P2 (vicinities of the fundamental harmonic f*, T* = 1/f* ~ 131 months) reflects the main time and amplitude characteristics of the cycles. Row P3 adjusts the branches of growth and decline. Component P4 transforms the smooth relief of cycles by means of "quasi-biennial oscillations", local maxima appear and the main maximum can shift, i.e. cycles become more individual. The high-frequency residue P5 includes the annual and 155-d harmonics.

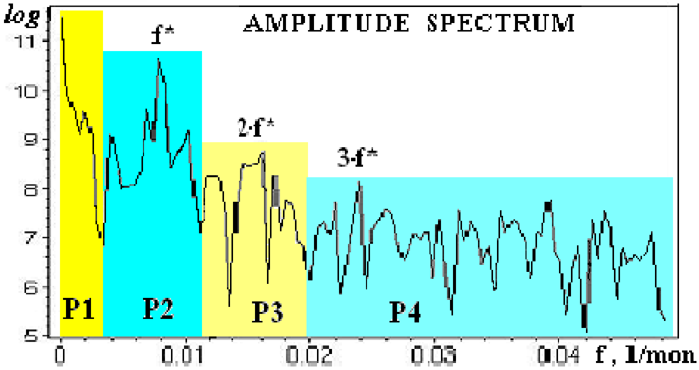

Fig. 2. Spectrum W_{old}, axis OX – inverse months

Each of the components P2 – P4 can be described by the $A(t) \times EXP[j \times \Theta(t)]$ - type template, and the Hilbert transformation [4] can be used to specify the time dependencies A(t), Θ(t). This enables to describe the signal with a slowly varying "envelope" and "instantaneous frequency" («A(t) – F(t)»).

"Quasi-biennial oscillations"

Let us demonstrate this approach by comparing the characteristics of the quasi-biennial oscillations (P4 ≥ 1.66 < T < 4.26) of the Greenwich series of areas S and a series of Wolf numbers $W(= W_{old})$ within the time interval 1874 ÷ 1976. The temporal dynamics of the "instantaneous" frequencies of these series is displayed at the top of Fig. 3, at the bottom – the dynamics of the envelopes and the date – along the OX axis. One can see the close dynamics of all the parameters of quasi-biennial oscillations of the series under consideration until 1975, and further there are known problems in the registration of a series of areas. Note the moment of transition to the Wolfer system in 1894 for the Wolf numbers (indicated by a rectangle). The continuity of the "instantaneous" frequency F[P4(W)] is broken and the envelope amplitude of the Wolf numbers A[P4(W)] is transformed, the smoothness of these parameters (F[P4(S)], A[P4(S)]) in a number of areas is maintained. In a number of cases, a change in the ratio of the amplitudes of the envelopes can be noted, but the nature of the temporal dynamics, as a rule, coincides.

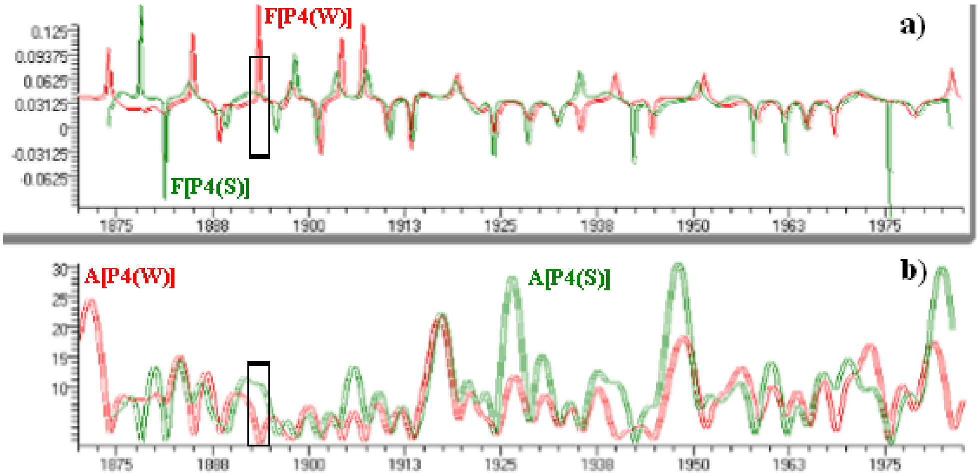

Fig. 3. Review of the parameters of quasi-biennial oscillations:
(a) – the "instantaneous" frequencies of W (F[P4(W)]) and S (F[P4(S)]);
(b) – the envelope amplitude of W (A[P4(W)]) and S (A[P4(S)])

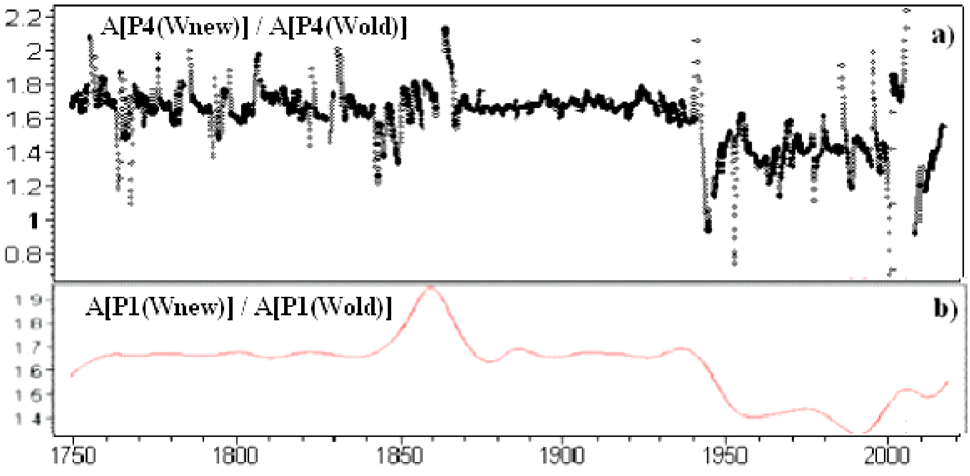

Fig. 4. (a) – relationship $A[P4(W_{new})]/A[P4(W_{old})]$; (b) – relationship $A[P1(W_{new})]/A[P1(W_{old})]$. Axis OX – date.

Let us apply this approach when comparing the quasi-biennial oscillations of the new and old versions of the series of Wolf numbers. Fig. 4a represents the amplitude ratio of the envelopes of "quasi-biennial oscillations" W_{new} and W_{old}, which is compared (Fig. 4b) with the ratio of the long-period components of these series [1]. In accordance with the W_{new}/W_{old} values ratio (Fig. 1), there are four areas with different conversion factors:

I – a series of restored values from 1749 to 1849;

II – from the beginning of a reliable series, part of cycle 9, and before the beginning of cycle 11;

III – interval with cycles 11 ÷ 17;

IV – the interval from cycle 18 and to the end is characterized by the most complex transformation.

When comparing the "quasi-biennial oscillations" W_{new} and W_{old}, it can be seen that there is the most stable connection between them only on the interval **III**. On the **II** interval we have a growing trend in contrast to the proportionality of W_{new} and W_{old}. The remaining intervals are characterized by a variety of all situations. We also note that the new version retains the "effect" associated with the transition to the Wolfer system.

Conclusion

The nature of transformation of the quasi-biennial oscillations and the long-period component of the series of Wolf numbers (1749 ÷ 2015) is quite different. Actually, old artefacts were added with new ones in the transformed series. A real assessment of the proposed method to count Wolf numbers can be obtained by comparing both versions of a series and a number of areas starting from June 2015.

References

1. Shibaev, A. The characteristics of old and new versions of month Wolf numbers range and there uniformity are compared. In: Proceedings of 12th Scientific Conference with International Participation "Space, Ecology, Safety" SES'2016, Sofia, Bulgaria, 2017, 73–76.
2. Ivanov-Holodnyj, G. S., V. S. Chertoprud. Kvazidvuhletnie variacii polnogo potoka izluchenija Solnca: ih projavlenie v variacijah stratosfernogo vetra i skorosti vrashhenija Zemli. Solnechno-zemnaja fizika. 2008, 12, 2, 291–92. (in Russian)
3. Shibaev, I. G. Estimation of the Reconstructed Part of the Wolf Series and the Possibility of Its Correction. Solar System Research, 2008, 42, 1, 63–72. DOI: 10.1134/S0038094608010073.
4. Bendant, J., A. Pirsol. Random Data. Analysis and Measurement Procedures, John Wiley & Sons, Inc., NY, 1986.

ТРАНСФОРМАЦИЯ ХАРАКТЕРИСТИК КВАЗИДВУХЛЕТНИХ ВАРИАЦИЙ В НОВОЙ ВЕРСИИ РЯДА ЧИСЕЛ ВОЛЬФА

И. Шибаев

Резюме

С введением с июня 2015 г. новой методики оценки чисел Вольфа W (или WSN — Wolf Sunspot Number) проведена коррекция этого ряда с января 1749 г. по май 2015 г., т.е. предложена новая версия ряда WSN. Наибольшая трансформация коснулась циклов достоверной части ряда с 1849 г., что явно отразилось в их амплитудной коррекции и, соответственно, длиннопериодной компоненте ряда, определяющей эпохи максимума/минимума солнечной активности.

Квазидвухлетние вариации, присутствующие в магнитном поле Солнца и в полном потоке его излучения, также проявляются в ряде параметров ионосферы Земли и важно оценить степень их трансформации при таком переходе. Данная работа сопоставляет характеристики частотного интервала квазидвухлетних вариаций обеих версий ряда.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e03

SUBSTORMS MANIFESTATION AT HIGH AND MID-LATITUDES DURING TWO LARGE MAGNETIC STORMS

Veneta Guineva[1], Irina Despirak[2], Natalia Kleimenova[3]

[1]Space Research and Technology Institute – Bulgarian Academy of Sciences
[2]Polar Geophysical Institute, Apatity, Murmansk region, Russia
[3]Schmidt Institute of the Earth Physics RAS, Moscow, Russia
e-mail: v_guineva@yahoo.com

Keywords: *Substorms, Geomagnetic storms, Auroral latitudes.*

Abstract

The dynamics of magnetic substorms at high and middle latitudes during two severe geomagnetic storms: on 17 March 2015 and on 22–23 June 2015 has been analyzed. The storms were rather similar: both storms were a result of the solar wind Sheath impact and both storms were characterized by a strong intensity ($SYM/H_{min} < -200$ nT). We studied the magnetic substorms during these storms on the base of the INTERMAGNET and IMAGE networks data. The attendant solar wind and Interplanetary Magnetic Field (IMF) parameters were taken from the OMNI database. The spatial-temporal dynamics of three substorms was studied in detail: at 17:29 UT and at 22:55 UT during the first storm and at 18:33 UT during the second storm. The substorms on 17.03.2015 originated during the main storm phase, and the onset of the substorm on 22.06.2015 followed the storm sudden commencement (SSC) of the second storm. All three substorms were characterized by a sharp poleward expansion of the westward electrojet simultaneously with a slower motion to lower latitudes. They were observed also at middle and low latitudes as positive magnetic bays. The westward electrojet reached ~71° CGMLat during the first two substorms and surpassed 75° CGMLat during the third substorm. Therefore, the first two events were "classical" substorms, and the third one – an "expanded" substorm. We suggested that this behavior is related to the different solar wind conditions: the "classical" substorms developed under magnetic cloud (MC) conditions, and the "expanded" – under the Sheath region effect.

Introduction

Substorms are a characteristic event at auroral latitudes. It is well known that during the substorm expansion phase, the westward electrojet propagates fast poleward, usually by a series of jumps. Depending on the magnetic activity, the electrojet could reach latitudes well above the typical location of the night side auroral oval [e.g., 1–10]. Thus, when the electrojet moves to geomagnetic latitudes higher than 75°, the so called "expanded" substorm forms [11]. However, it is generally accepted that under highly disturbed conditions, for example, under enhanced magnitude of the Interplanetary Magnetic Field (IMF) negative B_Z

component, the oval equatorward boundary shifts as well down up to ~50° geomagnetic latitudes. So, in such conditions, the magnetic substorms can be observed at middle and even low latitudes as positive magnetic bays [e.g., 12]. Akasofu, Chapman and Meng [13] assumed that the positive bay was created by the low-latitude return currents from the westward electrojet. Later on Akasofu and Meng [14] and Meng and Akasofu [15] explained the positive bays as a result of the field aligned currents. The mid-latitude positive bays are usually observed in the substorm expansion phase and actually they are caused by the substorm current wedge [16, 17].

The goal of our paper is to study the interplanetary and geomagnetic conditions suitable for the substorms activity at middle and low latitudes and their possible relationship with the substorms at high latitudes analyzing the magnetic disturbances during two large magnetic storms: on 17 March 2015 and 22–23 June 2015.

Data

We used the magnetic data from the IMAGE and INTERMAGNET networks. From the IMAGE set, we considered data from the meridional chain stations Suwalki (SUW) - Ny Ålesund (NAL), situated in the longitudinal range 98° ÷ 112° CGMLon, and covering the latitudinal range from 52º to 75º CGM lat. The list of the IMAGE stations and their coordinates is given at http://space.fmi.fi/image/www/index.php?page=stations. The chosen INTERMAGNET stations are in the longitudinal range of 92° ÷ 104° CGMlon, from 35° to 64° CGMlat. The magnetic observatories names and coordinated can be found at the INTERMAGNET site http://www.intermagnet.org/data-donnee/dataplot-eng.php? type=xyz.

The westward electrojet development was estimated by the time evolution of the equivalent ionospheric currents, computed by the Finish Meteorological Institute (FMI) on-line tool for 22.06° lon. (~112° CGMLon) (http://space.fmi.fi/MIRACLE/iono_1D.php#form). The solar wind and Interplanetary Magnetic Field (IMF) parameters were provided by the OMNI database (https://cdaweb.sci.gsfc.nasa.gov/cgi-bin/eval1.cgi) and by the catalog of large-scale solar wind phenomena (ftp://ftp.iki.rssi.ru/omni/) [18].

Results

Interplanetary and geomagnetic conditions

The interplanetary and geomagnetic conditions during the examined events are presented in Fig. 1. From up to down, the following quantities are shown: the magnitude of the interplanetary magnetic field (IMF) B_T, the IMF B_Z, the flow velocity V_X, the plasma density, temperature, pressure (P), and the AE, SYM/H and K_P geomagnetic indices. The considered storms were the largest ones during the present solar cycle 24.

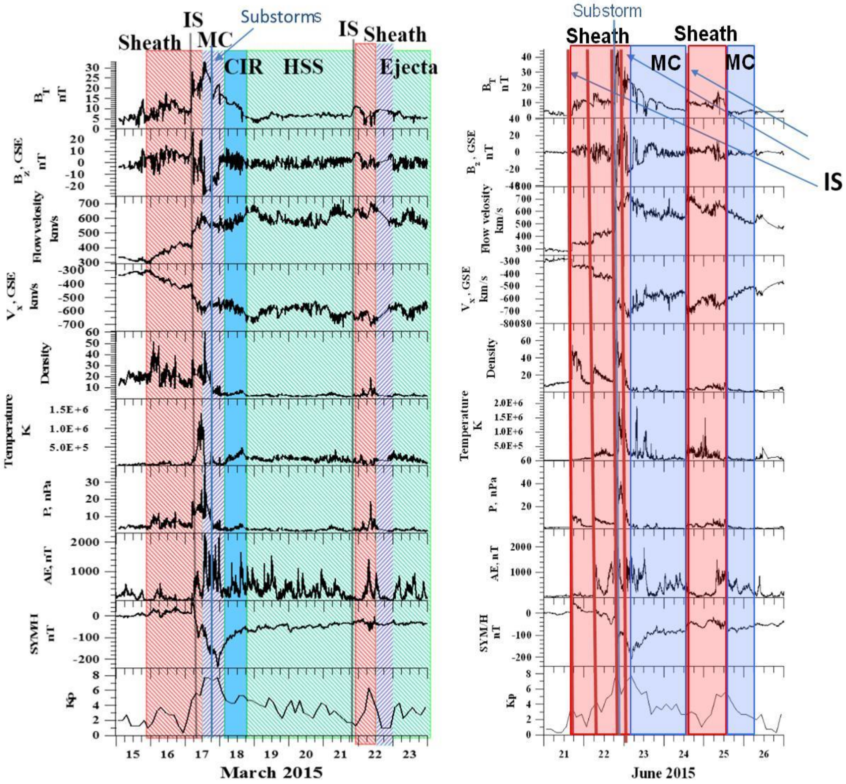

Fig. 1. Interplanetary and geomagnetic conditions during the storms on 15 March 2015 and 22–23 June 2015. The structures in the solar wind are marked by rectangles in different colours and inscribed in the upper part of the figure. The moments of interplanetary shocks (IS) arrivals are indicated by straight vertical lines. The time of the substorms during the main storm phases are marked by blue vertical lines.

The geomagnetic storm on 17 March 2015 (St. Patrick storm) was caused by a solar flare and the associated coronal mass ejections (CMEs) on 15 March 2015. The storm sudden commencement (SSC) was initiated by the formed large interplanetary shock (IS) in the sheath region. SYM/H jumped from 16 to 66 nT.

The storm was a severe one (of level G4), and the G3/G4 conditions were sustained for ~12 hours. The main phase continued ~18 hours. SYM/H fell down to –235 nT. The B_Z component of the IMF reached –30 nT and was retained ~ –20 nT for ~6.5 hours.

The storm on 22–23 June 2015 (the summer solstice storm) originated during variable solar wind conditions, when a consecution of three CMEs reached the Earth. At the third interplanetary shock the IMF B_Z turned from positive to negative and dropped to –40 nT, at that time the storm sudden commencement occurred with a sudden impulse from –20 nT to 69 nT. This storm was also a severe (G4) storm, the level of moderate-severe storm was retained for about 7 hours. The main phase lasted about 9 hours. SYM/H_{min} was –208 nT. The IMF B_Z was sustained ~ –20 nT for about 6 hours.

Both considered storms were similar to each other: they were SHEATH-caused storms, initiated by interplanetary shocks in the SHEATH region, they were very intensive, of level G4, they had clearly expressed storm sudden commencements, two-step main phases and long lasting recovery phases (Fig. 1).

Three substorms have been studied in detail: two substorms, registered during the main phase of the first storm (with their onsets at 17:29 UT and 22:55 UT on 17 March 2015), and one substorm generated during the initial phase of the second storm at 18:33 UT on 22 June 2015. The substorms of 17 March 2015 are presented in Fig. 2 and Fig. 3, and the substorm of 22 June 2015 – in Fig. 4 and Fig. 5. In Fig. 2 and Fig. 4, the equivalent ionospheric currents (upper panels) and the X-component of the magnetic field at the IMAGE latitudinal chain SUW-NAL (bottom panels) are given for the substorms on 17 March 2015 and 22 June 2015, respectively. The upper panels demonstrate the westward electrojet geographic latitude dynamics, estimated at the 22.06° geographic longitude. In Fig. 3 and Fig. 5 the magnetic field X-component at the selected INTERMAGNET stations during the considered substorms is presented. In the figures, the magnetic station location is arranged by the latitude. The substorm onsets are indicated by the red vertical lines (determined by TAR NUR and PEL stations data).

The values of the IMF B_T, IMF B_Z and solar wind parameters were averaged for 1.5 hours before the substorm onsets.

Substorm at 17:29 UT on 17 March 2015

This substorm has originated during the main storm phase, at the time of the magnetic cloud (MC) in the solar wind (see Fig. 1). The averaged parameter values were: B_T = 23 nT, B_Y = 2.0 nT, B_Z = –19 nT, V_X = –570 km/s. At the substorm onset, SYM/H was -176 nT. The westward electrojet moved fast to the Nord from ~56°÷62° to ~69° CGMlat at ~17:50 UT. After that, at ~18:05 UT, a new northward jump occurred and the electrojet reached ~72° CGMlat. A slower movement to the South was observed as well (Fig. 2, upper). The disturbances in the X-component begun at NUR (56.89 CGMlat.). They are clearly expressed to the North, to BJN (71.45° CGMLat) as well as to the South, to BRZ (52.30° CGMlat) (Fig. 2, bottom panel). At the lower latitudes, a positive bay in the X-component was observed at all mid-latitude stations to the South from HLP (50.70° CGMlat) (Fig. 3). It lasted about 20 min.

This positive bay could be seen even at the equatorial latitudes, at the station Adis Abeba (AAE), at 5.22° CGM lat. (not shown in Fig. 3).

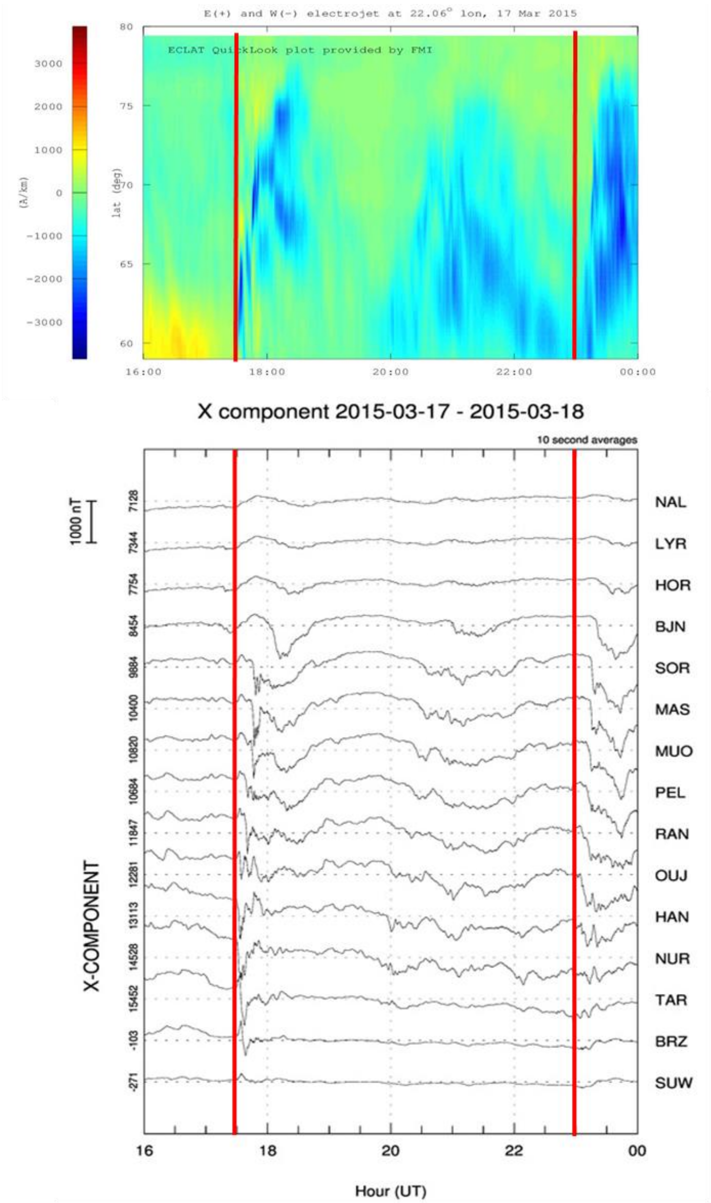

Fig. 2. Equivalent ionospheric currents (blue- negative, red -positive) – upper panel, and the X-component of the magnetic field at the IMAGE latitudinal chain SUW-NAL during the first two examined substorms on 17 March 2015 (bottom panel)

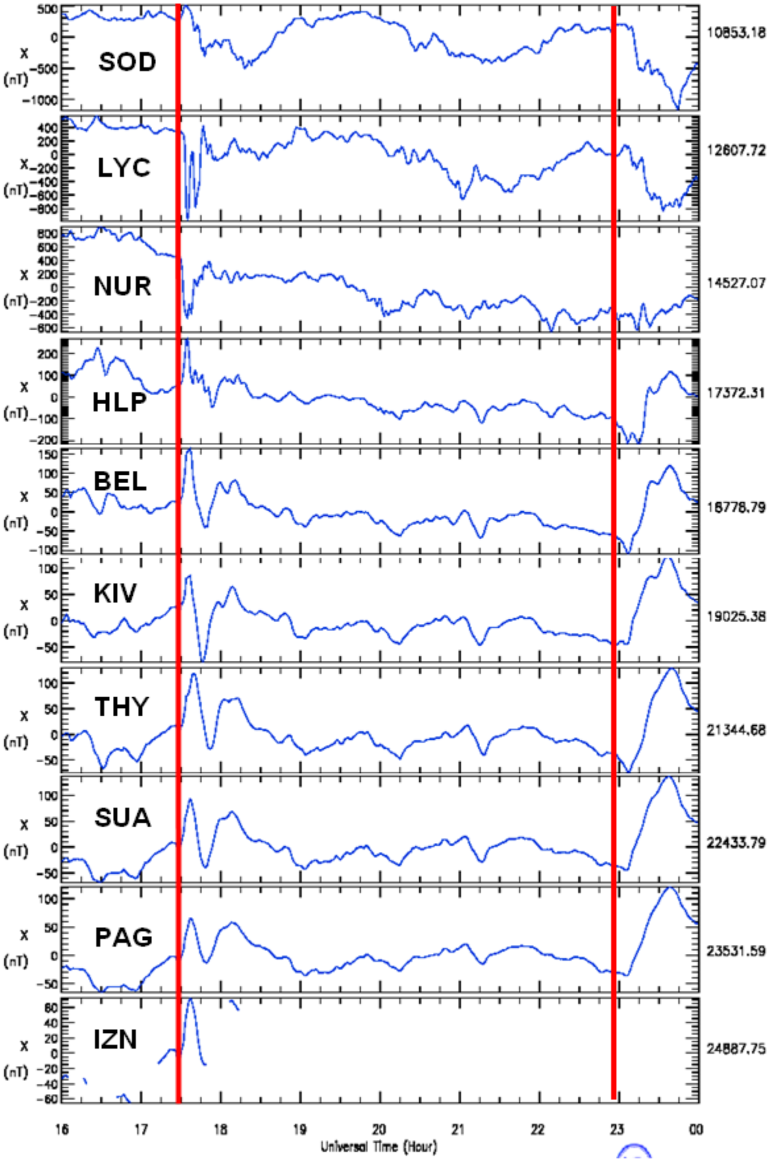

Fig. 3. X-component of the magnetic field at the selected INTERMAGNET stations during the examined substorms on 17 March 2015

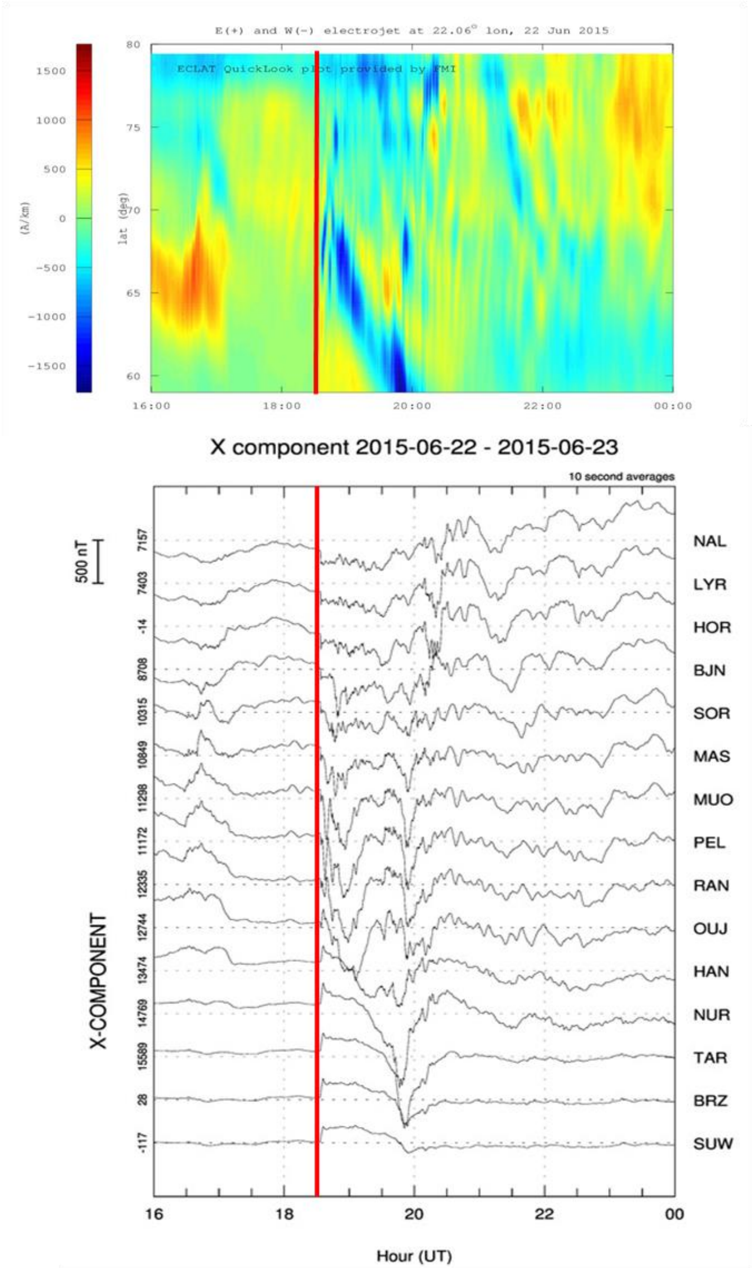

Fig. 4. Equivalent ionospheric currents (blue- negative, red -positive) – upper panel, and the X-component of the magnetic field at the IMAGE latitudinal chain SUW-NAL during the substorm on 22 June 2015 (bottom panel)

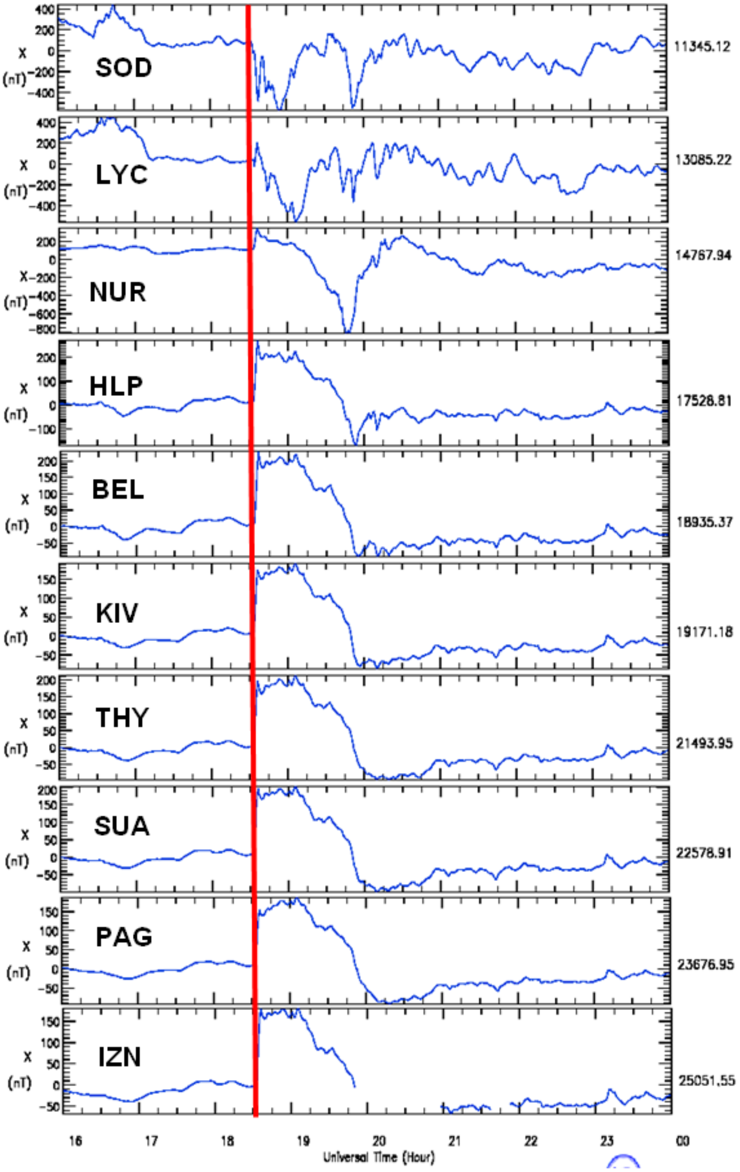

Fig. 5. X-component of the magnetic field at the selected INTERMAGNET stations during the substorm on 22 June 2015

Substorm at 22:55 UT on 17 March 2015

The second examined substorm on 17 March 2015 was developed also during the MC, in the main storm phase, close to the SYM/H_{min}. The following average IMF values were recorded: IMF B_T = 20.45 nT, IMF B_Y = –10 nT, IMF B_Z = –15 nT, V_X = –550 km/s. At the substorm onset, the SYM/H = –161 nT. The westward electrojet drifted fast to the North, from ~54° to ~72° CGMlat (Fig. 2, upper panel). The strong disturbances in the X-component begun at TAR (54.47° CGMlat), reached BJN (71.45° CGMlat) to the North and were observed up to BRZ (52.30° CGMlat) to the South (Fig. 2, bottom panel). A positive magnetic bay was registered at first at HLP (50.70° CGMlat) as well as in all mid-latitude stations to the South from HLP (Fig. 3), and also at the equatorial latitudes (AAE, not presented here). It lasted about 1 hour.

Substorm at 18:33 UT on 22 June 2015

This substorm was originated during SHEATH in the solar wind. Its onset was observed in the time when a shock wave (IS), third in this disturbed period, impacted the magnetosphere (Fig. 1, right panel). The shock arrival was characterized by a sharp increase of the solar wind parameters: the dynamic pressure jump was from 5 to about 60 nPa, the velocity X-component increased from 450 km/s to 700 km/s, the proton density – from 15 to 60 cm^{-3}, and the temperature – from $2*10^5$ to $1.4*10^6$ K. The magnitude of the IMF B_T enhanced from 10 to 45 nT, and the IMF B_Z turned southward at 18:39 UT and reached –40 nT at 19:22 UT. Prior to the onset, the average IMF and solar wind parameter values were: IMF B_T= 9.57 nT, IMF B_Y= –6 nT, IMF B_Z= –1.1 nT, V_X= –435 km/s. The fast decrease of the IMF B_Z and the change of its direction provoked the storm sudden commencement (SSC) at 18:33 UT. The SYM/H value sharply increased from –20 nT to 88 nT, after that decreased and at 19:18 UT became negative. Then the main storm phase began. The substorm onset followed the SSC, its development was in progress during the storm initial phase and continued further in the main phase.

The westward electrojet moved fast to the North from 62° ÷ 67° CGMlat at 18:33 UT and after a jerk reached the CGM latitudes of 75° and more. Simultaneously, the electrojet shifted to the South, to the CGM latitudes < 57° at 19:40–20:00 UT (the upper panel in Fig. 4). The perturbations in the X-component began at PEL (63.55° CGMlat), reached NAL (75.25° CGMlat) to the North and BRZ (52.30° CGMlat) to the South by the IMAGE latitudinal chain (bottom panel in Fig. 4). A positive magnetic bay was seen at the mid-latitude stations (Fig. 5) and equatorial stations (AAE, not presented here). The positive bay was registered at all stations southward from HLP (50.70° CGMLat). The bay lasted about 1.5 hours and was characterized by a sharp increase, followed by a gradual decrease.

Discussion

The considered substorms originated during the rather similar severe geomagnetic storms. One of its resemblances was a noticeable display of positive magnetic bays at middle and low latitudes. However, its onsets and further development have been observed under different interplanetary and geomagnetic conditions, which lead to the different onset locations and the different spatial dynamics of the westward electrojet, as well as to the differences in the substorms extent and the behavior of the middle and low latitude positive magnetic bays.

The substorms of 17 March 2015 occurred during MC, in the time of the main storm phase, under disturbed conditions, as indicated by the corresponding averaged IMF B_Z, V_X, and SYM/H values. The substorm onsets were located at ~57° and ~54° CGMlat, respectively, corresponding to an expanding auroral oval. The third substorm onset of 22 June 2015 has happened during SHEATH, and followed the interplanetary shock and the SSC. The average IMF B_Z and V_X values suggested relatively quiet interplanetary conditions prior the substorm. Perhaps, for that reason, the auroral oval was not so expanded as in the first two events and the substorm onset was at higher CGM latitude, at ~63–64°. (Note, the substorms of 17 March 2015 developed in the main storm phase).

In the first two events, the sharp motion of the west electrojet could be observed to the North direction up to ~70–71° CGMLat (upper panel of Fig. 2). The strong X-component magnetic perturbations on the ground reached 71° CGMlat (bottom panel of Fig. 2), a slower drift to the South was registered simultaneously as well. Such behavior is typical for the “classical” substorms.

During the substorm of 22 June 2015, the considerable movement of the westward electrojet to the South and North was observed (Fig.4, upper panel). The significant travel of the substorm to the South has happened, probably, due to the change of the IMF B_Z sign from positive to negative up to –40 nT. After the second jump of the electrojet to the North, its progress surpassed the 75° CGMlat. The electrojet center reached the station LYR (75.12° CGMlat). Such substorm behavior allows ranking this substorm among the “expanded” substorms [11].

The positive magnetic bays observed at the middle latitudes during the first two substorms, were nearly symmetric, and the duration of the perturbation was about 20 min and 1 hour, correspondingly. The positive bay during the third substorm was characterized by a sharp increase, as a result of the association of the substorm onset with the IS and SSC, and by a gradual decrease later.

The boundary between the negative and positive bays was observed in the latitude range of 50 ÷ 56° CGMlat (between the stations HLP and NUR). According to the McPherron et al. [12] scheme, this boundary could be mapped between the electrojet location and the field aligned currents during the considered substorms.

Conclusion

In this work we analyzed the strongest geomagnetic storm in the current 24th solar cycle – the storm of 17 March 2015 (Ap = 108) [19, 20]. It, together with the storm of 8 September 2017 (Ap = 106), represents the two extreme (G4 – level) manifestations of the geomagnetic activity of the 24th cycle during solar maximum and minimum respectively [19–22].

Also examined is the 2015 summer solstice storm of 22–23 June (Ap = 72), which is the sixth major geomagnetic storm (also G4 – level) of solar cycle 24 https://www.spaceweatherlive.com/.

Our main contributions are as follows:

- The middle and low latitudes substorms demonstrate the positive sign of the magnetic X- component. The magnetic bay sign changed from negative to positive between 50° and 56° CGMlat (between HLP and NUR sta-tions);
- The clear effect of the magnetic storm Sudden Commencement (SSC) was expressed by the rapid substorm shift from the auroral to low latitudes and the sharp increase of the substorm intensity on 22 June 2015. The larger amplitude and longer duration of the positive magnetic bay on 22 June 2015 are, probably, due to its development in the SHEATH versus the development in the MC of the substorms on 17 March 2015.
- It is seen that certain interplanetary conditions (SHEATH + IS) during the storm on 22 June 2015 led to a substorm that manifested itself at low latitudes (positive bays), and also at high latitudes (so called “expanded” substorms);
- The substorms during the storm on 17 March 2015 were observed at low and auroral latitudes too, but without the high-latitude expansion, perhaps, this is connected with the development of these substorms during the magnetic cloud (MC). Thus, they appear “classical” substorms.

The research conducted here will be expanded to other strong storms of the 24th solar cycle, for example the G4 – Severe geomagnetic storm on September 7–8 2017 and other interesting cases.

References

1. Akasofu, S.-I. The development of the auroral substorm, Planetary and Space Science, 1964, 12, 4, 273–82. https://doi.org/10.1016/0032-0633(64)90151-5
2. Feldstein, Y. L. and G. V. Starkov. Dynamics of auroral belt and geomagnetic disturbances, Planetary and Space Science, 1967, 15(2), 209–229.
3. Kisabeth, J. L. and G. Rostoker. The expansive phase of magnetospheric substorms: 1. Development of the auroral electrojets and auroral arcs configuration during substorm, J. Geophys. Res., 1974, 79, 7, 972–84.

4. Wiens, R. G. and G. Rostoker. Characteristics of the devel opment of the westward electrojet during the expansive phase of magnetospheric substorms, J. Geophys. Res., 1975, 16, 2109–28.
5. Sergeev, V. A., A. G. Yakhnin, and N. P. Dmitrieva. Substorms in the polar cap are effect of high-speed streams of the solar wind, Geomag. Aeron., 1979, 19, 1121–22. (in Russian)
6. Wang, H., H. Lühr, S. Y. Ma, and P. Ritter. Statistical study of the substorm onset: its dependence on solar wind parameters and solar illumination, Ann. Geophys., 2005, 23, 2069–79. doi:10.5194/angeo-23-2069-2005
7. Despirak, I. V., A. A. Lubchich, A. G. Yakhnin, and H. K. Biernat. Poleward expansion of the westward electrojet depending on the solar wind and IMF parameters, Geomagnetism and Aeronomy, 2008, 48, 3, 284–92.
8. Tanskanen, E. I., T. I. Pulkkinen, A. Viljanen, K. Mursula, N. Partamies, and J. A. Slavin. From space weather toward space climate time scales: Substorm analysis from 1993 to 2008, J. Geophys. Res., 2011, 116, A00I34. doi:10.1029/2010JA015788
9. Clausen, L. B. N., S. E. Milan, J. B. H. Baker, J. M. Ruohoniemi, K. H. Glassmeier, J. C. Coxon, and B. J. Anderson. On the influence of open magnetic flux on substorm intensity: Ground- and space-based observations, J. Geophys. Res., 2013, 118, 2958–69. doi:10.1002/jgra.50308
10. Despirak, I. V., A. A. Lubchich, and N. G. Kleimenova. Polar and High Latitude Substorms and Solar Wind Conditions, Geomag. Aeron., 2014, 54, 5, 619–26.
11. Despirak, I. V., A. A. Lubchich, and N. G. Kleimenova: High-latitude substorm dependence on space weather conditions in solar cycle 23 and 24 (SC23 and SC24), J. Atmos. Sol. Terr. Phys., 2018, 177, 54–62**.**
12. McPherron, R. L., C. T. Russell, and M. P. Aubry. Satellite studies of magnetospheric substorms on August 15, 1968. 9. Phenomenological model for substorms, J. Geophys. Res., 1973, 78, 3131.
13. Akasofu, S.-I., S. Chapman, and C. I. Meng. The polar electrojet, Journal of Atmospheric and Terrestrial Physics, 1965, 27, 11–12, 1275–1305. doi:10.1016/0021-9169(65)90087-5
14. Akasofu, S.-I. and C. I. Meng, A study of polar magnetic substorms, J. Geophys. Res., 1969, 74, 1, 293–313. doi:10.1029/JA074i001p00293
15. Meng, C. I. and S.-I. Akasofu, A study of polar magnetic substorms .2. 3-dimensional current system, J. Geophys. Res., 1969, 74, 16, 4035–53. doi:10.1029/JA074i016p04035
16. McPherron, R. L., Chu, X., The mid-latitude positive bay and the MPB index of substorm activity, Space Science Reviews, 2017, 206, 1–4, 91–122. doi:10.1007/s11214-016-0316-6
17. McPherron, R. L. and X. Chu, The midlatitude positive bay index and the statistics of substorm occurrences, J. Geophys. Re.: Space Physics, 2018, 123, 2831–50. doi:10.1002/2017JA024766
18. Yermolaev, Yu. I., Nikolaeva, N. S., Lodkina, I. G., Yermolaev, and M. Yu., Catalog of large-scale solar wind phenomena during 1976–2000. Cosmic Research (Engl. Transl.), 2009, 47, 81–94.

19. Tomova, D., P. I. Y. Velinov, and Y. Tassev, Energetic evaluation of the largest geomagnetic storms of solar cycle 24 – on March 17, 2015 and September 8, 2017 during solar maximum and minimum respectively. C. R. Acad. Bulg. Sci., 2017, 70, 11, 1567–78.
20. Tomova, D., P. I. Y. Velinov, and Y. Tassev, Comparison between extreme solar activity during periods March 15–17, 2015 and September 4–10, 2017 at different phases of solar cycle 24, Aerospace Res. Bulg., 2017, 29, 10–29.
21. Tassev, Y., P. I. Y. Velinov, D. Tomova, and L. Mateev, Analisis of extreme solar activity in early September 2017: G4 – Severe geomagnetic storm (07–08.09) and GLE72 (10.09) in solar minimum, C.R. Acad. Bulg. Sci., 2017, 70, 10, 1437–44.
22. Velinov, P. I. Y. and Y. Tassev, Long term decrease of stratospheric ionization near the 24-th solar cycle minimum after G4 – Severe geomagnetic storm and GLE72 on September 8–10, 2017, C.R. Acad. Bulg. Sci., 2018, 71, 8, 1086–94.

ПРОЯВА НА СУББУРИ НА ВИСОКИ И СРЕДНИ ШИРИНИ ПО ВРЕМЕ НА ДВЕ СИЛНИ МАГНИТНИ БУРИ

В. Гинева, И. Деспирак, Н. Клеймьонова

Резюме

Анализирана е динамиката на магнитните суббури на високи и средни ширини по време на две силни геомагнитни бури, на 17 март 2015 г. и 22 юни 2015 г. Двете бури са доста подобни: и двете са резултат от въздействието на Sheath област в слънчевия вятър, и двете се характеризират с висока интензивност ($SYM/H_{min} < -200$ nT). Ние изучихме магнитните суббури по време на тези бури на основата на данните от мрежите станции INTERMAGNET и IMAGE. Съпътстващите параметри на слънчевия вятър и междупланетното магнитно поле (ММП) бяха взети от базата данни OMNI. Пространствено-временната динамика на три суббури беше изучена подробно: суббурите от 17:29 UT и 22:55 UT през първата буря и от 18:33 UT през втората буря. Суббурите на 17 март 2015 г. възникнаха през главната фаза на бурята, а началото на суббурята на 22 юни 2015 г. беше след внезапното начало (SSC) на втората буря. И трите суббури се характеризират с рязко разширяване към полюса на западния електроджет едновременно с по-бавно движение към по-ниски ширини. Те бяха наблюдавани също така на средни и ниски ширини като положителни магнитни „заливи”. Западният електроджет достигна ~71° CGMlat през първите две суббури и задмина 75° CGMLat през третата суббуря. Следователно, първите две събития са „класически” суббури, а третото – „разширена” суббуря. Ние предполагаме, че това поведение е свързано с различните условия в слънчевия вятър: „класическите” суббури се развиват при магнитен облак (MC), а „разширените” – под въздействието на Sheath областта.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e04

THE GLOBAL TEMPERATURE ANOMALIES RELATED TO THE SLOWDOWN OF ATMOSPHERIC CO2 CONCENTRATION OBSERVED FROM 1939 UP TO 1950

Rolf Werner, Veneta Guineva

Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: rolwer52@yahoo.co.uk

Keywords: *Global temperatures, Atmosphere, Temperature anomalies,* CO_2*.*

Abstract

It is very well known that carbon dioxide (CO_2) accumulated in the atmosphere is the main climate driver. The first precise direct continuous measurements of the atmospheric CO_2 concentration were provided from Keeling since 1958 at Mauna Loa Observatory. The measurements are going on up to day. Law Dom ice core drilling was started in 1969 by the Australian ANARE program.

The slowdown of the World economic development during the World War I, the Great Depression and the World War II lead to a deceleration of the CO_2 emissions. The integration of the total CO_2 emissions using the impulse response function concept shows that the observed slowdown of the CO_2 emission is not sufficient to explain the CO_2 plateau and additional CO_2 sinks are necessary. Based on multiple regression models adjusted global temperatures were determined by removal of temperature influences other than related to CO_2. The adjusted temperatures follow close the CO_2 radiation term. The difference between the estimated adjusted temperature time evolution with and without the CO_2 slowdown and also the short time trends demonstrate very clear the close relation between the temperature change and the CO_2 radiative forcing. It is shown that the slowdown of the CO_2 emission in the period from 1939 to 1950 and the related CO_2 concentration in the atmosphere, caused at least partially by human activities, generate slower increase of the temperature anomalies. Consequently CO_2 is the leading variable of the relation surface temperature – CO_2.

Introduction

Global warming is usually perceived as the increase of the average global temperature over a long period of time, usually 30 years or more. The global temperatures inferred from direct observations show a rise of approximately 0.7 °C during the 20th century. The temperatures however do not increase continuously. Longer periods of warming alternate with periods in which the temperature does not increase or a cooling is observed. The global temperature series are characterized also by variations on the scale from decades to years associated with El Niño events, with volcano eruptions and with changes in the solar activity. Lean and Rind [1] explained 76% of the temperature variance by indices describing anthropogenic forcing, El Niño/ Southern Oscillation (ENSO) events, volcano activity and solar irradiance changes using multiple regression method analysing the monthly combined land and ocean surface temperature series (HadCRUT3 vcgl) for the period 1889–2006. Foster and Rahmstorf [2] have performed multiple regressions of five different temperature series for the time span of 1979–2010. They have constructed adjusted data sets by removal of the influence of ENSO, volcanic eruptions and the total solar irradiance on the temperatures. They stated that the trends of the adjusted series were linear after 1998 when the greenhouse gases concentrations increased as before, but the rise of the global temperatures seems to be slower in comparison with the period before. It is well known that the oceans are thermally inert and provide an important memory of the climate variations. The inter-decadal Pacific oscillation signal has quasiperiodic variations of the time scale of 15–30 years and is connected to other teleconnections.

The Atlantic multi-decadal oscillation (AMO) was detected by Schlesinger and Ramankutty [3]. It is defined by the Sea surface temperature (SST) anomaly pattern in the Northern Atlantic, where the global warming influence was removed using different methods (see [4] for more details). The AMO shows a periodicity of approximately 65–70 years. The reasons of AMO multi-decadal periodicities are not fully understood until now and they are discussed controversially. Some authors state that the origin of AMO is an intrinsic mode in the dynamic of the ocean-atmosphere system without external forces. It is shown that AMO is driven by the Atlantic meridional overturning circulation [5]. Knudsen et al. [6] explicitly reject the hypothesis that AMO was forced by changes in the solar activity. Other authors found a link with the solar activity long-term changes at multidecadal scales, e.g. [7]. Different studies have demonstrated relations between regional weather phenomena and AMO ([8] and the citations herein). Rohde et al. [9] found an AMO signal in the residuals of the Berkeley Earth global land temperature series modelled by the logarithm of the CO_2 concentration and by volcanic sulphate emissions. For the first time Zhou and Tung [10] included the AMO index in a multiple regression model as a predictor and determined the global warming rate.

In the science community a wide consensus consists about the origin of the climate change by the increase of the man-made CO_2 emission in the atmosphere. The emitted CO_2 passes into the global carbon cycle, where about 45% of it remains in the atmosphere [11]. The remaining in the atmosphere CO_2 absorbs the long-wave radiation coming from the Earth surface. One part of the absorbed radiation is trapped as heat in the atmosphere and another part is re-emitted in all directions, downwards and upwards. This process is the basic understanding of the warming by greenhouse gases. Up to now there is not an exact proof of the real causes of the climate change by greenhouse gases. In the IPCC reports the term "evidence substantiated by simple physical models" was used. The probability of the man-made individual climate change elements is evaluated by climate experts. The evidence of man-made origin of the climate change is demonstrated also by certain patterns in the spatial, temporal or spatial-temporal domain, obtained by climate models and confirmed by observations. One of the most known patterns (or so called finger prints) is e.g. the increase of the temperature in the lower troposphere and the simultaneous decrease of the temperature in the stratosphere caused by CO_2 taking into account solar, volcanic and ozone temperature effects as well [12]. Another example is the substantial rising of the occurrence of the number of warm days per decade during the period 1951–2003. Simultaneously a slight decrease of the number of cold days and of the diurnal temperature range was observed meaning that the atmosphere warming is faster during the day. This fingerprint was found both by climate model simulations and by investigations of the experimental climate observations [13, 14]. The isotope signature of ^{13}C and ^{14}C in the atmosphere is changing (Suess ^{14}C and ^{13}C effect). ^{14}C isotope content in the atmosphere comes by galactic cosmic rays. The ^{14}C isotope has a decay time of more than 5 700 years. Therefor fossils do not contain ^{14}C. By the CO_2 release in the atmosphere due to fuel combustion the amount of ^{14}C decreases. The ^{13}C isotope concentration is reduced compared to the ^{12}C isotope atmospheric concentration, because millions of years ago fuel was formed by plants, accumulating more of the lighter ^{12}C isotope than of the heavier isotope ^{13}C. This change of the carbon isotopes composition in the atmosphere is a fingerprint of the human activity.

The main goal of this paper is to present a new evidence about the causality between the growing of the atmospheric CO_2 concentration and the temperature increase with the leading role of CO_2.

CO_2 emissions and concentration in the Earth atmosphere

The CO_2 concentration in the Earth atmosphere before the industrial revolution, during the Holocene was almost constant. With the beginning of the industrial revolution after the discovery of the first commercial steam-engines the CO_2 emissions were increasing slowly with the developing of the world economy. The CO_2 amount in the atmosphere is determined by a complicated global balance of the CO_2 exchange between the ocean, biosphere, atmosphere and lithosphere, involving diffusion, advection and dissolving processes. Simple empirical models approximate the carbon concentration as a sum of exponentially decaying pulse response functions, where the time decays are of the order of one up to some hundred years [15]. The global total emissions of CO_2 since 1850 are presented in Fig. 1. The global emissions of CO_2 up to about 1910 were dominated by changes in land use [16]. However the increase of the annual global CO_2 emissions is determined basically by the global emitted CO_2 from fuel combustion, as shown in Fig. 1. The exponential increase of the CO_2 emissions during the industrial revolution is interrupted by a period characterised by a slowdown of the emissions throughout the World Wars I and II (WWI and WWII) and the Great Depression [17, 18]. During this time the global annual land use CO_2 emissions are nearly unchanged at the level between 0.8 and 0.9 GtC/yr (see Fig. 4).

After the WWII the national economies rapidly restored and a boom of the worldwide economic development was registered. Sometimes this period is called Great Acceleration. The rapid economic growth was interrupted by the first and later by the second oil shocks. These economic developments are connected to the corresponding energy consumption and reflect in the CO_2 emissions (see Fig. 1).

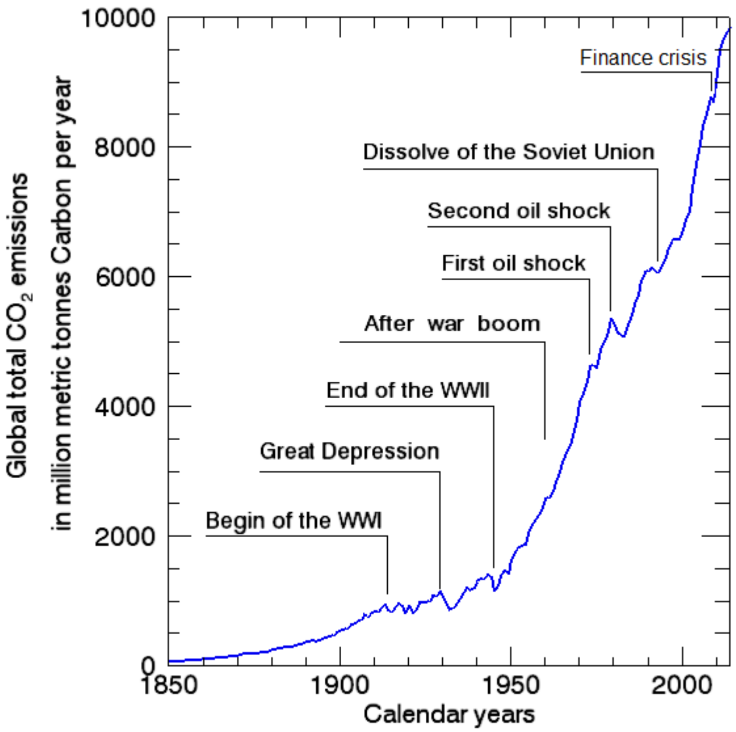

Fig. 1. Total CO_2 emissions including burning of fossil fuel and cement production at the globale scale. The figure is drawn based on the Carbon Dioxide Information Analysis Center (CDIAC) data (http://cdiac.ornl.gov/trends/emis/tre_glob_2013.html) and was redrawn from: http://wiki.bildungsserver.de/klimawandel/index.php/ Datei:FossileEnergie1850–2007.jpg#file.

The concentration of CO_2 accumulated in the atmosphere shows not or only slow increase after 1939 and doesn’t achieve the value of 1939 until 1950 as is seen in the CO_2 data compilation from Hansen (http://www.climateaudit.info/ data/hansen/giss_ghg.2007.dat, see also [19]).

MacFarling Meure et al. [20] stated that the CO_2 stabilization between 1940 and 1950 is a notable feature of the ice core measurements and it was verified by newer high density measurements. They confirm this stabilization at 310–312 ppm between 1940 and 1955 (see Fig. 1 in the previously cited paper [20]). For comparison in Fig. 2 the atmospheric CO_2 concentration obtained by different methods throughout different campaigns is presented. Indirect determination of atmospheric CO_2 by analysing air bubbles trapped in polar ice at the Low Dom station in the East Antarctic as well as direct atmospheric CO_2 observations at Mauna Loa beginning in 1958 and at Cape Grim (Tasmania, Australia) from 1978 up to 2015 are included (see also [21]). The uncertainties of the ice core measurements are 1.2 ppm [22].

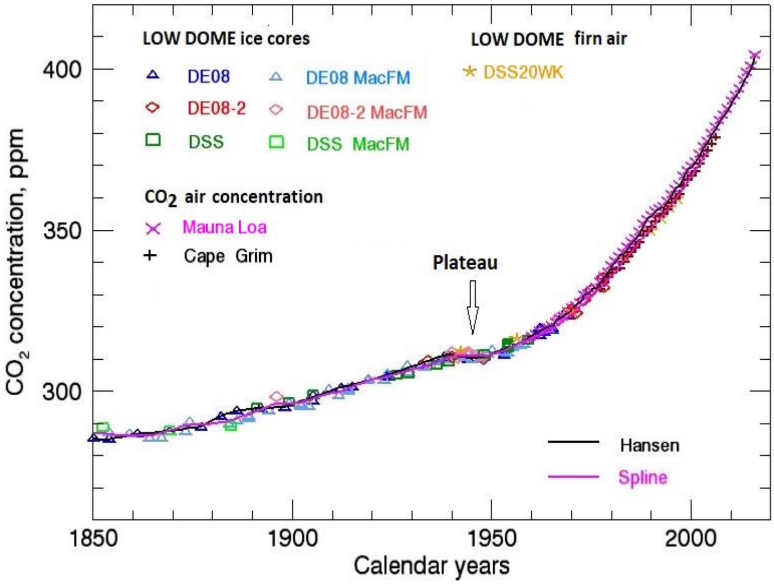

Fig. 2. Increase of the CO_2 concentration in the Earth's atmosphere established by indirect measurements in ice cores DE08, DE08-2 and DSS at the Antarctic station Low Dom since 1850 and established annual means of the CO_2 concentration by direct measurements at Mauna Loa and at Cape Grim, where the CDIAC data and data of the CSIRO Marine and Atmospheric Research and the Australian Bureau of Meteorology were used. The continuous black line is the data compilation of [19], which was extended by means of regression up to 2016. The continuous line in magenta presents a spline attenuate variations with periods of < 20 years by 50% as given by [19], based on [22].

The direct CO_2 concentration measurements are much more accurate and with uncertainties better than 0.2 ppm. (https://www.esrl.noaa.gov/gmd/ccgg/aboutco2_measurements.pdf).

The ice core measurements are smoothed (e.g. by about 10 years for the DE08 core) due to diffusion processes filtering out short term atmospheric CO_2 variations. The atmospheric measurements at Cape Grim show slightly smaller CO_2 concentrations in comparison to the Mauna Loa results, due to the geographical locations – Cape Grim at 40°38' S and Mauna Loa observatory at 19°28' N. The CO_2 concentrations observed at Mauna Loa observatory are closer to the global mean CO_2 amounts in the atmosphere. Therefore these data and the extended by linear regression Hansen's data compilation are used in this paper.

The remaining in the atmosphere amount of CO_2 is given by the imbalance between the CO_2 sources and sinks. The CO_2 sources are the total CO_2 emission consisting mainly from the fuel combustion and the cement production, and also from the land use change, e.g. by deforestation. The trees are burned or let to rot, whereby the CO_2 storage in the plants is released to the atmosphere. Biomass burning immediately leads to increase of CO_2 concentration in the atmosphere. On the other hand by burning of vegetation the carbon sink is destroyed for a long time.

The inventory of CO_2 the ocean by solution of CO_2, the photosynthesis of terrestrial plants and oceanic phytoplankton represent the main CO_2 sinks. For more details see e.g. [23]. The relative part of the carbon emission remaining in the atmosphere is called Airborne Fraction (AF).

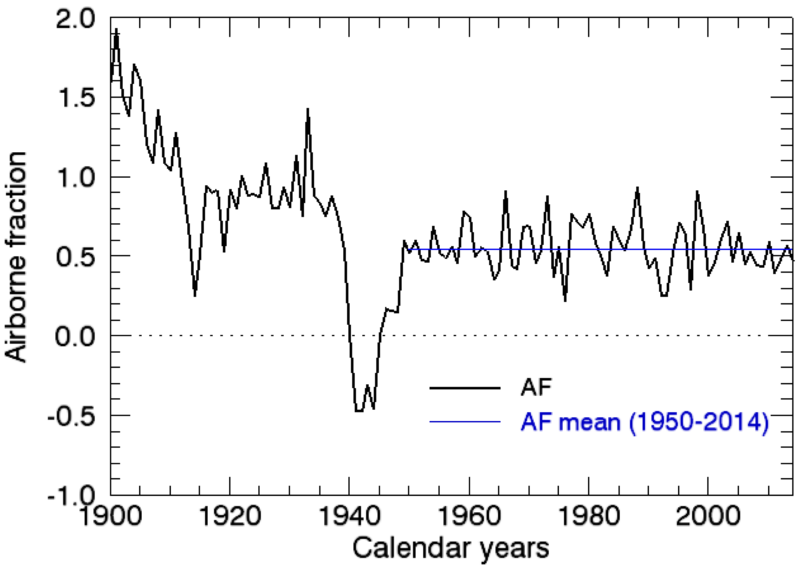

Fig. 3. Airborne fraction related to the CO_2 fuel combustion

The IPCC Fourth Assessment Report (p. 139) [24] follows Keeling [25], who defined the AF in relation to the CO_2 emission taking into account the fossil fuel and cement production. Later the estimation quality of the emission was improved by land use changes and they were included in the CO_2 emissions. Jones et al. [26] have shown that an exponential rise of the CO_2 emissions leads to a constant AF. Here the AF was calculated by the ratio of the first backward differences of the CO_2 and the total CO_2 fuel emission. From 1900 up to 1959 the AF shows variations partially caused by the measurements errors of CO_2 determination from ice cores or by small values of the Carbon emissions. However, strong AF values > 1 indicate that more CO_2 remained in the atmosphere than the amount added by Carbon fuel emission. The total fuel carbon emissions and the carbon emissions by land use are shown in Fig. 4. It is seen that up to about 1910 the emissions by land use are greater than Carbon fuel emission (see also [27]). Negative AF as observed between 1940 and 1945 is caused by a decrease of CO_2 in the atmosphere. At the same time the CO_2 emissions are almost constant. The negative AF or AF values near zero indicate an additional strong CO_2 sink (or they could be caused by higher measurement errors, as well). A significant trend in the AF would indicate a CO_2 feedback (e.g. decrease of the CO_2 ocean uptake). Since the beginning of the direct CO_2 concentration measurements the fraction of the fossil fuel emissions for the time interval up to 2014 shows variations around the mean of 0.55 over the time [24]. The observed here AF seems to be reasonable since 1950 (see Fig. 3). The mean value of the AF is 0.55 ± 0.12. The maximal deviations in 1973, 1988 and 1998 are related to anomaly CO_2 growth rates caused by strong ocean CO_2 uptakes variations [22, 25, 28, and 29]. It was established that the strong El Niño events in 1895–1898, 1911–1916 and 1940–1942 are in coincidence with an increase of the atmospheric CO_2 growth rate. The removal of atmospheric CO_2 by uptake into the ocean is strong, when the Southern oscillation (SO) shifts from its warm phase (El Niño event) to its cool phase (La Niña event) and coincided with a warm to cool phase change of the Pacific Decadal Oscillation (PDO) and with lower temperatures and progressive weakening of the Atlantic thermohaline circulation [20]. The sea-surface warming during El Niño reduced the upwelling of CO_2 rich water which resulted in a reducing of CO_2 outgassing and an enhancement of the CO_2 uptake, respectively [30]. The minimum of the AF in 1992/1993 is related to the Monte Pinatubo eruption and the maximum observed around 2003 is connected to anomalies during the strong heatwave observed in Europe [31].

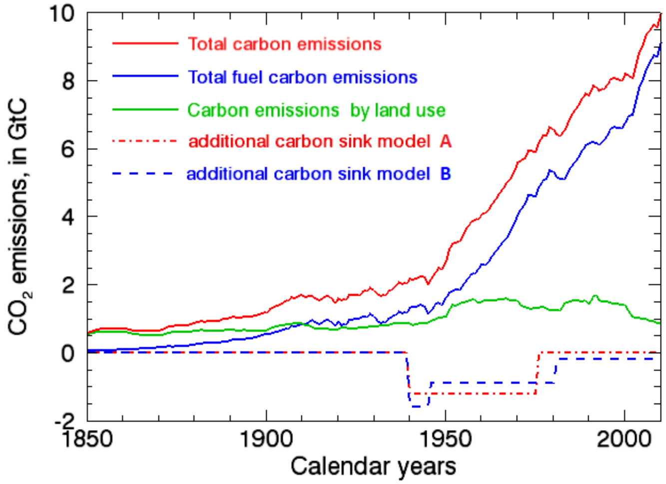

Fig. 4. Global total CO_2 emission by fuel combustion (blue line), the emission by land use changes (green line) and the total CO_2 emissions as the sum of the fuel combustion and land use change (red line). Models for additional carbon sinks (negative values for emissions) are drawn by dashed dotted red line for model A and by dashed blue line for model B.

In contradiction to the argument of strong ocean uptake, Rafelski et al. [27], Truderinger et al. [32] and Rubino et al. [33] concluded, that the plateau in the CO_2 atmospheric concentration after 1940 is likely caused by land air temperature decreasing over the Northern Hemisphere. Bastos et al. [34] have found that cropland abandonment in the Former Soviet Union, as a consequence of the WWII, could explain the CO_2 plateau, due to increase of CO_2 absorption by vegetation recovery and by increase of organic matter in soil.

More detailed studies have shown later that the human part on the additional CO_2 sink by the cropland due abandonment in the Former Soviet Union is about 6–10 % of the gap sink required to explain the plateau [35]. Bastos et al. [35] have outlined that it is likely decreases in agricultural areas during WWII might have also occurred in other regions and could additionally contribute to the sink gap. Such events may not be included in FSU–REF data due to the use of different sources of information as is the case, for example, of China. They claimed that the part of the additional sink of about 60% has a natural source. The reason of the CO_2 stabilization up to now is not fully understood.

The response of the concentration of CO_2 in the atmosphere to changes in the carbon emissions can be estimated in the first order using the concept of the Impulse Response Function (*IRF*) (or Green's function), where CO_2 can be represented by the sum of earlier anthropogenic emissions e at time t' multiplied by the fraction remaining still airborne after time $t-t'$, given by the IRF [36]:

$$CO_2 = c \cdot \int_{t_0}^{t} e(t')IRF(t-t')dt' + CO_2(t_0), \tag{1}$$

where the IRF is given by a sum of exponential functions

$$IRF(t) = a_0 + \sum_{i=1}^{n} a_i \cdot \exp\left(\frac{-t}{\tau_i}\right) \qquad for\ t \geq 0,$$

and a_0 is the part which remains permanently in the atmosphere and a_i are the fractions associated with the time scale τ_i. *IRF* is not an invariant function, but depends on the magnitude of the carbon emissions [37]. If the carbon emissions are given in GtC then the factor *c* is about 0.47 ppmv/GtC to obtain the CO_2 concentration in ppmv. This constant can be used for tuning to achieve the best adjustment. The *IRF* concept was developed for cost effective estimations of the atmospheric CO_2 concentration as response to different future anthropogenic emission scenarios and describes the CO_2 ocean uptake. Biological sources and sinks are not included in the model to obtain the *IRF* [37]. Using *IRF* parameters of the standard Bern SAR model the integration of the CO_2 emissions (see unfccc.int/resource/brazil/carbon.html) gives a result close to the obtained one by

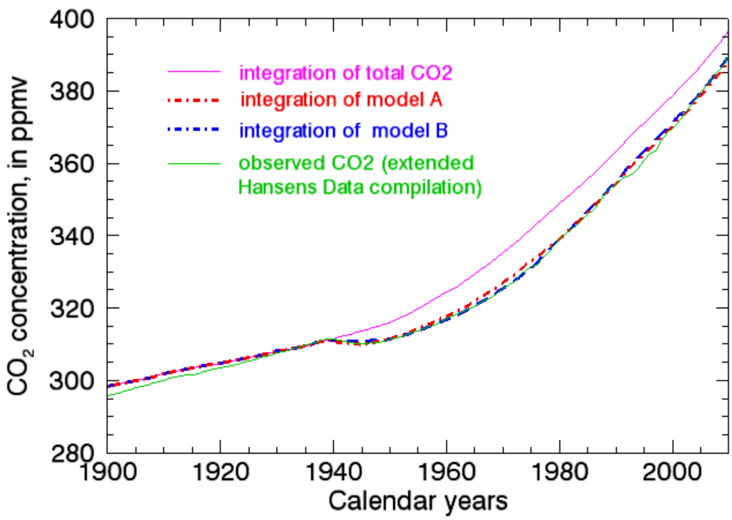

Fig. 5. Results of the integration of the total CO_2 emissions (defined as the sum of the total CO_2 fuel emission and the emission by change of land use) (continuous magenta line), the integration of the total CO_2 emissions reduced by additional sinks model A (dashed dotted red line), and model B (dashed dotted blue line) in comparison with the observed CO_2 concentration (using the extended Hansen's data compilation) (continuous green line)

ice core, firn and atmospheric measurements (as obtained by the extended Hansen's data compilation). By the integration of the original data consisting of the sum of the total fuel emissions and the land use emissions no plateau can be observed (the magenta line in Fig. 5). The slowdown of the emissions during the period between the beginning of the WWI and the end of the WWII lead only to a stabilization of the grow rate from about 0.5 Gt carbon (0.24 ppmv CO_2) per year up to approximately 1.0 Gt carbon (0.47 ppmv CO_2). After this the grow rate fast increase up to 4 Gt carbon (1.9 ppmv CO_2) per year. However if the total carbon fuel and land used emissions during the interval between 1940 and 1975 are reduced yearly by 1.2 Gt carbon (see Fig. 4, sink model A) then the integration results form a plateau very close to the ones obtained by the ice core and firn air CO_2 measurements (the dashed dotted red line in Fig. 5). (The total sink corresponds to about 43 GtC, see Fig. 4). Here of course the precise shape of the function describing the carbon reduction cannot be reconstructed, by reason of the insensitivity of the integration. A very like result for the total CO_2 emissions is formed by a strong carbon reduction of 1.6 Gt carbon per year during 1940–1945 following by a permanent sink of 0.9 Gt carbon per year up to 1980 and after this a sink of 0.2 Gt carbon (see Fig. 4, sink model B), with total carbon sink of about 47 GtC (the dashed dotted blue line in Fig. 5). The goal of this work was not to obtain a good adjustment to the measured atmospheric CO_2. The aim was to demonstrate that the plateau can be observed only, if the carbon reduction takes place suddenly by a sink formed by one longer impulse or by one stronger short impulse followed by a stepwise sequence of weaker impulses. Of course this is true only in the framework of the used model (based on IRF). It has to be mentioned that from the slowdown of CO_2 emissions directly follows an equivalent slowdown of the radiation forcing function [38].

In the following section the influence of the CO_2 concentration observed by measurements on one hand and estimated by the integration of observed total carbon emissions (the sum of the total fuel emissions and the land use emissions) on the other hand, on the temperature is estimated using linear multiple regression model.

Regression model

The long-time global temperature trends are determined by the concentration of the carbon dioxide (CO_2) in the atmosphere. In our regression model the CO_2 data compilation from Hansen is used (http://www.climateaudit.info/data/hansen/giss_ghg.2007.dat). The data, covering the period 1850–2006, were extended to 2016 adding the obtained annual CO_2 by regression of the Hansen data against the CO_2 measurements provided by the Mauna Loa observatory (ftp://ftp.cmdl.noaa.gov/ccg/co2/trends/co2_annmean_mlo.txt).

To study the influence of different factors on climate Werner et al. [39] were used the following linear regression model:

$$(2) \qquad T_{obs} = const + \beta_1 * \ln(CO_2 / 280\, ppmv) + \beta_2 * AMO + \beta_3 * PDO + \beta_4 * TSI + \\ + \beta_5 * SO + \beta_6 * AOD + \varepsilon,$$

where at the left hand side T_{obs} is the observed annual temperature anomaly and the regressors at the right hand side are the logarithm of CO_2 referred to its pre-industrial value of 280 ppmv, the Total solar irradiances (TSI), and the indices of the AMO, of the Pacific decadal oscillation (PDO), of the Southern oscillation (SO) and of the Aerosol optical depth (AOD).

To avoid collinearity in the regression other greenhouse gases except CO_2 were not included in the regression because the time evolution of their atmospheric concentrations is similar to the one of CO_2. Moreover other radiative active gases as methane (CH_4), halocarbons and nitrous oxide (N_2O) have a higher warming potential per molecule than CO_2, but their amounts in the atmosphere are much smaller. So the climate change is driven mainly by CO_2 [40]. The CO_2 forcing is defined as $\Delta F = 5.35 * \ln(CO_2 / 280 ppmv)\, Wm^{-2}$ [40]. In our computations, the used time interval was limited from 1900 up to 2010, because the temperature observations are more accurate since the beginning of 1900 than before and the carbon emissions are available up to 2010. The climate sensitivity can be easily determined by the linear regression equation with the help of the relation $\Delta T = \lambda\, \Delta F$. Werner et al. [39] have shown that the residuals in Eq. 1 can be described by an AR(1) model and the time series size can be replaced by its effective number [41]. By means of the effective sample size the effective standard deviations and the Students' effective t-values were calculated. The temperature sets of the leading climate centres as NOAA and Met Office are close to each other and do not give different results. In view of this, here only one temperature set, namely the HadCRUR4 temperature set of the UK Met Office is used.

In Table 1 the results are summarized including only the statistically significant regressors at the confidence level of 0.95, where the critical t-value is about 1.96. The corresponding to the 1-σ error effective *t*-value, taking into account the residuals autocorrelation, was denoted as t_{eff}. A stepwise regression with backward elimination of the most non-significant term was performed. To estimate the temperature anomalies, two models were applied. In the first model the real atmospheric CO_2 concentration was used as regressor, and an explained variance of 0.926 was achieved. In the second model the total CO_2 emissions obtained from equation 1 were used as regressor, an explained variance of 0.923 was observed. Both models distinguished not significantly one from the other and both statistic models describe the observed temperature anomalies very well.

Table 1. Results of the stepwise regression Eq. 2

note		β_1	β_2	β_4	β_5	R^2
Model 1 Atmospheric CO_2	Regr. coeff. σ $t_{eff.}$	2.734 0.097 22.20	0.488 0.038 10.10	0.058 0.017 2.62	-0.028 0.006 -3.53	0.926
Model 2 Integrated total CO_2 (fuel and Land use)	Regr. coeff. σ $t_{eff.}$	2.483 0.090 22.10	0.533 0.038 10.70	0.040 0.018 1.74	-0.027 0.006 -3.22	0.923

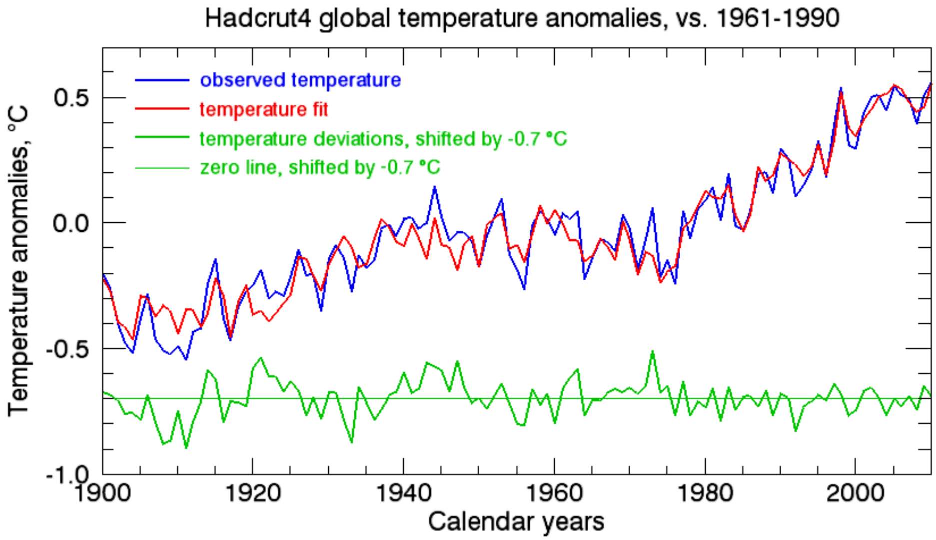

Fig. 6. The global annual temperature anomalies by the HadCRUT4 data set (blue line), the ones by the fitted using the regression equation (2) for model 1 temperatures (red line) and the differences between the observed temperatures and the fitted temperatures (green line). For better eyesight, the deviations were shifted down by 0.7 °C.

The obtained results for the fitted temperature anomalies compared to the observed ones and the computed temperature deviations for model 1 are shown in Fig. 6. (There are shown only the results for model 1, because the results for model 2 are very close to the ones of model 1). It is seen that the cooling and warming periods are very well captured by the model. The long-time trend is determined by the CO_2 term and the AMO influence in both models. These are the main regressors, describing the observed global temperature anomalies. The multiple correlation taking in consideration only these two terms is about 0.951 for model 1 and 0.953

for model 2 with an explained variance of approximately 0.905 and 0.907, respectively. In comparison to the first model in the second one the influence of the CO_2 term is somewhat smaller which is compensated by a stronger impact of the AMO. The temperature change by doubling the CO_2 content would be 1.85 °C, corresponding to a climate sensitivity of 0.5 K/Wm^{-2}, based on equilibrium change in global mean temperature. The temperature influence of TSI is small, about 0.06 °C, when TSI is changed by 1 W/m^2, corresponding to a change of the solar activity from its minimum to its maximum (or vice versa). The SOI influence is also low, but it is more important at regional scales (not shown here). It has to be mentioned, that in a model without CO_2 the high autocorrelation of the residuals leads to a significant influence only of the AMO.

By the regression equations, the adjusted temperature, describing the CO_2 influence on the temperature anomalies, can be easy calculated by removal of the temperature impacts of AMO, TSI and SOI.

Adjusted Temperature estimation

To demonstrate the relationship between the CO_2 concentration and the global temperature anomaly, here, as it was mentioned above, adjusted temperatures were calculated by removal of the influences of AMO, TSI and SOI.

(3) $$T_{adj} = \beta_1 * \ln(CO_2/280\ ppmv) + T_{obs} - T_{fit}\ ,$$

where $T_{obs} - T_{fit}$ is equivalent to the residuals ε.

The obtained adjusted temperatures T_{adj} follow very close the time evolution of the CO_2 concentration (see Fig. 7). It has to be pointed out that a climate pause caused by CO_2 after 1998 is not noticeable. If the temperatures would be not influenced by CO_2 and the temperature anomalies increase would be generated only by a linear trend the residuals would show a CO_2 induced signal. The residuals $T_{obs} - T_{fit}$ (which are equal to the differences between the adjusted temperatures and the CO_2 temperature impacts) are very close for both models. The maximal deviations of the difference of the temperature fits of both models are of the order of about 0.04 °C and they are smaller than the non-explained variance of the fits themselves but show clearly a CO_2 induced signal, like the difference of the model CO_2 impacts on the temperatures (see Fig. 7).

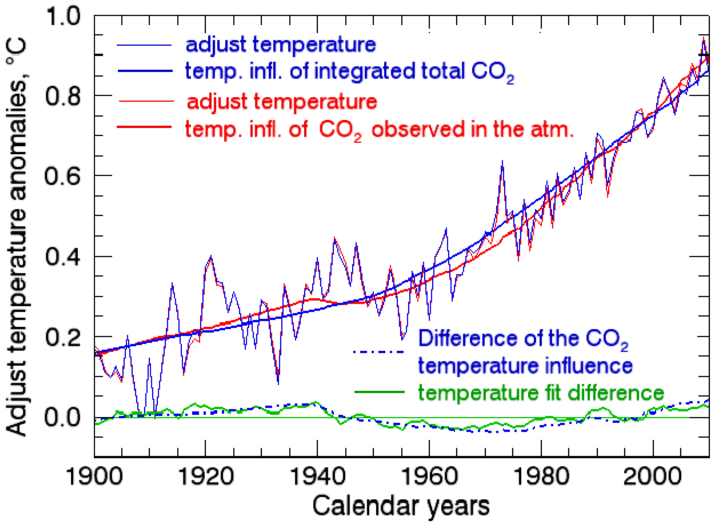

Fig. 7. The adjusted by equation (3) temperatures (thin blue and thin red line) and the CO_2 influence on temperature (thick blue and thick red line) by both regression models (see Table 1). The difference of adjusted temperature fits by the two regression models is presented by a thick green line. The dashed blue line shows the difference of the CO_2 influence on the temperature obtained by the two regression models.

For illustration of the slowdown observed in the CO_2 concentration and temperature anomalies it was assumed, that the world economy gathered the same speed of development as after 1950 immediately after 1939, so the growth rate of CO_2 after 1939 would be the same as after 1950. Then the adjusted temperature T_{adj}^* can be estimated replacing CO_2 by:

(4) $$CO_2^*(t) = CO_2(t - 10) \quad \text{for} \quad t > 1939.$$

Based on the piecewise linear regression investigations of the long-lived radiation factors Estrada et al. [38] have identified a strong breakpoint in the CO_2 radiative forcing about 1960 related to the post-war economic expansion with data sets up to 2010. The authors proposed a radiation forcing of CO_2 without the slowdown (see Supplementary information in [38]) like CO_2 forcing described here. However in [38] temperature changes due to the shifted CO_2 concentrations were not recalculated. Here the results for $T_{adj}(CO_2)$ and $T_{adj}^*(CO_2^*)$ are shown together with the original data in Fig. 8 using the regression model A.

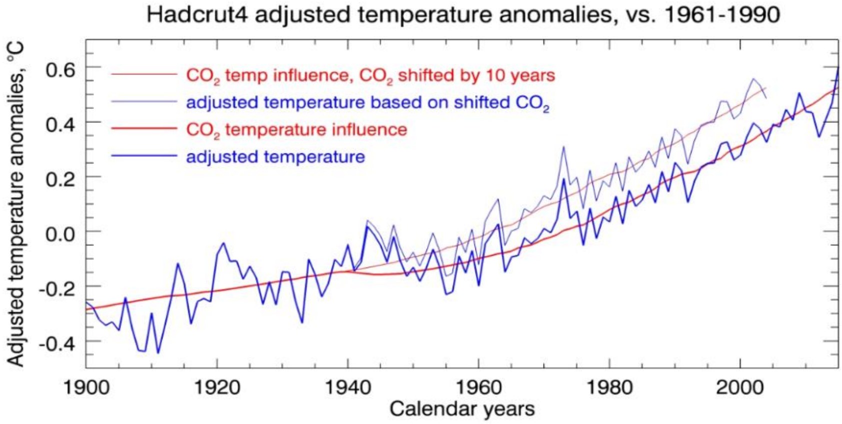

Fig. 8. Adjusted by equation (3) temperature anomalies (thick blue line) and the CO_2 temperature influence (thick red line, regression model 1) and the temperature and the CO_2 term developments assuming that the CO_2 grow rates observed after 1950 would be expected immediately after 1939 (shown by thin lines).

Mainly the additional CO_2 sink lead to a decrease of the CO_2 concentration growth rate, which caused a deceleration of the adjusted temperature anomalies. As a result of the CO_2 slowdown the temperature increase decelerates and it can be speculated that without the CO_2 slowdown the temperature today probably would be approximately 0.15 °C higher. This effect can be clearly seen only in the adjusted temperatures. For the observed temperature anomalies the effect is covered by the other temperature impacts, mainly by the AMO influence.

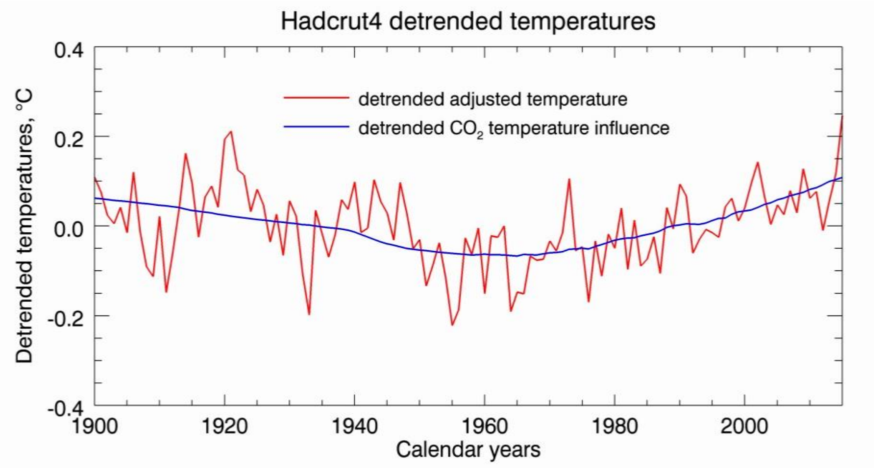

Fig. 9. Time evolution of the linear detrended adjusted temperatures and the linear detrended CO_2 temperature influences (for the regression model 1)

The relation between the adjusted temperature and the CO_2 radiation forcing at shorter time scales determined by the linear detrended series is shown in Fig. 9. It is to be noted again the close connection between the detrended adjusted temperature and the detrended CO_2 term. The close linear relation between the CO_2 temperature influence and the adjusted temperature is outlined also by the scatter plot (Fig. 10). Any deviation from a linear relation cannot be detected. Due to the at least partially man-made CO_2 slowdown and the close relation to the temperature changes at long time and shorter time scales we can conclude that the CO_2 is the leading variable of the relation between the temperature changes and CO_2 after the Pre-industrial Era.

Of course the statistical proof of causality and the proof of the human impact on climate change are complicated [42, 43]. As it is seen in Fig. 1 the residuals are not homoscedastic. The variance in the time interval up to approximately 1975 is evidently greater than after that time and the autocorrelation is also not constant over the whole time interval. The statistic investigations are more complex because the global annual temperature anomalies series have structural breaks. In the case of filtered temperature series generated by removal of the AMO temperature influence the global temperatures have a structural break approximately in 1970, as well [38, 44].

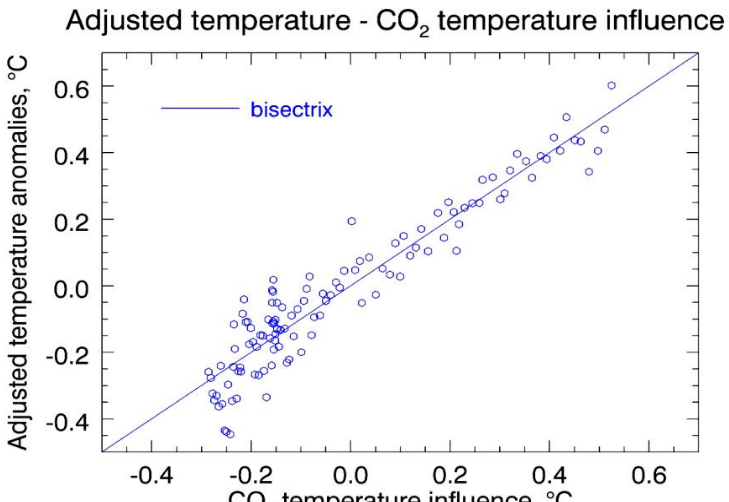

Fig. 10. Scatter plot of the CO_2 temperature influences and the adjusted temperatures

The aim of the present paper is not the statistical proof of the cause-effect relationship. Here the main goal was to demonstrate the evidence of the relation-ship between the CO_2 radiative forcing and the temperature change where the leading variable during the Industrial Era is CO_2.

Summary and conclusions

The rapid economic development after 1950 was accompanied with enormous consumption of oil up to the first oil shock in 1973. Since this time by different economic and politic reasons the growth of the emissions was decelerated. The simultaneous CO_2 ocean uptake shows high variations forced by strong ENSO events. The concentration of the atmospheric CO_2 increases continuously up to now. Due to the man-made effect of the depression of the World economic development during the World War I and after it, the Great Depression and the World War II the CO_2 emission was slowdown. The integration of the total CO_2 emission using the IRF concept shows that the observed slowdown of the CO_2 emission is not sufficient to explain the plateau of the atmospheric CO_2 concentration. To account for the CO_2 plateau one or more additional sinks are necessary, as it was obtained before by solution of the inverse problem and climate models. The AF indicates an additional strong CO_2 sink as well. Here the observed CO_2 content in the atmosphere was obtained by the integration of two different models of additional CO_2 sinks. The results show that the integration is not sensitive to the explicit form of the sinks. Furthermore based on regressions models by subtraction of temperature impacts excluding these of CO_2 adjusted temperatures were determined. It was demonstrated that the adjusted temperature and the detrended adjusted temperature follow very close the course of the radiative CO_2 forcing and the detrended course of the radiative CO_2 forcing, respectively. That means that both the adjusted temperature and the radiative CO_2 forcing have the same long time trend and the same trend at shorter time scales. Moreover a CO_2 signal was found in the adjusted global temperature anomalies and consequently in the global temperatures.

Assuming that the CO_2 grow rates would not decelerate after 1939 and the economy would have the same development as after 1950 the global temperature anomalies would be approximately 0.15 °C higher than the tempera-tures today.

In the used here time interval after the preindustrial era the atmospheric concentration was determined mainly by the energy consumption. It was shown that the CO_2 slowdown observed between 1939 and 1950 was at least partially man-made. Therefore the CO_2 is the leading variable in the relationship with the global temperature change.

The CO_2 slowdown and the deceleration of the temperature increase at the same time visible in the adjusted temperatures can be considered as a fingerprint of the human impact on the climate warming process.

References

1. Lean, J. L., D. H. Rind. How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006, Geophys. Res. Lett., 2008, 35, L18701. doi:10.1029/2008GL034864
2. Forster, G. and S. Rahmstorf. Global temperature evolution 1979–2010, Env. Res. Lett., 2011, 6, 8-xx. http://iopscience.iop.org/article/10.1088/1748-9326/6/4/ 044022/ meta
3. Schlesinger, M. E. and N. Ramankutty. An oscillation in the global climate system of period 65-70 years, Nature, 1994, 367, 723–26. doi:10.1038/367723a0
4. Enfield, D. B. and L. Cid-Serrano. Secular and multidecadal warmings in the North Atlantic and their relationships with major hurricane activity, Int. Jour. Climatol., 2010, 30, 2, 174–84. doi:10.1002/joc.1881
5. Wang, C. and S. Dong. Is the basin-wide warming in the North Atlantic Ocean related to atmospheric carbon dioxid and global warming? Geophys. Res. Lett., 2010, 37, L08707. doi:10.1029/2010GL042743
6. Knudsen, F.M., M.-S. Seidenkranz, B. H. Jacobsen et al. Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years, Nature communications, 2011. doi:10.1038/ncomms1186
7. Weng, H. Impacts of multi-scale solar activity on climate. Part II: Dominant timescales in decadal-centennial climate variability, AAS, 2012, 4, 887–908. doi:10.1007/s00376-012-1239-0
8. Knight, J. R., Folland, C. K., and Scaife, A. A. Climate impacts of the Atlantic Multidecadal Oscillation, Geophys. Res. Lett., 2006, 33, L17706. doi:10.1029/2006GL026242
9. Rohde, R., R. A. Muller, R. Jacobsen et al. A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, Geoinfor Geostat: An Overview 1:1. doi:10.4172/2327-4581.1000101
10. Zhou, J. and K. K. Tung. Deducing multidecadal anthropogenic global warming trends using multiple regression analysis, J. Atmos. Sci., 2013, 70, 1, 3–8. doi:10.1175/JAS-D-12-0208.1
11. Ballantyne A. P, C. B. Alden, J. B. Miller, P. P. Tans, and J. W. White. Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years, Nature, 2012, 488, 7409, 70-2. doi:10.1038/nature11299
12. Jones G. S., S. F. B. Tett , and P. A. Stott. Causes of atmospheric temperature change 1960–2000: A combined attribution analysis, Geophys. Res. Lett., 2003, 30, 5, 1228. doi:10.1029/2002GL016377
13. Braganza, K., D. J. Karoly, and J. M. Arblaster. Diurnal temperature range as an index of global climate change during the twentieth century, Geophys. Res. Let., 2004, 31, L13217. doi:10.1029/2004GL019998
14. Alexander, L. V., X. Zhang, T. C. Peterson et al. Global observed changes in daily climate extremes of temperature and precipitation, J. Geophys. Res., 2006, 1, 111. doi:10.1029/2005JD006290
15. Siegenthaler, U. and F. Joos. Use of a simple model for studying oceanic tracer distribution and the global carbon cycle, Tellus, 1992, 44B, 186–207.

16. Siegenthaler, U. and H. Oeschger. Biospheric CO_2 emissions during the past 200 years reconstructed by deconvolution of ice core data, Tellus, 1987, 39B, 140–54. doi:10.1111/j.1600-0889.1987.tb00278.x
17. Houghton, R. A., Balancing the global carbon budget, Ann. Rev. Earth Planet. Sci., 2007, 35, 313–47.
18. Friedrich, J. and T. Damassa. The History of Carbon Dioxide Emissions, 2014. http://www.wri.org/blog/ 2014/05/history-carbon-dioxide-emissions
19. Hansen, J., M. Sato. Greenhouse gas growth rates. PNAS, 2004, 101, 46, 16109–114. http://www.pnas.org/content/suppl/2004/11/02/0406982101.DC1/06982 SupportingText.html
20. MacFarling Meure, C., D. Etheridge, C. Trudinger et al. The Law Dome CO_2, CH_4 and N_2O Ice Core Records Extended to 2000 years BP, Geophys. Res. Lett., 2006, 14, L14810. doi:10.1029/2006GL026152
21. Langenfelds, R. L., L. P. Steele, M. V. Van der Schoot et al. Atmospheric methane, carbon dioxide, hydrogen, carbon monoxide and nitrous oxide from Cape Grim flask air samples analysed by gas chromatography. In: Baseline Atmospheric Program Australia. 2001–2002, ed. J.M. Cainey, N. Derek, and P.B. Krummel (editors), Melbourne: Bureau of Meteorology and CSIRO Atmos. Res., 2004, 46–47.
22. Etheridge, D. M., L. P. Steele, R. L. Langenfelds, R. J. Francey, J.-M. Barnola, and V. I. Morgan. Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn, J. Geophys. Res., 1996, 4115–28. doi:10.1029/95JD03410
23. Heinze, C., S. Meyer, and N. Goris. The ocean carbon sink – impacts, vulnerabilities and challenges, Earth Syst. Dynam., 2015, 6, 3R.L.27–358. doi:10.5194/esd-6-327-2015
24. AR4 Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007, 996 pp. on p. 139.
25. Keeling, C. D., T. P. Whorf, M. Wahlen, and J. v. D. Pflicht. Interannual extremes in the rate of atmospheric carbon dioxide since 1980, Nature, 1995, 375, 666–70. doi:10.1038/375666a0
26. Jones, C. D, P. M. Cox, and C. Huntingford. The atmospheric CO_2 airborne fraction and carbon cycle feedbacks, 2007. https://www.esrl.noaa.gov/gmd/co2conference/posters_pdf/jones1_poster.pdf
27. Rafelski, L. E., S. C. Piper, and R. F. Keeling. Climate effects on atmospheric carbon dioxide over the last century, Tellus B, 2009, 61, 5, 718–31. doi:10.1111/j.1600-0889.2009.00439.x
28. Joos, F., R. Meyer, M. Bruno, M. Leuenberger. The variability in the carbon sinks as reconstructed for the last 1000 years, Geophys. Res. Let., 1999, 26, 10, 1437–40. doi:10.1029/1999gl900250
29. Francey, R. J., P. P. Trans, C. E. Allison, I. G. Enting, J. W. C. White, and M. Troller. Changes in oceanic and terrestrial carbon uptake since 1982, Nature, 1995, 373, 326–30. doi:10.1038/373326a0

30. Ishii, M., H. Y. Inoue, and T. Midorikawa. Spatial variability and decadal trend of the oceanic CO_2 in the western equatorial Pacific warm/fresh water, Deep-Sea Res. Pt. II, 2009, 56, 591–606, doi:10.1016/j.dsr2.2009.01.002
31. Jones, C. D. and M. P. Cox. On the significance of atmospheric CO_2 growth rate anomalies in 2002–2003, Geophys. Res. Lett., 2005, 632, L14816. doi:10.1029/2005GL023027
32. Truderinger, C. M., I. G. Enting, P. J. Rayner, R. J. Francey. Kalman filter analysis of ice core data 2. Double deconvolution of CO_2 and $\delta^{13}C$ measurements, J. Geophys. Res., 2002, 107, D20, 4423. doi:10.1029/2001JD001112
33. Rubino, M., D. M. Etheridge, C. M. Trudinger et al. A revised 1000 year atmospheric $\delta^{13}C$-CO_2 record from Law Dome and South Pole, Antarctica, J. Geophys. Res.: Atmosphere, 2013, 118, 8482–99. doi:10.1002/jgrd.50668
34. Bastos, A., P. Ciais. J. Barichivich et al. Re-evaluatingthe1940s CO_2 plateau, Biogeosciences, 2016, 13, 4877–97. doi:10.5194/bg-13-4877-2016
35. Bastos, A., A. Peregon. É. A. Gani et al. The contribution of land-use change versus climate variability to the 1940s CO_2 plateau: Former Soviet Union as a test case, Biogeosciences Discuss., doi:10.5194/bg-2017-26
36. Joos, F., R. Roth, J. S. Fuglestvedt et al. Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis, Atmos. Chem. Phys., 2013, 13, 2793–2825. doi:10.5194/acp-13-2793-2013
37. Maier-Reimer, E. and K. Hasselmann. Transport and storage of CO2 in the ocean - an inorganic ocean-circulation carbon cycle model, Climate Dynamics, 1987, 2, 2, 63–90. doi:10.1007/BF01054491
38. Estrada, F., P. Perron, and B. Martínez-López. Statistically derived contributions of diverse human influences to twentieth-century temperature changes, nature geoscience, Advanced online publication, Supplementary information, 2013. doi:10.1038/NGEO1999
39. Werner R., D. Valev, D. Danov, V. Guineva, and A. Kirillov. Analysis of global and hemispheric temperature records and prognosis, Adv. Space Res., 2015, 55, 12), 2961–973. doi:10.1016/j.asr.2015.03.005
40. Myhre, G., E. J. Highwood, K. P. Shine, and F. Stordal. New estimates of radiative forcing due to well mixed greenhouse gases, Geophys. Res. Lett., 1998, 25, 2715–2818. doi:10.1029/98GL01908
41. Thiébaux, H. J. and F. W. Zwiers. The Interpretation and Estimation of Effective Sample Size, Journal of Appl. Meteor, 1984, 5, 800–11. doi:10.1175/1520-0450(1984)023<0800:TIAEOE>2.0.CO;2
42. Stern, D. I. and R. K. Kaufmann. Detecting a global warming signal in hemispheric temperatures series: A structural time series analysis, Climatic Change, 2000, 47, 411–38. doi:10.1023/ A:1005672231474
43. Kaufmann, R. K. and D. I. Stern. Cointegration analysis of hemispheric temperature relations, J. Geophys. Res., 2002, 107, D2, 4012. doi:10.1029/2000JD000174
44. Werner R., D. Valev, D. Danov, and V. Guineva. Study of structural break points in global and hemispheric temperature series by piecewise regression, Adv. Space Res., 2015, 56, 11, 2323-34. doi:10.1016/j.asr.2015.09.007

ГЛОБАЛНИ ТЕМПЕРАТУРНИ АНОМАЛИИ, СВЪРЗАНИ СЪС СПАД НА КОНЦЕНТРАЦИЯТА НА CO_2 В АТМОСФЕРАТА, ОТ 1939 Г. ДО 1949 Г.

Р. Вернер, В. Гинева

Резюме

Известно е, че въглеродният диоксид (CO_2), акумулиран в атмосферата, е главният драйвер на климата. Първите точни непрекъснати преки измервания на концентрацията на атмосферния CO_2 са проведени от Keeling от 1958 г. в обсерваторията Mauna Loa. Измерванията продължават до днес. Сондирането на ледения слой в Low Dom започва през 1969 г. с програмата ANAPE.

Интегрирането на тоталната емисия на CO_2 с помощта на импулсна апаратна функция показва, че наблюдаваният спад на емисията на CO_2 не е достатъчен, за да обясни полученото плато и е необходимо да се включат допълнителни потоци. На основата на мулти-регресионни модели бяха определени изчистените „напаснати" глобални температури чрез изключване на други температурни влияния, освен свързаните със CO_2.

Изчистените „напаснати" температури на хода на радиацията на CO_2, разликата между времевото развитие на оценените уточнени температури със спада на CO_2 и без него, а също и кратковременните трендове показват много ясно тясната връзка между температурната промяна и парниковия ефект от CO_2. Показано е, че спадът на емисията на CO_2 в периода от 1939 г. до 1950 г. и свързаната с нея концентрация на CO_2 в атмосферата, породена поне отчасти от човешката дейност, предизвиква по-бавно нарастване на температурните аномалии през този период. Следователно CO_2 е водещата променлива в зависимостта: температура на повърхността – CO_2.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e05

COSMIC RAY AND SOLAR ACTIVITY INFLUENCES ON LONG-TERM VARIATIONS OF CAVE CLIMATE SYSTEMS

Alexey Stoev, Penka Stoeva

Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: stoev52@abv.bg, penm@abv.bg

Keywords: *Cosmic ray, Solar activity, Cave climate, Temperature changes*

Abstract

During the analysis of solar activity impact on climate, the emphasis is placed on temperature changes. Earth's atmosphere is a dynamical system with a complex variability in space and time. Due to the fact that caves in Karst preserve the long term environmental changes, the investigation of the in-caves' atmospheric parameters and their variations with time becomes very important in the last quarter of century.

In this paper we investigate the temporal evolution of the temperature and pressure of the ground atmospheric layer in the region of two Bulgarian caves: Snezhanka (Pazardjik region) and Uhlovitsa (Smolyan region), during the period 2005–2017. We show that thermal and mass exchange of the caves' air with the environment has significant temporal variations. On annual basis the thermodynamical parameters of the observed caves behaves as a barotropic fluid, in which the air density depends only on atmospheric pressure. As a result, the temporal evolution of in-caves' pressure and temperature change synchronously with time. The observed 11-year signal could be attributed to the heliospheric modulation of galactic cosmic ray (GCR) intensity, which modulates the ozone and humidity near the tropopause and correspondingly the strength of the atmospheric greenhouse effect. Our study helps to clarify the influence of helio-geophysical factors on the state of the lower atmosphere.

Introduction

It has been known for a long time that solar activity and disturbances it causes in the interplanetary environment affect different processes in all layers of the Earth's atmosphere. Among the other external factors influencing Earth's climate are galactic cosmic rays (GCRs) and space dust, the latter of which controls the total energy transferred from the Sun to the Earth's atmosphere. With no doubt, the comfortable conditions supporting the existence of life on the planet, are obviously ensured by the amount of the total solar irradiance (TSI) reaching the planetary surface. The instrumental measurements, however, do not support the idea for a significant deviation of TSI from its recently observed values, while paleoclimatic records give evidences for dramatic changes of Earth's climate during the planetary evolution. Moreover, the reconstruction of TSI from 7000 BC up to 500 AD shows that the amplitude of its variations does not exceeds 1% [1].

Much bigger changes are observed in solar UV radiation (in the interval 175÷200 nm (Schumann-Runge bands) ~6%. This motivates some scientists [2] to suggest that the heating of the stratosphere by UV radiation can be dynamically transported down to the troposphere. Meanwhile, it has been revealed the increased climate sensitivity to the chemical and dynamical conditions in the upper troposphere – lower stratosphere region [3, 4]. The external forcing, which easily reach these altitudes are GCRs, what gives a hint that they could somehow influence the surface climate. One of the proposed mechanisms includes GCR influence on the clouds' cover, which could modulate the amount of solar radiation reaching the planetary surface [5]. The more generalized relationship is given by Lev Dorman [6]:

(solar activity cycles + long-term changes in the geomagnetic field) → → (CR long term modulation in the Heliosphere + long term variation of cutoff rigidity) → → (long term variation of clouds covering and aerosols + atmospheric electric field effects) → climate change.

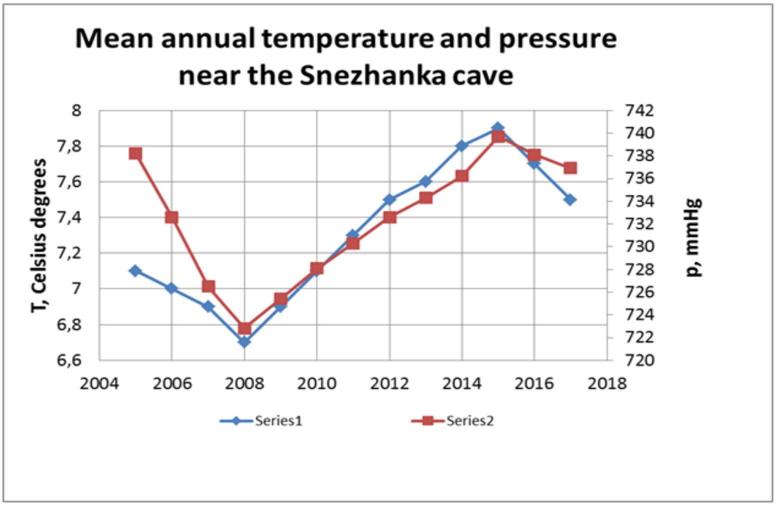

*Fig. 1. Mean annual temperature **T** (series 1) and barometric pressure **p** (series 2) near the Snezhanka cave (Pazardjik region, Bulgaria) in the period 2005–2017*

However, the CERN laboratory experiment reveals that the impact of cosmic radiation in cloud's formation is negligible [7]. Meanwhile, it has been proposed a new hypothesis [8] according to which the GCR influence on the near tropopause ozone drives corresponding variation of the water vapour in the upper troposphere. The impact of the latter ensures 90% of the greenhouses power of the whole water content in the atmosphere [9], modulating in such a way the climate conditions near the planetary surface.

Experimental modelling of complex monitoring for sustainable development of protected Karst territories

The aim of the work is to design experiment and create a model for complex monitoring of protected Karst territories as a basis for specialized research, management and sustainable development and education. The Karst territories were identified and measurements of the different parameters required for the complex monitoring model were made.

Studying microclimate in caves and influence of solar and geomagnetic activity on processes there, we have investigated relations between cave air temperature and: sunspot number, Ap max index, surface temperature and pressure near about the caves and found very good correlation [10].

In this work we want to investigate the influence of CR on cave microclimate. We compare average annual changes in pressure and temperature of the ground atmosphere in the region of caves Uhlovitsa (Smolyan) and Snezhanka (Pazardzhik), Bulgaria in the period 2005–2015 and Cosmic Ray (CR) variations from the neutron monitor in Athens, Greece (geographic latitude 37.97° N, geographic longitude 23.78° E, altitude 260 m a.s.l.).

The studies of the microclimate in caves give additional perspective on the variations of climatic parameters with time. This paper presents information for variations of near surface temperature and pressure measured near two Bulgarian caves, but close to their entrances, where the exchange with the outside environment is still quite good. The examination of the records reveals the existence of quasi-decadal variations in the temperature and pressure during the observed period 2005–2017. The results and possible explanation of these variations are presented in the following sections of the work.

Data and results

The temperature and pressure has been measured by automatic meteorological stations in the region of two Bulgarian caves Snezhanka (Pazardzhik region) and Uhlovitsa (Smolyan region) – close to the cave entrances, during the period 2005–2017. The results are presented on Fig. 1 and Fig. 2.

Data of Galactic Cosmic Ray variability has been taken from NMDB: Real-Time Database for high-resolution Neutron Monitor measurements (http://www01.nmdb.eu/). We use such a station which is located closest to South Bulgaria where the studied caves are. This is the neutron monitor in Athens, Greece (geographic latitude 37.97° N, geographic longitude 23.78° E, altitude

260 m a.s.l.). The instrument is Standard 6-NM64 neutron monitor with geomagnetic threshold – effective vertical cut-off rigidity (Epoch 2000.0) 8.53 GV (see http://www01.nmdb. eu/station/athn/).

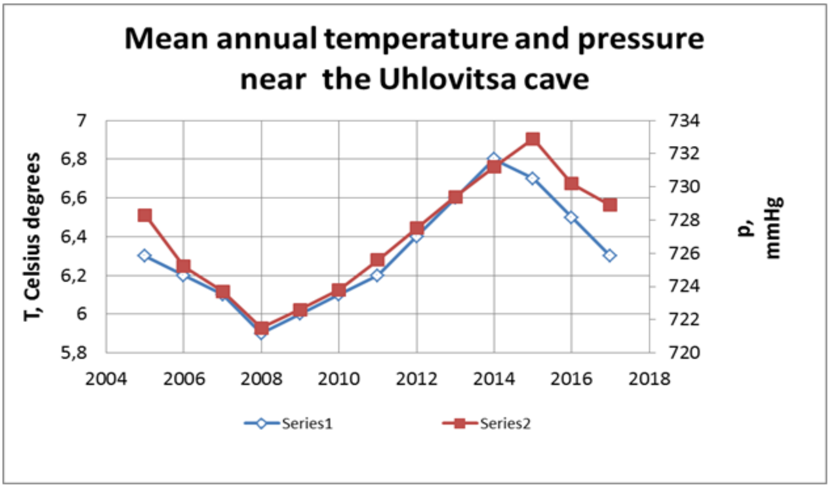

*Fig. 2. Mean annual temperature **T** (series1) and barometric pressure **p** (series 2) near the Uhlovitsa cave (Smolyan region, Bulgaria) in the period 2005–2017*

Sunspot numbers and planetary Ap index are taken from the SILSO (Sunspot Index and Long-term Solar Observations, URL: http://www.sidc.be/silso/datafiles) of World Data Center for the production, preservation and dissemination of the international sunspot number, situated in Royal Observatory of Belgium, Brussels.

On Fig. 3a are shown the courses of GCRs (Athens Neutron Monitor) and the monthly smoothed sunspot numbers (SSN). On Fig. 3b are presented the monthly smoothed Ap index and SSN during the period 2001–2018.

Analysis and interpretation

We have compared the mean annual values of air temperature and pressure in the region of two Bulgarian caves – Snezhanka (Pazardzhik region) and Uhlovitsa (Smolyan region) (Fig. 1 and Fig. 2). It is easily noticed that the temperature and pressure have a well pronounced minimum during the period of low solar activity, followed by a corresponding peak near the maximum of sunspot's numbers. Due to the modulation of GCR intensity by the 11-year solar cycle, in period of minimum solar activity more GCR are allowed to penetrate the heliosphere and correspondingly their intensity, measured in the Earth's environment, is maximal (see Fig. 3). Oppositely, when the Sun is more active, the higher density of solar wind and the stronger magnetic field carried by them act as a barrier for less energetic GCR – disrupting their propagation in the heliosphere.

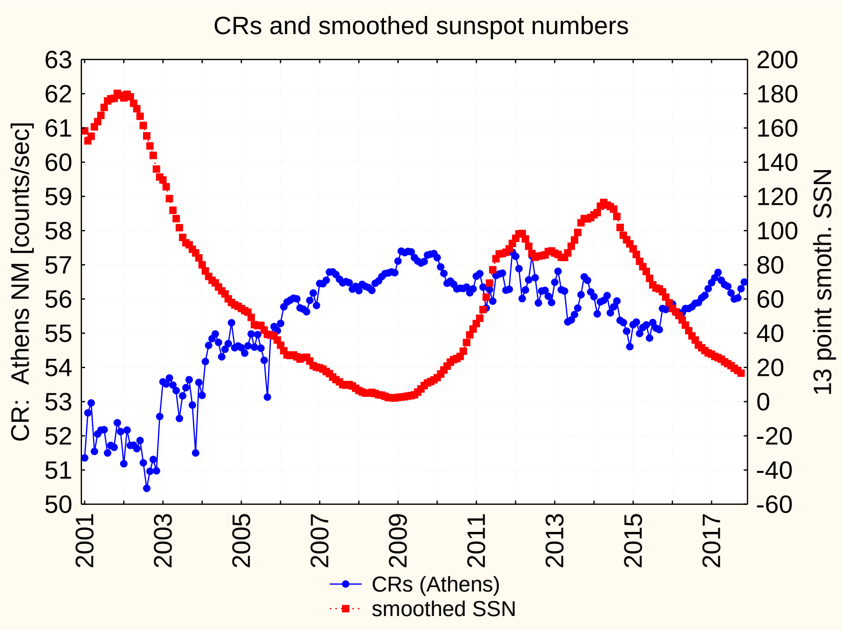

Fig. 3a. Variations of smoothed sunspot numbers (SSN) (http://www.sidc.be/silso/datafiles) and neutron component of cosmic rays during the period 2001–2018

On Fig. 3a is clearly shown the correlation between solar and geomagnetic activity, while Fig. 3a demonstrates the anti-correlation between solar activity and the intensity of galactic cosmic rays. The time lag or the relaxation of natural processes relative to solar activity is also noticeably seen on Fig. 3.

The existence of solar signal in climate parameters is reported from many authors and several hypotheses are attempting to explain this quasi-periodicity. An intriguing moment in our data is the covariance between surface air temperature and pressure. This means that the caves are filtering the baroclinic disturbances, which are typical for the real troposphere, especially at mid-latitudes. More specifically, in baroclinic atmosphere the relation between pressure and temperature in not linear, due to the fact that besides from pressure the atmospheric density depends also on the temperature of air masses. Temperature and pressure covariates with time (see Fig. 1 and Fig. 2). This suggests that ground atmospheric layer near the cave entrances behaves as a barotropic fluid, in which air density depends only on the atmospheric pressure.

The next question, which should be elucidated, is the decadal variability of the temperature and pressure with time. As mentioned in the introduction, the amplitude of the quasi-decadal variations of TSI is less than 1% [1], which suggests either existence on amplifying mechanism(s) of solar electromagnetic radiation reaching the planetary surface, or impact from other external influence. There are a lot of studies reporting for an existence of a relation between GCR and climatic and ionization state of the environment [11, 12].

A plausible hypothesis for GCR influence on the Earth's climatic conditions is presented in [8, 13] and an explanation of our results with it is described in the following section.

Explanation of results with mechanism of Kilifarska

The relation between energetic particles and middle atmospheric ozone has been reported in several studies [14, 15]. However, Kilifarska [8, 13] has revealed for the first time the important role of the secondary electrons, produced by GCRs in the Regener-Pfotzer maximum [16], for activation ofion-molecular reactions and ozone production just above the tropopause [17]. Moreover, the GCR effect is not homogeneously distributed over the globe, due to the spatially heterogeneous geomagnetic field [18]. This specificity of GCR forcing is able to explain the regional character of climate variability, which couldn't be said for the other mechanisms.

According to [8, 13] the O_3 variations above the tropopause affect its temperature and consequently – the moist adiabatic lapse rate. For example, during the minimum of solar activity – the increased level of GCR should produce more ozone in the lower stratosphere, which warms the tropopause. This would stabilize the upper troposphere, due to the increased wet adiabatic lapse rate. However, the vertical motions in a stably stratified atmosphere are strongly suppressed [19]. Consequently, after some time the upper troposphere will become drier and its greenhouse power will be reduced. Taking into account that 90% of the greenhouse effect of the whole water vapour in the atmosphere is due to the upper tropospheric water vapour [9] – one might expect a surface cooling in periods of high CR intensity (respectively low solar activity).

Coming back to our Fig. 1 and Fig. 2, we could attribute the depression in near surface temperature and pressure in both caves to the increased ozone density in the lower stratosphere and consequently – to the increased static stability and reduced humidity in the upper troposphere. This, in turn, diminishes the greenhouse effect of the water vapour cooling the Earth's surface. Oppositely, the temperature-pressure peaks, during the maximum of 24th solar cycle, could be attributed to the depleted O_3 density and increased water vapour near the tropopause, which warms the surface through the strengthened greenhouse effect.

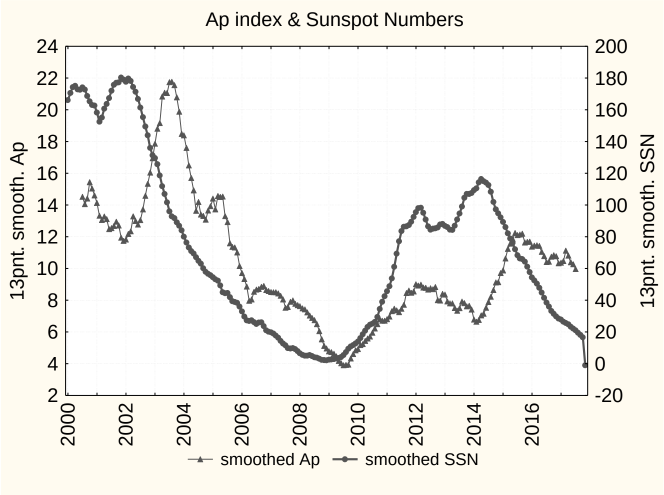

Fig. 3b. Variations of smoothed sunspot numbers (SSN) (http://www.sidc.be/silso/datafiles) and planetary geomagnetic Ap index during the period 2001–2018

Discussion and conclusion

In the last quarter of the century, the problem of the impact of the low atmosphere with its dynamic baric and temperature fields on the microclimate of caves in Karst was particularly interesting. An important aspect in the study of the change of the thermodynamic state of the cave atmosphere is the search for processes directly or indirectly affecting the air temperature in the zone of constant temperatures of the caves. The paper shows that models of ground atmosphere circulation and thermo and mass exchange with the air volume in the caves have energetically significant spatio-temporal variations. Their main thermodynamic parameters are mainly related to the dynamic efficiency of atmospheric movements in the low atmosphere. In turn, they are strongly influenced by the rapid change of Earth's cosmic rays and the evolution of the actual solar cycle.

Subject of this work is the detected relationship between annual changes in the temperature and pressure of the ground atmospheric layer near the entrances of Snezhanka (Pazardjik caves) and Uhlovitsa (Smolyan region) caves, Bulgaria in the period 2005–2017.The preliminary data analysis reveals the existence of an 11-year solar signal. This temperature variability could be attributed to the influence of GCR on the secondary ionisation in the Regener-Pfotzer maximum [15] and furthermore on the lower stratospheric ozone. The latter, in turn, affects the humidity beneath the tropopause and consequently – the strength of the greenhouse effect, because more than 60% of it belongs to the water vapour [20], and particularly to its amount in the upper troposphere [9]. The ozone – humidity variations and their imprint on the surface temperature seems to be a key factor for understanding the temporal and spatial specificity of recently observed climate changes [13, 21, 22].This is like that because the changes in solar luminosity are small and slowly [23] to be a factor in the observed climate variations.

Acknowledgements

We acknowledge the NMDB database (www.nmdb.eu) founded under the European Union's FP7 programme (Contract no. 213007), and the PIs of individual neutron monitors for providing data.

References

1. Shapiro, A. I., W. Schmutz, E. Rozanov, M. Schoell, and M. Haberreiter. A new approach to the long-term reconstruction of the solar irradiance leads to large historical solar forcing. Astronomy&Astrophysics, 2011, 529, A67, 1–8, doi:10.1051/0004-6361/201016173.
2. Kilifarska, N. A. and J. D. Haigh. The impact of solar variability on the middle atmosphere in present-day and pre-industrial atmospheres, J. Atmos. Solar-Terr. Phys, 2005, 67, 3, 241–49.
3. Kilifarska, N. A. Solar variability and climate–UTLS amplification of solar signal, Sun and Geosphere, 2006, 1, 1, 51–55.
4. Damiani-Badache, C., J. P. Rozelot, K. Coughlin, and N. Kilifarska. Influence of the UTLS region on the astrolabes solar signal measurements, Mon. Not. Royal Astron. Soc., 2007, 380, 2, 609–14, doi:10.1111/j.1365-2966.2007.12079.x.
5. Svensmark, H. Cosmic rays and Earth’s climate. Space Sci. Rev., 2000, 93, 175–85.
6. Dorman, L. CR and other space climate factors influenced on the Earth's climate change. Electronic Scientific Edition Almanac Space and Time Special issue 'The Earth Planet System', Earth in Space, COST 803 program, 2012, 1, 1, UDK 524.1: 58.057:57.045:551.581.1.
7. Dunne, E. M., et al. Global atmospheric particle formation from CERN CLOUD measurements, Science, 2016, 354, 6316, 1119–24, doi:10.1126/science.aaf2649.
8. Kilifarska, N. A. Bi-decadal solar influence on climate, mediated by near tropopause ozone, J. Atmos. Solar Terr. Phys, 2015, 136, 216–30.
9. Inamdar, A. K., Ramanathan V., and Loeb N. G. Satellite observations of the water vapour greenhouse effect and column longwave cooling rates: Relative roles of the continuum and vibration-rotation to pure rotation bands, J. Geophys. Res., 2004, 10, D06104, doi:10.1029/2003JD003980.
10. Stoeva, P. and Stoev A. Cave air temperature response to climate and solar and geomagnetic activity. Memorie della Società Astron. Ital., 2005, 76, N4, 1042–47.
11. Eroshenko, E., P. I. Y. Velinov, A. Belov, V. Yanke, E. Pletnikov et al. Relationships between neutron fluxes and rain flows, Adv. Space Res., 2010, 46, 637–41.
12. Velinov, P. I. Y., N. A. Smirnova, and V. A. Vlascov. Hybrid quadri-ionic model of the lower ionosphere, Adv. Space Res., 1984, 4, 1, 123–30.

13. Kilifarska, N. A. Hemispherical asymmetry of the lower stratospheric O3 response to galactic cosmic rays forcing, ACS Earth and Space Chemistry, 2017, 1, 80–88, DOI: 10.1021/acsearthspacechem.6b00009.
14. Tassev, Y., Kilifarska, N., and Tomova, D. Statistical analysis of solar proton flux influence on thermodynamics of middle atmosphere in the north hemisphere, C. R. Acad. Bulg. Sci., 2014, 67, 1, 95–100.
15. Kilifarska, N. A. Long–term variations in the stratospheric winter time ozone variability – 22 year cycle, C. R. Acad. Bulg. Sci., 2011, 64, 6, 867–74.
16. Velinov, P. I. Y. and L. Mateev. Effects of galactic cosmic rays and high energy particles on the parameters of the global atmospheric electrical circuit, Geomagnetism and Aeronomy, 1990, 30, 4, 554–57.
17. Kilifarska, N.A. An autocatalytic cycle for ozone production in the lower stratosphere initiated by galactic cosmic rays, C. R. Acad. Bulg. Sci., 2013, 66, 2, 243–52.
18. Velinov, P. I. Y, L. Mateev, and N. Kilifarska. 3-D model for cosmic ray planetary ionisation in the middle atmosphere, Ann. Geophys., 2005, 23, 9, 3043–46.
19. Young, J. A. Static stability, in: J. R. Holton, J. A. Curry, J. A. Pyle (Editors), Encyclopaedia of Atmospheric Sciences. Academic Press, 2003, 5, 2114–20.
20. Schmidt, G. A., R. A. Ruedy, R. L. Miller, A. A. Lacis. Attribution of the present-day total greenhouse effect, Journal of Geophys. Res., 2010, 115, D20106, DOI:10.1029/2010JD014287.
21. Kilifarska, N., V. Bakhmutov, and G. Melny. Energetic particles influence on the southern hemisphere ozone variability, C. R. Acad. Bulg. Sci., 2013, 66, 11, 1613–22.
22. Kilifarska, N. and Tassev Y. Ozone profile response to the series of coronal mass ejections and severe geomagnetic storm in September 2017, C. R. Acad. Bulg. Sci., 2018, 71, 5, 663–71.
23. Nikolov, T. and N. Petrov, Main factors influencing climate change: A review, C. R. Acad. Bulg. Sci., 2014, 67, 11, 1455–76.

ВЛИЯНИЕ НА КОСМИЧЕСКИТЕ ЛЪЧИ И СЛЪНЧЕВАТА АКТИВНОСТ ВЪРХУ ДЪЛГОСРОЧНИТЕ ИЗМЕНЕНИЯ В КЛИМАТА НА ПЕЩЕРНИТЕ СИСТЕМИ

А. Стоев, П. Стоева

Резюме

При анализа на влиянието на слънчевата активност върху климата акцентът се поставя върху температурните промени. Земната атмосфера е динамична система със сложни изменения в пространството и времето. Поради факта, че пещерите в Карст запазват дългосрочните промени в околната

среда, изследването на атмосферните параметри в пещерите и техните вариации във времето става много важно през последната четвърт на века.

В тази статия изследваме еволюцията на температурата и налягането на приземния атмосферен слой във времето, измерени близо до входовете на две български карстови пещери: Снежанка (област Пазарджик) и Ухловица (област Смолян), през периода 2005–2017. Ние показваме, че топлинният и масовият обмен на въздуха на пещерите с околната среда има значителни изменения във времето. На годишна база термодинамичните параметри на наблюдаваните пещери са като на баротропичен флуид, при който плътността на въздуха зависи само от атмосферното налягане. В резултат на това, налягането и температурата в пещерите се променят синхронно. Наблюдаваният 11-годишен цикъл може да се дължи на хелиосферната модулация на интензивността на галактическите космически лъчи (GCR), които, от своя страна, модулират озона и влажността в близост до тропопаузата и съответно – силата на парниковия ефект. Нашето изследване помага да се изясни влиянието на хелио-геофизичните фактори върху състоянието на ниската атмосфера.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e06

INVESTIGATION OF THE SIMILARITY ALGORITHM OF THE SATELLITE IMAGES STORAGE SYSTEM FOR STABILITY ON THE BASIS OF HAAR WAVELETS ACCORDING TO TIKHONOV

Maretta Kazaryan[1], Mikhail Shahramanian[2], Svetoslav Zabunov[3]

[1]North Ossetian State Medical University, marettak@bk.ru
[2]Research Institute "AEROKOSMOS", 7283963@mail.ru
[3]Space Research and Technology Institute — Bulgarian Academy of Sciences, e-mail: svetoslavzabunov@gmail.com

***Key words**: Wavelet-Haar transformations, The Tikhonov regularization method, Waste disposal facilities (WDF), Waste disposal site (WDS).*

Abstract

A model of the automated space monitoring system (ASMS) for the presence of waste disposal facilities (WDFs) is being developed in this paper. One of the components of the model is the space imagery storage unit, which allows not only to improve the performance of environmental authorities, but also to enhance the effectiveness of WGF monitoring (space, aerological, and ground based). The article also investigates the problem of searching in space image archives of similarities based on Haar wavelet transformations for stability.

This paper aim at modelling of an ASMS and the application of the regularization method in the problem of finding similar space images in the archives of the simulated ASMS by means of discrete orthogonal transformations, in particular, wavelet transformations of Haar. We use the Tikhonov regularization method, elements of mathematical analysis, the theory of discrete orthogonal transformations, and methods for decoding cosmic images.

The result of the experiment, which confirms the Tikhonov regularization method for Haar wavelet transforms, based on the example of processing archive satellite images that are located in a data warehouse, is presented.

Introduction

In relation to the emergence of a variety of information technologies with their numerous capabilities, wide-range monitoring has been provided for a variety of phenomena and, accordingly, objects [1–9]. It is, first of all, the satellite systems of Earth observation and large flows of information entering the scientific laboratories. This led to the creation of new approaches and methods of organizing the work with information on remote sensing of the Earth (RSE), as well as the technologies for remote monitoring systems construction.

The major factors in the field of remote sensing that affect the development of monitoring systems are:

- Growth of the number of spacecraft, resulting in increased frequency and volume of information that can be utilized by the systems. It is now possible to organize monitoring of rapidly occurring processes.
- A large number of satellite systems are now equipped with modern technology that allows the research of quantitative information calibration of the phenomena and objects on Earth.
- The information is quite accessible, and this leads to its extensive use.
- The availability of remote sensing information has identified the need to increase the level of automation of satellite information processing and work with extremely large data archives.
- Data processing systems need to be optimized, forming basic information products and performing standard primary processing. The implementation of these procedures requires considerable computing resources.
- Effective work with extremely large archives and, accordingly, computing resources leads to the creation of new technological capabilities.
- To perform all the listed tasks, one must use various distributed computing resources.

In this paper, studies are carried out in the field of waste disposal facilities (WDFs), which will be considered while creating a specific remote monitoring system.

WDFs are a source of risk, associated with air and soil pollution, and pose a real threat to the population and the environment [10–12].

"The big business gurus bequeathed to their descendants: the most profitable investment is an investment in a person's natural needs. Among other things, a person in the process of his life leaves behind mountains of debris. Only the population of our country annually "produces" 35–40 million tons of solid domestic waste"[1].

It is almost impossible to control these objects, scattered on a giant territory, by means of traditional land-based methods, and the annual use of aerial observations becomes more and more expensive and can be carried out only in limited local areas. Under these conditions, there is no alternative to space

[1]http://www.rbc.ru/economics/23/08/2012/5703fbc99a7947ac81a6b045

monitoring as the most expeditious and cheap method for detecting and controlling the development of unauthorized places for the WDFs.

The current article initiates a project on the creation of an automated space monitoring system for the presence of WDFs in the territory of the Russian Federation designed to provide an operative survey of the study area, the detection and mapping of real and potential centres of unauthorized landfills, the forecast of their development, a preliminary assessment of the scale of disasters and possible consequences. It should be emphasized that such systems belong to the class of open systems. They cannot be built in a complete form, they are in constant development.

The initial project envisages the creation of a starter complex (core) of the system, including the space and ground segments. An effective monitoring system should comprise a space segment: spacecraft on circular solar synchronous orbits equipped with active and passive sounding instruments of different spatial resolution in the widest possible range of the electromagnetic spectrum. First of all, it is advisable to use such public space systems such as NOAA (resolution 1100 m) and EOS AM TERRA (resolution 250 m) for regular survey of state territory. Of the high-resolution systems, the meteorological data "Meteor-3M" (37 m resolution) and Indian satellites IRS 1C/1D (PAN resolution 5.6 m) are the most accessible. It is these spacecraft that form the backbone of the space segment of the system.

Since the research area is often restricted by a strong cloud cover, which prevents regular space surveys in the microwave range, it is necessary to use active sounding data, in particular, the Canadian satellite RadarSat-1, and Sentinel-1 (Copernicus). This satellite is equipped with radar, which allows obtaining high-quality images of the Earth's surface, regardless of the presence of cloud cover and time of day with a resolution of 8 to 100 m. In combination with space imagery in the visible and thermal ranges, this information is extremely useful for detecting and monitoring the development of unauthorized waste disposal facilities [9]. The basis of the ground segment infrastructure of the ASMs WDF is the remote sensing data acquisition system (RSAS). The most reasonable solution is to create duplicate universal receiving stations and a distributed network for processing space images. Creation of integrated automated WDF monitoring systems that unite informational modelling and control systems, software complexes and technical means of data collection and transmission on the basis of local computer networks into a single structure, and the development of perspective models and algorithms for forecasting foci, unauthorized WDF is an actual problem.

The functioning of space monitoring of waste disposal facilities (WDFs) can be represented using the following scheme (Fig. 1).

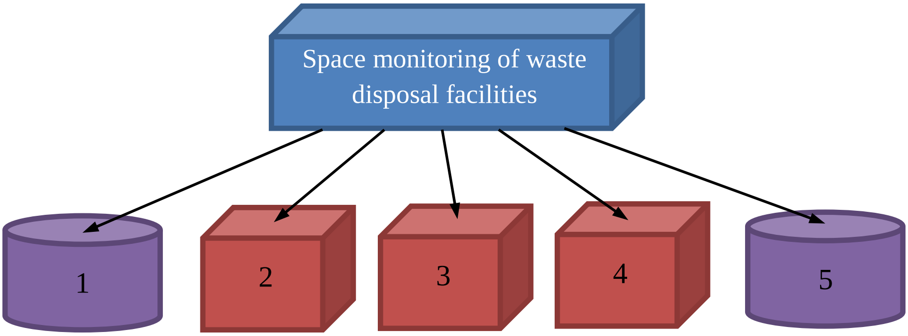

Fig. 1. Block diagram of WDF space monitoring

In Fig. 1 the blocks are:

1. Database of existing polygons of WDF, which are listed in the state register of solid domestic waste (SDW)[2] for the current time;
2. A subsystem for detecting unauthorized WDF[3] at the current time;
3. A subsystem for monitoring the rules for the design, operation and reclamation (DOR) of existing solid waste landfills [1, 2, 10];
4. A subsystem for estimating the parameters of the WDF and their environmental impact [2, 11, 12];
5. Subsystem of satellite monitoring.

The WDF space monitoring system functions as follows. In block II, space images are processed by means of automation of cosmic image processing in conjunction with detection of WDF[4]. Comparing the detection area of the WDF with the data of the database of solid waste landfills in block I, unauthorized WDFs are detected, and fixed in block II. In block IV, the WDF parameters are evaluated for both the WDF presented in Block I and the unauthorized WDFs presented in Block II. Using the data from Block V obtained through satellite monitoring methods, the data of Block I and IV is collected in Block III. There, the incoming information is further analysed, and violations of operating rules are identified for existing landfills and the development of appropriate environmental measures.

The proposed monitoring system for landfills and WDFs (Fig. 1), developed on the principle of extraterritoriality, allows the control the garbage storage sites to be taken to a new information level.

[2]http://www.airsoft-bit.ru/stati-po-ekologii/394-groro

[3]http:// www.ecobiocentre.ru/naturalist/google-svalka.doc

[4]https://cyberleninka.ru/article/n/metodika-avtomaticheskogo-detektirovaniya-komponent-obektov-zahoroneniya-othodov-po-kosmicheskim-izobrazheniyam

In addition to the above described system, it is possible to consider the creation of a system for supporting long-term data archives[5] along with and its information input. These systems are actively created in recent years to address both fundamental and applied problems. The main task of such systems is to collect and organize long-term data storage and to conveniently represent this data (import this data) into the information systems of various research projects.

We will consider the use of ready-made software for storage, output and input of monitoring data as an alternative to creating software for such systems. The most important thing of the inner workings of such systems is the electronic Earth's map of the site on Earth being under research. The electronic map has a primary interface, it identifies the objects that are of interest to us, in this case, these are solid waste objects, authorized and unauthorized, and sampling points, through which soil survey is carried out, etc. The operation of the system will look like this: some kind of GIS-system with open source. Then the map and data are loaded into it.

The drawback of this approach is that GIS systems are either not adapted to Russian or this is done poorly. This disadvantage introduces some discomfort in the process of work, so it is advisable for one to conduct their own development of ASMS SWF. When creating such an information system, you must have the following components at hand: database management systems, a web server, a programming language development environment and, of course, a web browser.

The question of obtaining an electronic map can be solved in two ways. The first one is to use Yandex Maps or Google Maps services. You can optionally download the map fragment of interest in a raster format and convert it to a vector format, so you can create an interactive map, for example, using the *Inkscape* application. This method has certain specificity – in spite of a fairly simple manufacturing method, such a map does not provide information on the composition of the soil of the investigated section of the earth.

The second method is the use of specialized maps on paper. Complete information is provided, including soil data, etc. One could scan such maps and convert them to vector format. However, this approach seems rather cumbersome due to the fact that paper maps are not always in good condition and special computer programs for their improvement have to be applied.

Thus, summarizing the above reasoning, we may conclude that for storing and processing of the earth monitoring data, it is possible to adapt applications to the tasks, leading to a significant reduction in time.

---

[5]http://www.ntsomz.ru/ks_dzz/nkpoi/catalog_service

The next block on the information stream is the block for image processing, i.e. space images[6, 7]. With the widespread implementation of digital communication systems, the urgency of recovery problems solving, reduction of additive noise and reconstruction of multidimensional signals, obtained with photo and video cameras and transmitted through communication channels, is increasing[8]. In practice, the images are often distorted by noise, which happens during the formation and transmission of signals through the communication channel. When digital images are obtained, the noise source can be a CCD detector-spectrometer, as well as fluctuation processes in photo sensors[9]. The results of image restoration are widely used in automatic signal processing systems, in digital photo and video recording systems and machine vision. The use of multimedia and television digital systems increases the urgency of solving the image reconstruction problems in the reconstruction of static and dynamic two-dimensional signals. Reconstruction of images is an important area of application of modern digital information processing systems in obtaining a reliable estimate for visual and especially for automatic analysis. In most cases, when solving the reconstruction task, it is required to estimate the missing pixel values of images and video sequences, and also to select the most "similar" section from a large number of already existing ones. The solution of this task involves retouching and restoring of the missing fragments of images when removing scratches, defects, unnecessary inscriptions, etc.

Currently, the technical implementation of digital systems is intensively developing, their speed and energy efficiency are increasing. In this case, methods and algorithms for signal and image processing are used.

Formulation of the problem

As a practical example of the above discussion, let us consider the following problem, which has an applied character: the search and storage of space images in archives. In describing the digital images below, we will mean space images (SIs).

The task is to create a similarity search system (SSS) and to search for "similarity" in the archives of images storing WDFs based on Haar wavelet transforms and investigating this problem for stability [21–23].

[6]http://window.edu.ru/resource/028/76028/files/PosobieERS.pdf

[7]http://files.lib.sfu-kras.ru/ebibl/umkd/54/u_course.pdf

[8]http://www.dissercat.com/content/korrektsiya-tsifrovykh-kosmicheskikh-izobrazhenii-na-osnove-verifitsiruyushchego-modelirovan

[9]http://www.dissercat.com/content/rekonstruktsiya-smazannykh-i-zashumlennykh-izobrazhenii-metodami-regulyarizatsii-i-usecheniy

Basic concepts

An important step in constructing a mathematical model of the concept of "similarity" for storing space images is the measurement of the appearance of the image boundaries using the Haar J-step discrete wavelet transformation (DWT) [13–15].

As it is well known, one of the problems in image processing is to find a way to efficiently represent an image or SI in a compact form. In modern practice of spectral analysis, recently acquired signals of a special kind, namely, wavelets have become quite popular. Since we are studying the SIs, we are primarily interested in two-dimensional discrete wavelet transforms. For being ascertained [16], the two-dimensional wavelet transforms are based on one-dimensional wavelet transforms that do not depend on the number of rows and columns of the image. By virtue of this rule, we consider, basically, the horizontal and vertical directions of the wavelets. Specifically, Haar's wavelets [17, 18].

1. These are piecewise constant functions defined on intervals having different scales and, furthermore, they take two values {–1; +1}.

2. The Haar's maternal wavelet, which has a unit scale and zero offset, is a function that takes the value +1 on the interval [0; 1/2) and –1 on the interval [1/2; 1).

The Haar transformation (HT) is one of the simplest basic wavelet transforms [19]. Let $f = f_1, \ldots, f_N$ – the one-dimensional discrete signal. The HT divides the signal into two components, the average and the difference. Suppose we have a sub-signal $a^1 = (a_1, a_2, \ldots, a_{N/2})$ consisting of mean values. It is defined as follows:

(1) $$a_n = \frac{f_{2n-1} + f_{2n}}{\sqrt{2}}, \; n = 1,2,\ldots,N/2$$

The detail signal $d^1 = (d_1, d_2, \ldots, d_{N/2})$ at the same level is determined

(2) $$d_n = \frac{f_{2n-1} - f_{2n}}{\sqrt{2}}, \; n = 1,2,\ldots,N/2$$

By means of these values, two new signals are formed:

(3) $$a = \{a_n\}, \; n \in Z; \; d = \{d_n\}, \; n \in Z.$$

The first of these signals represents a coarsened copy of the original, and the second is informative or detailed for the original signal.

(4) $$f_{2n-1} = a_n + d_n, \; f_{2n} = a_n - d_n, \; n \in Z.$$

A similar partition can be performed with respect to the vector a. Let's consider an example with numbers. Let

$$(5)\qquad I = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \\ 8 & 9 & 0 & 1 \\ 2 & 3 & 4 & 5 \end{pmatrix}$$

Let us apply HT to the given matrix, using the principle of division by rows, thus we obtain the following matrix:

$$(6)\qquad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 5 & \vdots & -1 & -1 \\ 9 & 13 & \vdots & -1 & -1 \\ 17 & 1 & \vdots & -1 & -1 \\ 5 & 9 & \vdots & -1 & -1 \end{pmatrix}$$

We apply HT to the given matrix, using the principle of column separation of the obtained matrix. We obtain the following:

$$(7)\qquad \frac{1}{\sqrt{2}} \begin{pmatrix} 10 & 18 & \vdots & -2 & -2 \\ 22 & 10 & \vdots & -2 & -2 \\ -8 & -8 & \vdots & 0 & 0 \\ 12 & -8 & \vdots & 0 & 0 \end{pmatrix}$$

Or in general form we get the following matrix:

$$(8)\qquad \frac{1}{\sqrt{2}} \begin{pmatrix} A & \vdots & H \\ \dots & \vdots & \dots \\ V & \vdots & D \end{pmatrix}$$

$$(9)\qquad A = \begin{pmatrix} 10 & 18 \\ 22 & 10 \end{pmatrix}, H = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix}, V = \begin{pmatrix} -8 & -8 \\ 12 & -8 \end{pmatrix}, D = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Here

A – area, including information about the global properties of the analysed SI;

H – horizontal area, including information on the horizontal components of the analysed SI;

V – vertical area, including information on the vertical components of the analysed SI;

D – diagonal component, including information on the diagonal components of the analysed SI.

When decomposing the image [20–24], the one-dimensional fast Haar transformation (FHT) is applied first to the rows, and then to the columns of pixel values of the input of the displayed matrix. As a result, we have a two-dimensional wavelet transform. As a result, the SI is divided into four equal parts LL, HL, LH, HH:

LL	*HL*
LH	*HH*

Fig. 2. Single application of a two-dimensional wavelet transform to a SI

LL – low-frequency wavelet coefficients (minor quadrant – reduced copy of the original image); HH – high-frequency wavelet coefficients. N-fold two-dimensional wavelet transform implies the application of N times the two-dimensional wavelet transform to the lower quarter of the matrix – LL.

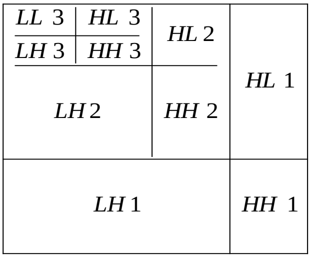

Fig. 3. Three-fold application of a two-dimensional wavelet transforms

Fig. 4. The image of the landfill

The inverse two-dimensional wavelet transform restores the lowest quadrant recursively. We take for the initial function of the space image, the image of the landfill (Fig. 4).

Consider a seven-step fibreboard using db1 (Haar). According to the structure, the database in which information on the waste disposal site (WDS) is stored [9] is presented as:

1. Common (name, type, location, comments, notes, etc.);
2. Geographical parameters (coordinates in different geographical projections, adjacent and including administrative-territorial units (AU), etc.);
3. Geometric (area, perimeter, litter concentration, accuracy of estimation, etc.) and other parameters.

Photographs and images, links to Internet resources, the degree of danger, the time of occurrence of t1 and the disappearance of t2 (their detection in images), notes are also included.

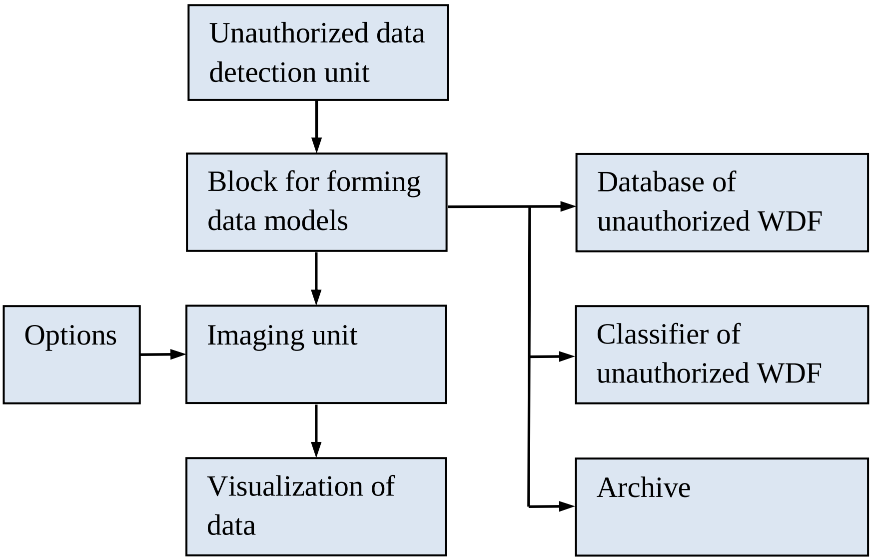

Fig. 5. Block diagram of obtaining a classifier and database of unauthorized WDS

Model of the WDS classifier is represented by the classification code as X.Y, where X – classification code OKATO and Y = ZZZZ – WDS sequence number composed of limiting the hierarchy AU X (unsaturated objects such as Balashihinsky district of Moscow region have Y = 0000). For a complete identification WDS injected additional code, identifying the WDS geographical location on the Earth surface in AB format, where A – latitude, B – longitude in HHHMMSS.SSSS format (HHH – degrees, MM – minutes, SS.SSSSS – seconds; 4 characters S set the maximum geographical accuracy of detection). The classifier stores the current WDS, i.e. those that exist at the current time t (t').

Each sequential WDS entered in the data model is assigned the sequence number N, which connects the database, the classifier and the WDS archive. Historically, WDSs that have disappeared by the time $t_2 < t$ are transferred from the classifier to the archive, with each new WDS assigned the next sequence number. The database stores information about all WDSs (current and historical). So, the principle of the SIs storing has been described above. The work in the database, storing the SIs, is carried out according to the following principle. Each SI is described in a text document containing administrative data (date, time of taking a picture, etc.). Also included are terms and phrases for key search. The documents found are linked directly to the snapshot, which can then be viewed. This is a normal, standard way of constructing a DBMS. However, to store the SIs this method does not provide a complete adequate description. Oral description of the SI, which must be found in the database, may not meet the expectations of the user. The archive, formed on this principle, does not allow the search for "likeness" of an image.

NGNs are systems that allow mechanisms to search for such images. They are relevant in connection with the active use of modern multimedia technologies. Hence, the purpose of the research is the development of the NGN project – the system for storing the SI, a description, and in particular, the WDS.

Thus, the database management system of the SI operates according to the following algorithm (Fig. 5):

- A space image is entered;
- A text document contains administrative information about the next digital image, i.e. date and time of image reception, then terms and phrases that characterize its content are entered, certain keywords are also input [9];
- Input the structure vector of the image.

Let *h, v, d* be the horizontal, vertical, and diagonal components of the wavelet-Haar transformation of the original image or SI.

Definition 1:

The structural vector of the image *f* is:

(10) $$x^f = \left(\sigma_{1h}^2, \sigma_{1v}^2, \sigma_{1d}^2, a_1, \ldots, \sigma_{Jh}^2, \sigma_{Jv}^2, \sigma_{Jd}^2, a_J \right),$$

where $\sigma_{1h}^2, \sigma_{1v}^2, \sigma_{1d}^2$ are the variances [13], $a_j = \dfrac{\sigma_{1v}^2}{\sigma_{1h}^2}$ $(j = 1, \dots, J)$ is the measure of anisotropy. If greater than 1, the structure is oriented vertically; if less than 1, the structure is oriented horizontally.

Definition 2:

Let two pictures f_1, f_2 be given. Images f_1, f_2 are similar, if for, $\forall \varepsilon > 0 \quad \rho(f_1, f_2) < \varepsilon$ where

(11) $$\rho(f_1, f_2) = \left\| x^{f_1} - x^{f_2} \right\| = \sqrt{\sum_{i=1}^{4J} \left(x_i^{f_1} - x_i^{f_2} \right)^2}$$

Algorithm of similarity of two images

Into the DBMS array of space images f_i $i = 1, \dots,$ both administrative data and structural vectors tied to the image are input. Let g be the desired image and search for similar ones in the database are required. The query of the database is carried out according to the following principle:

- Calculation of the structural vector x^g;
- Sorting images f of the database on the principle: $\left\| x^g - x^{f_1} \right\| \le \left\| x^g - x^{f_2} \right\| \le \left\| x^g - x^{f_3} \right\| \le \dots$. As a result, a "response directory" is formed;
- The recipient image directory $f_1, f_2, f_3, \dots$ is given to the user.

To improve the qualitative set of the "response catalogue" it is necessary to use the procedure of "relevant feedback" in accordance with the following algorithm:

- Step 1. Let the user identify the first images of the recipient catalogue as relevant (suitable) or irrelevant (unsuitable). Denote by – most not suitable. Denote by – a set of suitable images;
- Step 2. A new search for similar images starts, for this purpose we introduce a new structural vector

$$x^{new} = \frac{x^g}{\left\| x^g \right\|} - \frac{x^{f_{mis}}}{\left\| x^{f_{mis}} \right\|} + \sum_{f_i \in A^{rel}} \frac{x^{f_i}}{\left\| x^{f_i} \right\|},$$

where $\left\| x^g \right\| = \sqrt{\sum_{i=1}^{4J} \left(x_i^g \right)^2}$;

- Step 3. The response directory is generated, viewed visually. If the result suits the user – Stop, otherwise go to Step 1.

Investigation of the problem of searching similar satellite images on Tikhonov stability

According to [14], we distinguish correctly defined and incorrectly defined problems. We give the notion of well-defined for the problem introduced by J. Hadamard. The solution of any quantitative problem usually consists in finding a "solution" z from the given "initial data" *u*, $z = R(u)$. We shall regard them as elements of the metric spaces *F* and *U* with the distances between the elements $\rho_U(u_1, u_2)$, $\rho_F(z_1, z_2)$; $u_1, u_2 \in U$; $z_1, z_2 \in F$. The metric is determined by the statement of the problem. So, let the concept of "solution" be defined as and to each element there corresponds a unique solution from the space *F*.

Definition 3:

The problem of determining a solution $z = R(u)$ from the space *F* with respect to the initial data $u \in U$ is said to be stable on the spaces (*F*, *U*) if for $\forall\ \varepsilon > 0\ \exists\ \delta(\varepsilon) > 0$, from the inequality $\rho_U(u_1, u_2) \le \delta(\varepsilon)$ follows that $\rho_F(z_1, z_2) \le \varepsilon$, where $z_1 = R(u_1),\ z_2 = R(u_2);\ u_1, u_2 \in U;\ z_1, z_2 \in F$.

Definition 4:

The problem of determining a solution *z* from a space *F* with respect to "initial data" *u* from a space *U* is called a metric space (*F, U*) correctly posed on a pair if the following conditions are satisfied:

1. For any there exists a solution *z* from the space *F*;
2. The solution is uniquely determined;
3. The problem is stable on the spaces (*F, U*).

Tasks that do not meet the listed requirements are called incorrectly delivered.

General algorithm of similarity of two images with regularizing matrix

We give a mathematical formulation of the problem of finding similar images based on Haar wavelets using the regularization matrix. A two-dimensional wavelet transform is a one-dimensional, one-dimensional wavelet transform of rows and columns of this matrix. Firstly, one-dimensional wavelet transforms of each line are performed, after which the converted string is written to its original position. Elements are numbered. Next, wavelet transforms are applied to all columns. The image decomposition is shown in Fig. 2.

In general form, the similarity algorithm for two SIs has the following form: Let $X(i,j)$ $i,j=\overline{1,\ldots,N}$ be the original image or SI. The process of forming a distorted image can be represented as the output of some linear system.

The mathematical model of the process of forming a noisy image has the form:

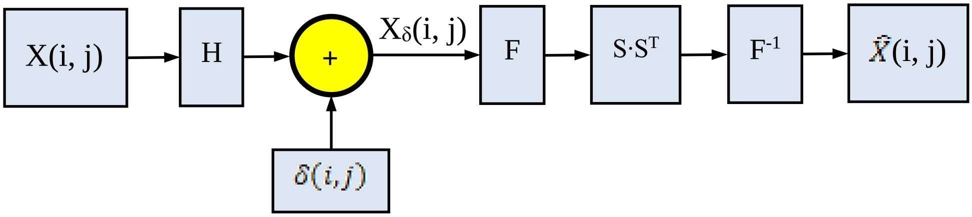

Fig. 6. Model of distorted image formation

The distortion process is modelled as a function of H, which together with additive noise $\delta(i,j)$ acts on the original image of $X(i,j)$ and generates a noisy image $X_\delta(i,j)$

(12) $$X_\delta(i,j)=H[\mathrm{X(i,\ j)}]+\delta(i,\ j)\ .$$

The restoration of the image comprises the construction of the approximation $\tilde{\mathrm{X}}(\mathrm{i,\ j})$ of the original image $X(i,j)$. The higher the approximation accuracy of the image $X(i,j)$ by the function $\tilde{\mathrm{X}}(\mathrm{i,\ j})$, the more is known about the operator H and noise $\delta(i,j)$ [13, 14]. A mathematical model for reconstructing a distorted image is an ill-posed task.

To conduct the experiment, consider a simplified version of the mathematical model:

(13) $$X_\delta(i,j)=\mathrm{X(i,\ j)}+\delta(i,\ j)$$

The problem of noise suppression consists of obtaining an estimate from the total observed signal, as close as possible to $X(i,j)$. We use the method of wavelet transforms with truncation of high-frequency parts of the spectrum using the Tikhonov regularization method.

Description of the algorithm

We will carry out the experiment in the following way. Let's give a mathematical statement of the problem of removing non-informative pixels of the SI with regularization [13]. Let:

- *X(i,j)* $i\,,j\;=\overline{1,\ldots,N}$ be the initial image with data dimensions , considered as the result of some random process with certain properties;
- *F* – discrete orthogonal transformation (Walsh, Fourier, wavelet-Haar, etc.),
- F^{-1} – the inverse transformation;
- *S* – a matrix of dimension choice $m{\times}N$ of rank *m*, $1\le m\le N$,
- R_α – a regularizing matrix in the task of processing the original image *X (i, j)* defined as follows:

(14) $$R_\alpha=\begin{bmatrix}\varphi(1,\alpha) & & & & & \\ & \varphi(2,\alpha) & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \varphi(N,\alpha)\end{bmatrix}$$

where: $\varphi(n,\alpha)$ are regularizing factors; $n=\overline{1,N}$; α is a regularizing parameter.

The problem consists in choosing, for given F_0, S_0, a regularizing matrix R_α such that the following condition holds (where ρ is a given metric):

(15) $$\tilde{X}(i,j)=F_0^{-1}\;S_0^T\;S_0\,F_0\,X(i,j)$$

Comment: For a fixed transformation F_0, arbitrary matrix *S*, $R_\alpha = I$, this problem, known as the band coding problem by means of the transformation *F*, was studied in [6].

We present a general algorithm for compressing two-dimensional signals with a regularizing matrix *R*:

- Step 1: the image of *X (i, j)* is transformed to *F*: $Y(i,j)=F\,X(i,j)$, let $\delta(i\,,\,j)$ be white noise;
- Step 2: the matrix δ of the component approximations $Y_\delta\,(i,j)$ is replaced by the selection operator *S* by a smaller vector $\tilde{Y}_\delta\,(i,j)$ that is to be transmitted over the communication channel, storage, etc. (the value $k=N/N_0$ is called the compression ratio).

- Step 3: an "extrapolation" is carried out using the matrix S^T i.e. on the receiving side, the resulting matrix is complemented to the dimension N (for example, all components except those selected are assumed to be 0).
- Step 4: the resulting matrix undergoes an inverse F^{-1} transformation;
- Step 5: Multiplication by a matrix of regularizing factors $R_{\square}$ is carried out.

As a result of these steps, the original vector is restored with errors:

(16) $$\varepsilon 1 = \rho_{l_2}\left(X(i,j), F^{-1}S^T SF\ \tilde{X}_\delta(i,j)\right), \qquad \varepsilon 2 = \rho_C\left(X(i,j), R_\alpha F^{-1}S^T SF\ \tilde{X}_\delta(i,j)\right)$$

The problem consists in choosing α and therefore $\varphi(n,\alpha)$ so that for the given k the following condition is satisfied: $\varepsilon 2 << \varepsilon 1$. The choice of α and $\varphi(n,\alpha)$ for the corresponding orthogonal transformations depends on the input data and the transformation structure. The form of the matrix R_α is defined for the continuous case and is used in a discrete interpretation of the image compression problem with regularization. Table 1 gives a specific form of the matrix R_α for various transformations.

Table 1. Specific form of the matrix R_α for various transformations

Fourier	$\varphi(\alpha,k) = \dfrac{1}{1+k^2\alpha}$	$k = \overline{1,\infty}$	[6]
Walsh	$\varphi(\alpha,k) = \dfrac{1}{1+k^p\alpha}$	$k = \overline{1,\infty}$; $p > 1/2$	[6]
Wavelet Haar	$\varphi(\alpha,k) = \dfrac{1}{1+(\alpha k)^\lambda}$	$k = \overline{1,\infty}$; $\lambda > 1/2$	[6]

The table does not indicate the exact values of α. The question arises of determining the values of α, for which the conclusions of the experiment are valid. The specific value of the regularizing parameter α is determined experimentally and is given in Table 2. Here, α_1 is the lower bound; α_2 is the upper limit of the change α.

Table 2. Experimental specific values of the regularizing parameter α

	COS	***WAVELET-HAAR***	***WALSH***	***FOURIER***
	0.5	*0.99*	*0.5*	*0.99*
	0.00005	*0.000001*	*0.000001*	*0.000001*

A chart is shown on Fig. 7 where the arrangement of compression errors with regularization *ε2* and without regularization *ε1* for the average experimental parameters is presented.

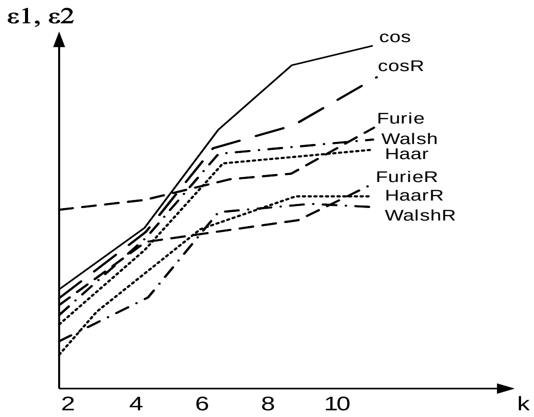

Fig. 7. Compression errors with regularization and without regularization for averaged experimental parameters

Conclusion

The project to create an automated space monitoring system for the presence of a WDS on the territory of the Russian Federation, designed to provide an operative survey of the studied area, the detection and mapping of real and potential foci of unauthorized dumps, the forecast of their development, a preliminary assessment of the scale of disasters and possible consequences is of great urgency. The current article explores the project model. The use of the data warehouse in the task of space monitoring of the WDS enables not only improving of the performance of environmental authorities, but also to enhance the monitoring effectiveness of the WDS (space, aerological and terrestrial).

References

1. Aristov, M. A. Monitoring of solid waste landfills and detection of spontaneous garbage disposal according to space survey data. GeoProfil.ru. № 2. http://internetgeo.ru/ uploads/ journals/geoprofile0209/monitoring_tbo.pdf
2. Pogorelov, A. V., S. V. Dulepa, D. A. Lipilin. Experience of space monitoring of landfills in the territory of the Krasnodar Territory. Geomatika 2013, № 4, pp. 64–71. http://geomatica.ru/pdf/2013_04/2013_4_64-71.pdf

3. Avetisyan, D., R. Nedkov. Determining the magnitude and direction of land cover changes in the semi-natural areas of Haskovo Region, Southeast Bulgaria. Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, IEEE, 2015, pp. 4637–40. DOI:10.1109/IGARSS.2015.7326862.
4. Stankova, N., R. Nedkov. Monitoring forest regrowth with different burn severity using aerial and Landsat data. Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, IEEE, 2015, pp. 2166–69. ISBN:978-1-4799-7929-5, ISSN: 2153-7003, DOI:10.1109/IGARSS.2015.7326233.
5. Stankova, N., R. Nedkov, I. Ivanova, and D. Avetisyan. Integration of multispectral and SAR data for monitoring forest ecosystems recovery after fire. Proc. SPIE 10444, Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017), 104441J, 10444, SPIE, 2017, ISBN:978-9963-697-24-3, DOI:10.1117/12.2277313
6. Nedkov, R. Quantitative assessment of forest degradation after fire using ortogonalized satellite images from Sentinel–2. Comptes rendus de l’Academie bulgare des Sciences, 71, 1, publishing house of BAS „Prof. Marin Drinov“, 2018, ISSN: 2367-5535
7. Nedkov, R. Normalized differential greenness index for vegetation dynamics assessmenT. Comptes rendus de l’Academie bulgare des Sciences, 70, 8, Prof. Marin Drinov Publishing House of Bulgarian Academy of Scinces, 2017, pp. 1143–46, ISSN:1310–1331.
8. Panayotova, D., R. Nedkov, M. Dimitrova, I. Ivanova, and M. Zaharinova. Eco-Monitoringal Investigation of Forest end Land-Used Area and Agro-Climate Characteristics in the Land of the Municipality of Kardzhali, Using Aerospace and GPS Data. Journal Biotechnology & Biotechnological Equipment Volume 23, 2009 - Issue sup1: XI Anniversary Scientific Conference, 23, sup1, Taylor & Francis, 2009, ISSN: 1310-2818, pp. 200–03, DOI:10.1080/13102818.2009.10818400
9. Avetisyan, D., R. Nedkov, E. Velizarova, and D. Borisova. SAR and optical data in land degradation processes estimation: a case study from Southeast Bulgaria. Proc. SPIE 10428, Earth Resources and Environmental Remote Sensing/GIS Applications VIII, 10428, SPIE, 2017, pp. 104281D-1-104281D-9. ISSN: 0277-786X, DOI:10.1117/12.2277945.
10. Aleksandrovna, L. N., M. L. Glinsky, A. A. Zubkov, V. V. Danilov, and L. G. Chertkov. Monitoring of the state of the subsoil on the LRW burial site. http://www.atomic-energy.ru/articles/2012/02/20/31086
11. Brovkina, O. V. Remote monitoring of anthropogenic disturbances in the taiga zone of the Northwest of Russia. PhD dissertation, 2011, 125 p. http://www.dissercat.com/content/distantsionnyi-monitoring-antropogennykh-narushenii-taezhnoi-zony-severo-zapada-rossii
12. Yannikov, I. M., M. V. Sleptsova, and K. K. Krivoshapkin. Organization of biomonitoring of chemically hazardous objects and landfills with the use of identification polygons. Bulletin of the North-Eastern Federal University. M.K. Ammosov, NEFU, Yakutsk, 2013. pp. 15–25. http://cyberleninka.ru/article/n/organizatsiya-biomonitoringa-himicheski-opasnyh-obektov-i-poligonov-tbo-s-ispolzovaniem-identifikatsionnyh-poligonov

13. Kazaryan, M. L., M. A. Schahramanian, and A.A. Richter. Space monitoring of the Earth and Haar wavelet transform. Future Communication Technology and Engineering, Chapter 61, Edited by Kennis Chan CRC Press 2015, pp. 291–94. Print ISBN: 978-1-138-02777-0. eBook ISBN: 978-1-315-69045-2. DOI:10.1201 / b18331-65
14. Kazaryan, M. L., M. A. Schahramanian, and A.A. Richter. Multidimensional Haar wavelet-transient research on the stability in the context of a satellite monitoring of the Earth problem//Serbian journal of electrical engineering (in press)
15. Kazaryan, M., M. Shakhramanyan, R. Nedkov, A. Richter, D. Borisova, N. Stankova, I. Ivanova, and M. Zaharinova. Research of the generalized wavelet transformations of Haar correctness in remote sensing of the Earth//Proc. of SPIE Vol. 10427-74 (Conference 10427: Image and Signal Processing for Remote Sensing Conferences: 11–14 September 2017, Warsaw, Poland), p. 99. https://spie.org/ERS/conferencedetails/image-signal-processing-remote-sensing#227857
16. Richter A. A., M. L. Kazaryan, M. A. Shakhramanian, D. Borisova, N. Stankova, I. Ivanova. Information modeling of waste disposal sites. Ecological Engineering and Environment Protection, 1, 2017, pp. 15–21, URL: http://ecoleng.org /archive/2017/1/15-21.pdf
17. Chui, C. K. Wavelets: A Mathematical Tool for Signal Analysis. SIAM, Philadelphia, 1997.
18. Daubechies, I. Ten Lectures on Wavelets. SIAM, Philadelphia, 1992.
19. Gonsales, R. C., R. E. Woods, S. L. Eddins. Digital Image Processing using Matlab, Pearson Prentice Hall, Upper Saddle River, NJ, 2004.
20. Korolyuk, V. S., N. I. Portenko, A. V. Skorokhod, A.F. Turbine, et al. A handbook on probability theory and mathematical statistics. M.: Science. The main edition of physics and mathematics, 1985, 640 p.
21. Tikhonov, A. N. and V. Ya. Arsenin. Methods for solving ill-posed problems. M.: Science. Home edition of physical and mathematical literature, 1979, 288 p.
22. Mallat, S. A wavelet tour of signal processing. N. Y.: Acad. Press, 1999, 851 p.
23. Nedkov, R. Orthogonal transformation of segmented images from the satellite Sentinel-2. Comptes rendus de l'Acad. bulg. des Sciences, 70, 5, Prof. Marin Drinov Publishing House of Bulgarian Academy of Sciences, 2017, ISSN:1310–1331, pp. 687–92.
24. Anuj, B. and A. Rashid. Image compression using modified fast Haar wavelet transform. World Appl. Sci. J. 2009. 7, 5. pp. 647–53.

ИЗСЛЕДВАНЕ НА СТАБИЛНОСТТА НА АЛГОРИТЪМ ЗА ПОДОБИЕ, ИЗПОЛЗВАН ВЪРХУ СИСТЕМА ЗА СЪХРАНЕНИЕ НА СПЪТНИКОВИ СНИМКИ ЧРЕЗ ХААР УЕЙВЛЕТИ, СПОРЕД ТИХОНОВ

М. Казарян, М. Шахраманян, С. Забунов

Резюме

Дистанционните изследвания на Земята позволяват получаването на информация със средна до висока пространствена разделителна способност от космически летателни апарати, с хиперспектрални инструменти на борда.

В настоящата статия е разработен модел на автоматизирана система за откриване на сметища чрез космически наблюдения (АСОС). Един от компонентите на модела е модулът за съхранение на космически снимки, който позволява не само да се подобри ефективността на институциите, занимаващи се с опазване на околната среда, но и също така да се увеличи производителността на мониторинга на сметища, базиран на космически изследвания, въздушни наблюдения и наблюдения от Земята. Освен това се изследва проблемът за търсене на подобия в архиви с космически снимки чрез използване на Хаар уейвлет трансформацията за стабилност.

Тук се използва регуляризационният метод на Тихонов, елементи на математическия анализ, теория на дискретните ортогонални трансформации и методи за декодиране на космически снимки.

Научна новост в разработката е предложеният модел на автоматизирана система за космически мониторинг, предназначена за откриване на сметища. Задачата за обработка на архивирани космически снимки, и по-конкретно търсенето на подобия в архивите от космически снимки, е изследвана от гледна точка на некоректно дефинирана задача.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e07

QUATERNION-BASED AUTOPILOT FOR DODECACOPTERS - PART II

Svetoslav Zabunov

Space Research and Technology Institute – Bulgarian Academy of Sciences

Keywords: *Unmanned Aerial Vehicle (UAV), Autopilot, Dodecacopter.*

Abstract

"Quaternion-based Autopilot for Dodecacopters - Part I" was published in Aerospace Research in Bulgaria journal, book 28/2016. The theoretical model at that time, evolved into a prototype autopilot system during the last two years. This small as per its dimensions, but large according to its capabilities, autopilot is now presented herein.

The autopilot is given the nomenclature identification "Z-Pilot Nano" and is currently at version 1.0 of its development. The device is scrutinized from the point of view of its major application platform – micro-drones weighing not more than 250 g. Furthermore, Z-Pilot Nano has been installed in Bulgarian Knight UAV and elaboration on this implementation is disclosed in the current article.

The autopilot is also applicable to other aircraft, as well as land and sea unmanned vehicles, due to its versatility and high performance parameters.

Introduction

The development of an autopilot for dodecacopters based on quaternion mathematical model began three years ago. The author published the Part I of this article series back in 2016 in Aerospace Research in Bulgaria journal, book 28. Further, the work continued and two years later resulted in the creation of a 12-rotor helicopter autopilot with miniature dimensions and weight. There were two motivations for creating the prototype such small in size and also versatile in capabilities:

1. The modern trend of using very small UAVs for various purposes [1–4].
2. The application of the autopilot to both dodecacopters and larger, by the number of rotors, drones.

The proposed autopilot is also suitable for land and sea unmanned vehicles. The autopilot prototype was named "Z-Pilot Nano" and has been mounted on a micro-dodecacopter model Bulgarian Knight (see Fig. 1).

The recently raised interest in micro unmanned aerial vehicles (micro-UAVs) is by no means random. Micro-drones are hard to detect and track by any means of physics or any technical instruments [5]. The manifested benefits of micro-drones are the driving force behind their extraordinary popularity during the last few years. There doesn't exist a single specific threshold of the total weight that, being below, a drone will be considered a micro-drone. Different scales of measure have been elaborated, but law changes, from not long ago, in some countries have established the 250 g total weight upper limit as the most often recognized criterion for classification of micro-drones. A similar upper limit of 25 g would generally classify a drone as a nano-drone.

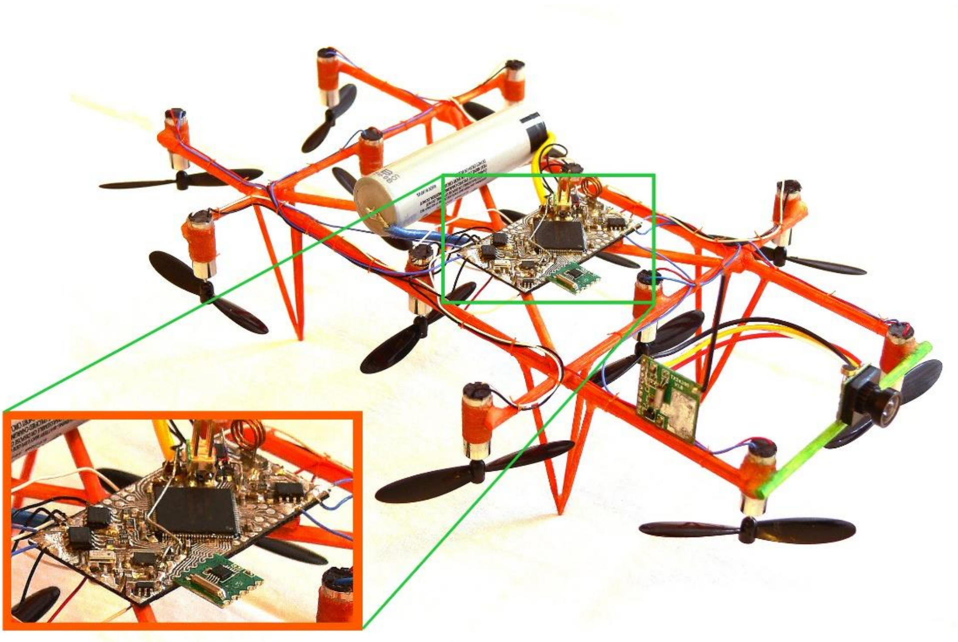

Fig. 1. Autopilot Z-Pilot Nano on board of Bulgarian Knight dodecacopter micro-drone. The aircraft total weight is 121 g and has dimensions: 280 × 185 × 40 mm.

Z-Pilot Nano prototype design

In Fig. 2, a block diagram of the Z-Pilot Nano design is presented. The heart of the system is a modern microcontroller with DSP capabilities from semiconductor division of Philips – NXP. Such devices are called Digital Signal Controllers (DSCs). The chip has a processor with ARM 32-bit architecture – Cortex-M4. The core of the processor is ARMv7-M. The processor is working at 120 MHz clock speed. Cortex-M4 architecture offers 3-stage pipeline with branch prediction and internal data and code cache memory. The RAM of the chip is 128 kiB (kibibyte) and the flash memory is 1 MiB (mebibyte). The chip is supplied with a single precision IEEE 754 compliant math coprocessor that executes most 32-bit floating point operations with 1 clock cycle latency.

The DSP capabilities offered by the device are:

- Single cycle 16/32-bit MAC (multiply-accumulate operation).
- Single cycle dual 16-bit MAC (multiply-accumulate operation).
- 8/16-bit SIMD arithmetic.
- Hardware Divide (2–12 Cycles).

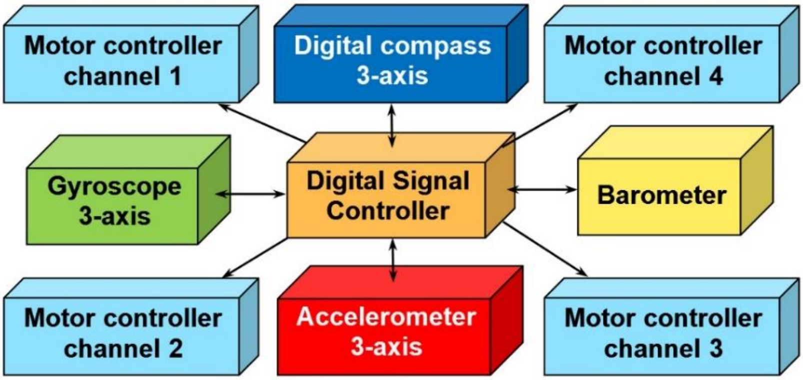

Fig. 2. Z-Pilot Nano principal block diagram

The employed digital signal controller offers a number of timers, analogue to digital converters, digital to analogue converters, communication interfaces and other peripherals. The chosen DSC variant occupies an SMD 100-pin LQFP package with dimensions of 14×14 mm (see Fig. 3). Device supply voltage range is from 1.71 to 3.6 V and its temperature range (ambient) is from –40 to +105 °C.

Fig. 3. NXP Digital Signal Controller implemented as the main IC in Z-Pilot Nano. The used variant is installed into an SMD LQFP 100-pin package with dimensions 14×14 mm.

The digital signal controller is communicating over I^{2}C interface with a number of devices on the PCB of the autopilot as denoted on Fig. 2 with two-way arrows. Further, the DSC is managing four motor controllers through one-way signals.

The inertial and altimeter sensors of the autopilot are of the MEMS type (microelectromechanical systems). The utilization of MEMS sensors into UAV autopilots has peaked lately [6–8].

Altimeter measurements are performed by means of a digital barometer IC – MPL3115A2 (Fig. 4 – left). This device employs a MEMS pressure sensor with range of 20 to 110 kPa. The pressure range translates to altitude range from –500 m to +12000 m from sea level. As already stated, the digital communication interface is I^2C. This device also implements an accurate thermometer. The sensor output accuracy is 20-bit for pressure/altimeter information and 12-bit for temperature data. The altimeter accuracy guarantees maximum altitude resolutions of 10 cm. Analogue signals are internally digitized by means of 24-bit high-resolution ADCs. There is an on-chip processing of the acquired data, which eliminates the need for compensation calculations from the DSC (host controller). MPL3115A2 offers a variety of operating modes that include engaging an interrupt signal, setting up a polling only mode, autonomous data acquisition, etc. The device is extremely low-powered and consumes only 40 μA per measurement per second. Supply voltage range is from 1.95 to 3.6 V. The package is SMD LGA type and has dimensions of 5.0×3.0×1.1 mm. Temperature range of operation is from –40 to +85 °C.

Fig. 4. ICs engaged in Z-Pilot Nano as sensors. From left to right: Digital barometer MPL3115A2; Digital IMU – LSM9DS1 top view; LSM9DS1 bottom view.

The other three sensor devices are packaged in one IC – LSM9DS1. This integrated circuit hosts a digital 3-axis gyroscope, a digital 3-axis accelerometer, a digital 3-axis magnetometer (digital compass), and a digital thermometer. All four devices communicate with the DSC through I^2C interface. The chip is powered by 1.9 to 3.6 V supply voltage and consumes not more than 4.6 mA of current. Operating temperature range is from –40 to +85 °C.

The gyroscope has angular rate range of ±245/±500/±2000 degrees per second. Its maximum output data rate (ODR) is 952 Hz and its accuracy is 16 bits.

The digital accelerometer acceleration range is ±2/±4/±8/±16 g and its maximum output data rate (ODR) is 952 Hz, again at 16-bit accuracy. The digital compass's range is ±4/±8/±12/±16 gauss, which equals ±0.4/±0.8/±1.2/±1.6 mT. Maximum magnetometer ODR is 80 Hz and data accuracy is 16 bits. The digital thermometer inside the LSM9DS1 device has temperature range from –40 to +85 °C and accuracy of 12 bits. The temperature sensor ODR is 59.5 Hz.

There are four motor controllers on board of the Z-Pilot Nano autopilot. Each motor controller is capable of delivering 3 A output current and thus manage 4 micro-motors of the 7×20 mm type or 3 micro-motors of the 8.5×20 mm type. A dual MOSFET transistor IC has been employed in the motor controllers – the si4564dy from Vishay, see Fig. 5. The IC contains one N-channel and one P-channel MOSFET transistor with on-state resistance of not more than 20 mΩ. Total switching capacitance of one motor controller's MOSFET ICs is 3215 pF. The energy loss for one switch on/switch off cycle is hence:

(1) $$E = \frac{C \times V^2}{2} \times 2 = 3215 \times 10^{-12} \times 3.3^2 = 35 \times 10^{-9}\ \mathrm{J}$$

Where the voltage transition was calculated at the regulated 3.3 V voltage supply of the controlling DSC chip.

Fig. 5. Dual MOSFET transistor IC si4564dy manufactured by Vishay, employed in motor controllers. Contains two MOSFETs – one N-channel and one P-channel transistor.

The switching frequency is 60 kHz. It follows that the switching power loss is:

(2) $$P = E \times 60000 = 2.1\ \mathrm{mW}$$

For all four motor controllers the switching power loss is equal to 8.4 mW. At 3.3 V supply voltage this power consumption translates to 2.5 mA current draw for all four motor controllers.

Table 1 summarizes all devices installed in the autopilot with their supply voltage range, current consumption, and operating temperature range.

There are still more devices of minor significance in the autopilot, such as voltage regulator, current sense circuit, filtering capacitors, etc. The latter are not elaborated on in this article for their implementation being trivial.

Table 1. Major parameters of the devices employed in Z-Pilot Nano

Device	Supply voltage range	Typical current consumption	Operating temperature range
Digital signal controller	1.71 ÷ 3.60 V	50.10 mA	-40 ÷ +105 °C
Barometric sensor	1.95 ÷ 3.60 V	0.04 mA	-40 ÷ +85 °C
Gyroscope	1.90 ÷ 3.60 V	4.00 mA	-40 ÷ +85 °C
Accelerometer and digital compass	1.90 ÷ 3.60 V	0.60 mA	-40 ÷ +85 °C
Motor controllers (four units)	2.60 ÷ 16.00 V	2.50 mA	-40 ÷ +85 °C
Overall voltage range:	2.60 ÷ 3.60 V		
Total typical current drain:		57.24 mA	
Overall temperature range:			-40 ÷ +85 °C

The autopilot is powered by a single cell Li-Ion battery and thus all voltage requirements in Table 1 are fulfilled.

Printed circuit board

The construction of Z-Pilot Nano has been establish around a printed circuit board with dimensions of 50×25 mm. The board offers general purpose input/output connections, as well as, general purpose analogue inputs. There is a connection for external I^2C devices; connections for motor wiring and a 6-pin socket for programming the digital signal controller (see Fig. 6).

All elements are surface mount. Some devices have leadless packages. Hence, the soldering process of the prototype requires special laboratory equipment for precise device positioning and heating.

The first prototype of the autopilot has been built on 1 mm thick printed circuit board. This thickness proved to be overdone and resulted in 5 g overall autopilot weight. Another prototype is under construction using 0.4 mm thick FR-4 board. The calculations show that the new prototype will weigh around 3 g. The autopilot, as mentioned earlier, is capable of driving 16 micro-motors of the

7×20 mm type. Managing a 16-rotor aircraft, the autopilot will have a negligible weight share of the aircraft weight. When the new prototype is installed in the Bulgarian Knight micro-drone it will occupy only 2.5 % of the total aircraft weight – the current prototype takes up the inacceptable 4.1 % of the total drone weight.

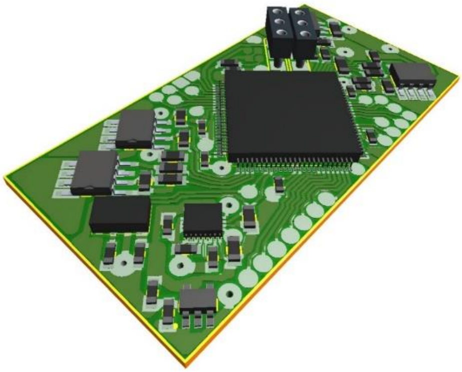

Fig. 6. The Z-Pilot Nano printed circuit board. The board dimensions are 50×25 mm, weight: 3-5 g (see text).

Conclusions

The author is constantly developing the Z-Pilot Nano and enhancing its characteristics. The roadmap is leading to the installation of the device into micro-drones of various types [9], most of which are multirotor drones with the number of rotors greater than or equal to 12. Further, Z-Pilot Nano is envisioned as the major controlling circuit of land and sea unmanned vehicles and robots, as well as different standalone devices implemented in measuring diverse parameters of the environment.

References

1. Kortunov, V. I., O. V. Mazurenko, A. V. Gorbenko, W. Mohammed, and A. Hussein. Review and comparative analysis of mini- and micro-UAV autopilots. 2015 IEEE International Conf. Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), 13-15.10.2015, DOI:10.1109/APUAVD.2015.7346622
2. Chao, H. Y., Y. C. Cao, and Y. Q. Chen. Autopilots for Small Unmanned Aerial Vehicles: A Survey. International Journal of Control, Automation, and Systems, 2010, 8, 1, 36–44.
3. Zhih, C. C., S. K. V. Ragavan, and M. Shanmugavel. Development of a simple, low-cost autopilot system for multi-rotor UAVs. 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 10-12 Dec. 2015, DOI:10.1109/ RAICS.2015.7488429

4. Coffey, T. and J. A. Montgomery. The Emergence of Mini UAVs for Military Applications. Defense Horizons Dec. 2002, No. 22.
5. Birch, G. C., J. C. Griffin, and M. K. Erdman. UAS Detection, Classification, and Neutralization: Market Survey 2015. SANDIA REPORT, SAND2015-6365, Unlimited Release, 1–74.
6. Esper, I. de M. and P. F. F. Rosa. Heading Controller for a Fixed Wing UAV with Reduced Control Surfaces Based on ANFIS. 2015 IEEE International Conference on Dependable Systems and Networks Workshops, 22–25 June 2015, DOI:10.1109/DSN-W.2015.27
7. Jang, J. S. and D. Liccardo. Automation of small UAVs using a low cost MEMS sensor and embedded computing platform. In 25th Digital Avionics Systems Conference, 2006 IEEE/AIAA, October 2006, 1–9.
8. Kortunov, V. I., I. Yu. Dybska, G. A. Proskura, and A. S. Kravchuk. Integrented MimiINS Based on MEMS Sensor for UAV Control. IEEE A&E System magazine. 2009, 41–43.
9. Zabunov, S., P. Getsov, and G. Mardirossian. XZ-4 vertical takeoff and landing multi-rotor aircraft. Asian Journal of Natural & Applied Sciences, 2014, 3, 4.

АВТОПИЛОТ ЗА ДВАНАДЕСЕТОКОПТЕРИ, БАЗИРАН ВЪРХУ КВАТЕРНИОНИ – ЧАСТ II

С. Забунов

Резюме

Статията “Автопилот за дванадесетокоптери, базиран върху кватернионни – част I” беше публикувана в списанието *Aerospace Research in Bulgaria*, книжка 28, година 2016. Теоретичният модел, създаден тогава, се разви и достигна прототипна фаза. Този малък според своите размери, но голям в зависимост от своите възможности, автопилот сега е представен в настоящата статия.

Автопилотът получи номенклатурна идентификация “Z-Pilot Nano” и текущата версия на разработката е 1.0. Устройството е подробно описано от гледна точка на своята основна платформа на приложение – микро-дроновете, тежащи не повече от 250 g. Z-Pilot Nano е инсталиран на микро-хеликоптера „Български рицар“, като цялата машина има тегло 121 g.

Платката на автопилота е подходяща както за различни видове безпилотни летателни апарати, така и за безпилотни наземни или водни машини, поради своята гъвкавост и висока производителност.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e08

ELECTRIC VERTICAL TAKE-OFF AND LANDING FIXED WING UNMANNED AIRIAL VEHICLE FOR LONG ENDURANCE OR LONG RANGE?

Dimo Zafirov

Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: zafirov@space.bas.bg

Keywords: *Long endurance, Long range, Joined wing, Ducted fan.*

Abstract

An analysis of requirements to electric vertical take-off and landing unmanned aerial vehicle with fixed wings is carried out in this article. These aircraft have to fulfil requirements of users and to be convenient for operation in any field conditions. Long flight duration and long flight range are important for most missions. Mathematical models for both cases are presented and it has been found that the requirements for the wing load are different.

It is recommended to use a type of UAV (Unmanned Aerial Vehicle) that is modular and allows performing flights with different configurations and payload depending on the mission in order to fulfil these requirements.

Notation

$\boldsymbol{C_D}$ – coefficient of drag force;

$\boldsymbol{C_L}$ – coefficient of lift force;

$\boldsymbol{D}$ – drag force of the aircraft;

$\boldsymbol{E_{bat}}$ – energy of the batteries;

$\overline{\boldsymbol{E_{bat}}}$ – specific energy of the batteries;

$\boldsymbol{F}$ – thrust of the propulsions;

$\boldsymbol{g}$ – acceleration of gravity;

$\boldsymbol{K}$ – glade ratio;

$\boldsymbol{K_e}$ – glade ratio by maximum endurance;

$\boldsymbol{K_R}$ – glade ratio by maximum distance;

$\boldsymbol{L}$ – lift force of the aircraft;

$\boldsymbol{m_0}$ – take-off mass;

$\boldsymbol{m_p}$ – mass of the payload;

$\boldsymbol{m_{bat}}$ – mass of the batteries;

$\boldsymbol{m_{empty}}$ – empty mass of the aircraft;

$\overline{\boldsymbol{m_{bat}}}$ – specific mass of the batteries;

$\overline{\boldsymbol{m_p}}$ – specific mass of the payload;

$\overline{\boldsymbol{m_{empty}}}$ – specific mass of the empty aircraft;

$\boldsymbol{P}$ – power;

$\boldsymbol{R}$ – distance of the flight;

$\boldsymbol{S}$ – wing area;

$\boldsymbol{t}$ – flight time;

$\boldsymbol{t_e}$ – endurance time;

$\boldsymbol{V}$ – air speed of the aircraft;

$\boldsymbol{V_c}$ – cruise speed of the aircraft;

$\boldsymbol{W_0}$ – take-off weight;

$\boldsymbol{\rho}$ – air density.

1. State of the Arts

The number of UAV has been increasing exponentially because of a large number of applications where they can be used. Different types of configurations, propulsors, sources of energy, and payloads have been used.

Different UAV, that perform vertical take-off and landing, have appeared recently. Most of them are multi-rotary UAV (copters) that are suitable for mission where stationary hanging or target tracking are required. These UAV possess low energy efficiency due to their low glade ratio ($K \approx 1$) that defines relatively short endurance. In addition, they cannot be used with strong wind.

Fixed wing UAVs possess better glare ratio and flight duration. However, they need even runways for take-off and landing and these aircraft are not capable to perform motionless hovering flight.

In this respect, designers have been working persistently on a combination of fixed wing aircraft and copter in order to achieve their advantages. In many cases, these UAV do not possess good weighting perfection because they are powered with different propulsors for vertical and horizontal flight. This increases their empty weight.

Electrical UAVs are favourites in terms of convenience in maintenance. These aircraft can be prepared for flight quickly and they possess a high level of airworthiness.

The article discusses the possibility for development of VTOL e-UAV that is convenient in maintenance. The aircraft will be able to take off and land vertically, and will be able to perform a motionless hovering flight. In addition, e-UAV will possess long endurance and long range.

2. Initial Requirements

Marketing analyses shows that customers would like to have UAV that possess long endurance or long range, vertical take-off and landing, vehicles that are able to perform motionless hovering flight, aircraft that are capable to localize and follow targets, or perform scanning of large areas. Payload mass of cameras and other equipment, all together with the gimbals, is less than 2 (two) kilograms. Customers require a high level of reliability and simple maintenance and flight operations.

Customers have different requirements according to flight missions. Long endurance is required with missions for radio signal retranslation or monitoring of small areas.

Long range is required with mission for:

- Surveying and Mapping;
- Precision Agriculture;
- Search and Rescue;
- Security;
- Logistics.

Some customers admit that this means maximum distance between UAV and GCS with sustainable receiving of video-information and transmission of control signals. Meanwhile, long range is required for most missions (for example – scanning).

Customers often look for long endurance UAVs for missions where scanning of large areas and remote objects are carried out and UAVs fly a great distance. With aircraft that perform horizontal take-off and landing, speeds providing long endurance and long range are almost equal so that the requirements are fulfilled by the same aircraft.

3. Requirements

Let us look at separate subsystems of Unmanned Aviation System (UAS) that fulfils customer requirements completely.

3.1. The structure

Aerodynamic and weight efficiency is defined by the aircraft configuration. One of possible configurations is the scheme Joined Wing that provides a great level of lift [1] and produces a low level of drag. Joined Wing scheme possesses high weight efficiency due to a high level of frame strength that the wing creates and due to the absence of landing gear.

The scheme provides the possibility for vertical take-off and landing of UAV only through a simple rotation of one or several propulsors. In this case, the thrust always has to pass through the centre of gravity of the aircraft. (Fig. 1).

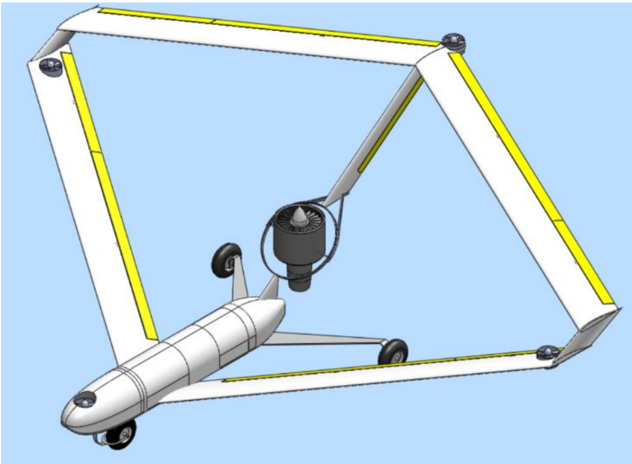

Fig. 1. UAV "Joined Wing" Configuration

3.2. Propulsors

Propellers, ducted fans, and jet engines are used for thrust (moving power) creation on UAV. A small number of military UAV are powered by jet engines. Marketing researches [2] show that a large number of UAV will be used for civilian applications. These aircraft are powered by air propellers or ducted fans.

Propellers for VTOL UAV possess great dimensions and they are inefficient during a horizontal flight. Aircraft that are powered by one type of propulsors for vertical flight and hanging and other type for horizontal flight have been used recently. However, this scheme makes their usage inefficient in a weighting aspect.

Ducted fans are used more frequently in comparison with air propellers due to their smaller dimensions at equal thrust. They work effectively during take-off, landing and at cruise velocity [3] Fig. 2.

Fig. 2. Ducted Fan Propulsor

In addition, ducted fans are effective during horizontal and vertical flight and also for control that improve the thrust efficiency of the aircraft.

3.3. Energy sources

Electric batteries are the most suitable source of energy for UAV and 96 % of UAV are electrical at the moment. Batteries have been developed very quickly and the specific energy of the new models has increased recently. NASA forecast [4] shows that the specific energy is expected to reach 1.200 Wh/kg in 2026 by going up more than four times. Specific energy is a basic feature of the electric battery defining flight endurance and range of UAV.

3.4. Electric motors

Electric motors that are made of neodymium alloys such as N50 and N52 possess a great specific power. Some electric motors, produced by SIEMENS, reach a specific power of 6 kW/kg that provide good thrust efficiency for UAV. In addition, revolutions of the motor can be controlled quickly and properly.

4. Basic features of e-VTOL UAV

Basic features of e-VTOL UAV are defined at the following flight modes (mission elements):

- Take-off;
- Climbing;
- Horizontal flight;

- Long endurance;
- Long range;
- Hovering flight;
- Descending;
- Landing.

Gimbals with mounted payload at weight of 2 kg are the most commonly used. We also place requirements e-VTOL UAV in development to possess maximum endurance or maximum range. These features can be achieved when the battery weight is maximum.

EASA standards for specific missions carried out by UAV require the take-off weight of aircraft not to exceed 25 kg. It is a limit value. When the take-off weight is more than 25 kg, UAV will belong to a certified category of UAV with higher requirements and expensive procedures of certification. Then:

(1) $$m_o = m_p + m_{bat} + m_{empty} \ [kg];$$

(2) $$1 = \frac{m_p + m_{bat} + m_{empty}}{m_0};$$

(3) $$\overline{m_{bat}} = 1 - \overline{m_{empty}} - \overline{m_p}.$$

The empty weight of UAV depends on many characteristics of the aircraft, materials and components [5–7]. In order to achieve a maximum value for t_e *or* R, m_{empty} has to be minimum.

Take-off mode, landing mode, and ascending mode will be short-term. The main part of the energy will be used for the horizontal flight – to achieve a maximum endurance t_e , or a maximum range R.

Maximum values for t_e *or* R depend on the amount of energy that can be provided by the batteries:

(4) $$E_{bat} = \overline{E_{bat}} m_{bat} \ [Wh],$$

where $\overline{E_{bat}}$ *[Wh/kg]* is the specific energy of the battery.

At the moment, the maximum value of specific energy of $\overline{E_{bat}} \approx 300$ Wh/kg is provided by batteries based on lithium. It is expected that they will be leaders in the future and their specific energy will reach a value of 1.200 Wh/kg until 2026.

In a horizontal flight of electric planes, following formulas are applicable:

(5) $$L = W_0 = C_L S \frac{\rho V^2}{2} \ [N]$$

(6) $F = D = C_D S \frac{\rho V^2}{2}$ [N]

(7) $K = \frac{L}{D} = \frac{W_0}{F} = \frac{C_L}{C_D}$

(8) $F = \frac{W_0}{K} = \frac{m_0 g}{K}$ [N]

(9) $P = FV$ [W]

(10) $E = Pt = FVt = FR$ [Wh].

4.1. Maximum endurance

For maximum endurance mode of flight:

(11) $L_e = W_0 = m_0 g = C_L S \frac{\rho V_e^2}{2}$ [N]

(12) $F_e = D_e = C_D S \frac{\rho V_e^2}{2}$ [N]

(13) $K_e = \frac{L_e}{D_e} = \frac{W_e}{F_e} = \frac{C_{L_e}}{C_{D_e}}$.

Maximum endurance mode is achieved when the ratio C_L^3 / C_D^2 is maximum and:

(14) $P_e = \frac{F_e V_e}{\eta}$ [W]

(15) $F_e = \frac{m_0 g}{K_e}$ [N]

(16) $V_e = \sqrt[2]{\frac{2L}{\rho C_{L_e} S}} = \sqrt[2]{\frac{2 m_0 g}{\rho C_{L_e} S}} = \sqrt[2]{\frac{2 g p_0}{\rho C_{L_e}}}$ [m/s]

(17) $E_A = P_e t = F_e V_e t = \frac{m_0 g}{K_e} R$ [Wh)

(18) $t_e = \frac{E_A}{P_e} = \eta \frac{\bar{E}_A m_{bat}}{F_e V_e} = \eta \frac{\bar{E}_A m_{bat}}{\frac{m_0 g}{K_e} V_e} = \eta \frac{\bar{E}_A \overline{m_{bat}} K_e}{g V_e}$ [s].

A maximum flight time can be achieved at a maximal value of the ratio $\frac{\overline{m_{bat}}}{V_e}$, that is reached at ${p_0}_{opt}$. When increasing p_0, V_e also increases so that the ratio quickly decreases. The maximum value will probably be achieved at a low speed V_e.

In order to accomplish a maximum flight time (endurance), UAV need to possess maximum energy E_A and respectively weight of the battery m_{bat}, highly effective propulsors with thrust F_e and UAV with a maximum glade ratio K_e at maximum value of the ratio C_L^3/C_D^2.

4.2. Maximum range

A maximum range is achieved when the glade ratio K_R is maximum and:

(19) $$E_A = P_R t \; [Wh]$$

(20) $$R = V_c t = V_c \frac{E_A}{P_R} = V_c \frac{E_{RA}}{F_R V_c} = \frac{\bar{E}_A m_{bat}}{F_R} = \frac{\bar{E}_A \overline{m_{bat}} m_0}{F_R} = \frac{\bar{E}_A \overline{m_{bat}} m_0}{\frac{m_0 g}{K_R}} = \frac{\bar{E}_A \overline{m_{bat}} K_R}{g} \; [m].$$

The maximum range will be as high as the available energy $\bar{E}_A \overline{m_{bat}}$ and the glider ratio by cruise speed K_R are higher. For one aircraft, the higher the cruise speed V_c is, the shorter the flight time t is, and vice versa. When developing e-VTOL UAV, engineers can select a higher value of wing load p_0 and respectively a smaller wing area. In this case, a lower value of empty mass and respectively a higher value of the cruise speed can be achieved. When accepted that m_0= 25 kg, the battery weight and therefore the flight range can be increased. The maximum value of p_0 can be determined by structure considerations and payload limits.

When developing e-VTOL UAV with a long range, we have to accomplish the highest glade ratio and select a high value of the wing load p_0 that is much higher than at horizontally take-off and landing UAVs. Meanwhile, maximum $\overline{m_{empty}}$ can be achieved by using non-standard balancing scheme, for example “Joined wing” configuration. In most missions of e-VTOLUAV, a higher cruise speed means a higher productivity.

Conclusion

The maximum endurance of flight can be achieved at a low value of the maximum endurance speed V_e and respectively at a low value of p_0. While maximum range of flight can be achieved at a high cruise speed V_c and respectively at a high value of p_0.

It is difficult to reach an acceptable compromise in the presence of such contradictory requirements for the wing load for both modes of flight of e-VTOL UAV. A good solution can be achieved by using a module configuration and wing systems with a different wing load.

Acknowledgements

Set up under the project 9иф-02-4/28.11.2018 “Development of an innovative drone with long endurance and high altitude”, financed by the National Innovation Fund of Ministry of Economic of Republic of Bulgaria.

References

1. Zafirov, D. Joined Wings Thrust Vectored UAV Flight Envelope, AIAA Atmospheric Flight Mechanic Conference, 2010, Toronto, Ontario, Canada, 2–5 August 2010, Volume 1, Curran Associates, Inc., pp. 153–61, ISBN:978-1-61782-333-6.
2. Unmanned Aerial Vehicles Market – Growth, Trends, And Forecast (2019–2024). 2018, Mordor Intelligence, URL: https://www.mordorintelligence.com/industry-reports/uav-market
3. Zafirov, D. and H. Panajotov, UAV Joined-Wing Test Bed. In: Proceedings of 4th CEAS in Linköping, 2013, Sweden, pp. 516–524, mISBN 978-91-7519-519-3.
4. https://nari.arc.nasa.gov/sites/default/files/attachments/TVF_Roadmap_v5.2.xlsx.
5. Stojkov, O. and I. Kamburov. Matematicheski model na eksploataciJata na sistemite za upravlenie na aviacionnoto v"or"zhenie, Nauchni trudove “VNVVU G. Benkovski”, 1983, No. 44, Military Publishing House, Sofia. (in Bulgarian)
6. Stojkov, O. Kompleksna obrabotka na informacijata poluchavana ot razlichni datchici. Nauchni trudove “VNVVU G. Benkovski”, 1988, No. 51. (in Bulgarian)
7. Stojkov, O. Optiko elektronni sistemi. NVU “V. Levski”, fakultet “Aviacionen”, Jubilejna nauchna sesija, 2005, Vol. 1, D. Mitropolija, pp. 164–170, ISBN 954-713-071-4 (in Bulgarian)

ЕЛЕКТРИЧЕСКИ ВЕРТИКАЛНО ИЗЛИТАЩ И КАЦАЩ БЕЗПИЛОТЕН САМОЛЕТ ЗА ГОЛЯМА ПРОДЪЛЖИТЕЛНОСТ И ДАЛЕЧИНА НА ПОЛЕТА?

Д. Зафиров

Резюме

В статията се прави анализ на изискванията към вертикално излитащи и кацащи електрически безпилотни самолети, които да отговарят на изискванията на потребителите и са удобни за експлоатация при всякакви теренни условия. За много от изпълняваните мисии е важно да се осигури голяма продължителност на полета или голямо изминато разстояние. Разгледани са математически модели за двата случая и е установено, че изискванията за крилното натоварване са различни. За да се изпълнят те, се препоръча използването на вариант на безпилотен самолет, който да е модулен и да позволява използването на различни конфигурации и полезни товари, в зависимост от мисията, която ще се изпълнява.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e09

SECONDARY POWER SUPPLY SYSTEM FOR SPACECRAFT POTENTIAL MONITOR DP-1 AND DP-2, "OBSTANOVKA" PROJECT, INTERNATIONAL SPACE STATION

Pavlin Gramatikov, Roumen Nedkov, Georgi Stanev

Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: pgramatikov@space.bas.bg, rnedkov@space.bas.bg

Keywords: *Spacecraft, International Space Station, Secondary power supply system.*

Abstract

Plasma Wave Complex is a scientific instrumentation for wave parameters measurements in the ISS environment, and is implemented in the OBSTANOVKA experiment on board of Russian segment of ISS. The device Spacecraft Potential monitor "DP-1"and "DP-2" (one part of Plasma-Wave Complex) was developed in IKI BAS and measured the potential of the hull no more than 3 m from the surface of the ISS at range ±200 V; 0 ÷ 500 Hz. There are block and functional diagrams of the "DP" and the secondary power supply system, designed to supply the measuring probe, analogue and digital circuit boards. The secondary power supply system for the device "DP-1" and "DP-2" is discussed here in detail.

1. Introduction

The "Charge-ISS" Project from the experiment "Obstanovka" is designed to study the charging of large-scale spacecraft of the International Station (ISS) in interaction with the Earth's Ionosphere plasma. Devices "DP-1" and "DP-2" (Fig. 2) are designed for measurements the electric potential of the ISS with the help of spherical sensors based on carbon fiber. Plasma-wave processes are components part of the Space Weather. This approach is based on one of the contemporary ideas in physics – view on the plasma, including plasma in the outer space as to the dynamic medium with the charged particles and the wide spectrum of plasma wave motions and heterogeneities [1]. The "Obstanovka-1" stage will be carried out to provide a data of electromagnetic fields and of plasma-wave processes occurring in the ISS near surface zone (NSC) to study the plasma component factors of near-Earth space (NES) [3]. On the Fig. 1 [2], is given block-diagram of Plasma-Wave Complex (PWC): Combined wave sensor (CWS); Flux gate magnetometer (DFM); Langmuir probe (LP); Plasma discharge stimulator (SPP); Correlating Electron Spectrograph (CORES); Radio Frequency Analyzer (Scorpion); Signal Analyzer and Sampler (SAS3); Data Acquisition and Control Unit (DACU); Block of Storage of Telemetry

Information (BSTM), inside ISS. To study the plasma environment parameters, including electromagnetic conditions (EMC) in the vicinity of ISS, PWC will measure fluctuation and gradients of magnetic field, parameters of electrostatic and electromagnetic fields, density and temperature of thermal plasma, thermal electron flows in the ISS near zone and also electric potential of the ISS itself [2]. The PWC is implemented as two measuring blocks with nearly identical sensors. The distance between them (3–15 m) will enable us to carry out simultaneous point-to point measurements necessary for the measurements of DC electrical fields and small-scale gradients in magnetic field.

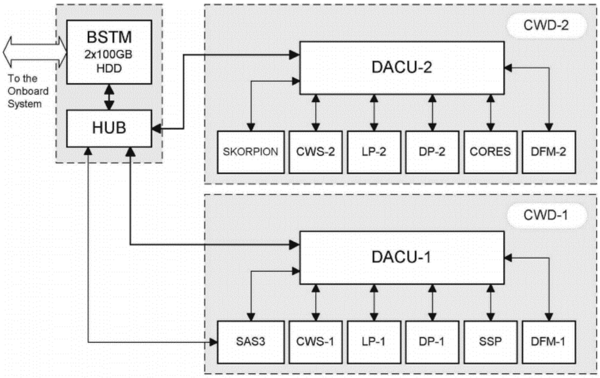

Fig. 1. Plasma-Wave Complex Block diagram [2]

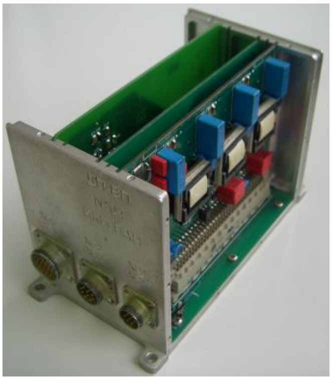

Fig. 2. Spacecraft potential monitor "DP-1"and "DP-2"

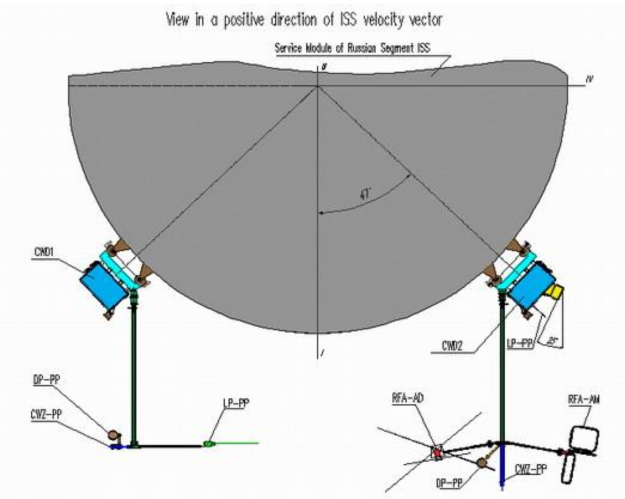

Fig. 3. Plasma-Wave Complex on the ISS board [6]

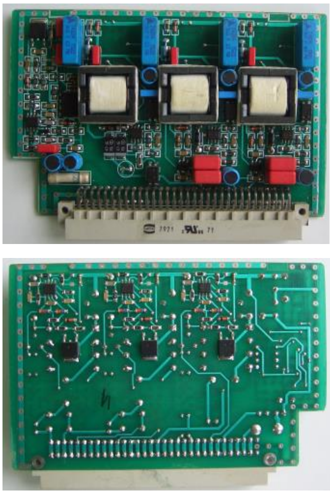

Fig. 4, 5. Secondary power supply system for "DP-1" and "DP-2"

The developed PWC scientific equipment is designed to measure in NES the following physical parameters: current parameters of thermal plasma (in two points); electrons and ions temperature, T_e, T_i; electrons and ions density, N_e, N_i; current electromagnetic parameters (in two points); DC electric and magnetic fields and currents; AC electric and magnetic fields and currents; current plasma potential and ISS potential; electrons spectra with energy range 0.01–10 keV; spectra of VLF electromagnetic fluctuations. For the study of discharge effects in NES plasma the PWC structure includes also the device for discharge stimulation [3].

2. Materials and methods

2.1. Device DP

The potential difference measurement between a probe and ISS body is the main scientific aim of the device DP. This allows us to study of the ISS electric charging processes and the time variation of the electric potential. The existence of two identical devices DP-1 and DP-2, which are mounted on every block CWD, allows us also to measure the spatial electric field in NSZ [3]. Each of them contains sensor – electronic Converter unit DP-PP (Primary Converter) and the electronic unit DP-SP (Secondary Converter). The device DP also provides an estimation of the contact layer resistance of the system plasma-probe. For Fig. 3 the designation DP-PP for the sensors of the device "DP" is accepted. Two devices "DP-1" and "DP-2" are mounted inside CWD1 and CWD2 respectively. The device DP is an electronic module which measures the potential difference in range ± 200 V, which is divided in two sub-ranges: ± 20 V and ± 200 V. The subranges are switched automatically. Device DP-SP (Fig. 2) consist from three boards: SPS-DP; analog processing and quantization board of input signals and microcomputer board. A 12-bit ADC (Fig. 3) provides potential difference measurements with resolution 10 mV (± 20 V) and 100 mV (± 200 V). The DP device measures and transmits information about the potential of the ISS body to the following information modes (measurements per second) [7]: fast 200; working: 100, 50, 20, and 12.5; base: 10, 8, 4, 2, and 1.25; slow: 1, 0.8, and 0.5; duty 0.33.

The probes of the device DP is the spherical collapsible structures with a diameter of 80 mm. Hemispheres are made of graphite and coated glass-carbon coating, which provides minimal variations in the value of the work electron output over the entire surface of the probes [7].

The DP instrument has passed functional tests and communication tests in the composition of the whole complex of scientific instruments in Russia and Hungary. The secondary source for powering a DP must meet all the requirements for scientific equipment for work on the outside of the ISS [5]. "DP-1" and "DP-2" instruments must undergo vacuum tests with simulated sun. The specific methodology shall be developed for each of the operating outside the ISS. The cutlery are tested in three modes:

– Permanent, with the inclusion of "sun" for 4.5 h;

– Periodically, tricikla with an imitation of the entry in the Shadow of the Earth (57 minutes "sun" and 36 minutes – shadow of the Earth);

– In the absence of a solar radiation, with an imitation of the Earth's radiation in the course of 4.5 h.

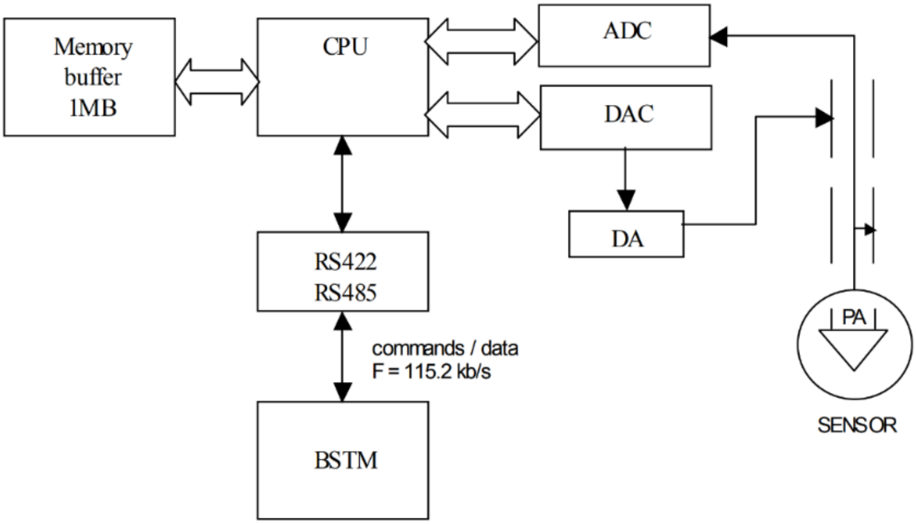

Fig. 6. Block diagram of Spacecraft potential monitor DP [3]

2.2. Secondary power supply system for device DP

The secondary power supply system has galvanic insulation between primary power lines and the secondary power lines. The connection with the main power supply is designed so that no instrument failure can influence other units. The instrument total power limit is ~ 4.7 W. The starting characteristics of the SPS-DP for the "DP" instrument were measured. In a Table 1 are given the electrical parameters of SPS-DP fly set (FS): FS-1, 2, 3, and 4, measured at maximum load (simulated by a resistive load exceeding the real load), idling and no secondary part. It is noticeable that with an increase in BN from 17 V to 36 V, the efficiency deteriorates by about 10 %. Functional scheme of SPS-DP for devices DP is given on the Fig. 7. The technical documentation for "DP" is compiled. The dimensions of the component of SPS-DP are 122 × 78 mm and the overall height of the components of the transformers side is 11 mm, Fig. 5.

Table 1. Parameters of SPS-DP, FS-1, 2, 3, and 4 for instrument "DP-1" and "DP-2"

SPS Fly Set 1, 2, 3, and 4	U-БМ [V]	17	19	23	29	34	36
FS-1, P_{out} = 2,1 W	efficiency [%]	68.13	68.24	67.56	63.21	59.05	57.28
FS-1, P_{out} = 2,1 W	power BN [mW]	3094	3089	3120	3335	3570	3680
FS-2, P_{out} = 2,1 W	efficiency [%]	68.13	68.49	67.89	63.76	59.05	57.44
FS-2, P_{out} = 2,1 W	power BN [mW]	3094	3078	3105	3306	3570	3670
FS-2, P_{out} = 2,1 W	efficiency [%]	67.39	67.65	67.39	63.76	59.05	57.41
FS-2, P_{out} = 2,1 W	power BN [mW]	3128	3116	3128	3306	3570	3672
FS-4, P_{out} = 2,1 W	efficiency [%]	67.83	68.49	68.40	64.33	59.62	57.98
FS-4, P_{out} = 2,1 W	power BN [mW]	3107.6	3078	3082	3277	3536	3636
FS-3, P_{out} = 0 mW	power BN [mW]	85	100.7	184	377	588.2	684
FS-4, P_{out} = 0 mW	power BN [mW]	85	100.7	184	377	588.2	684

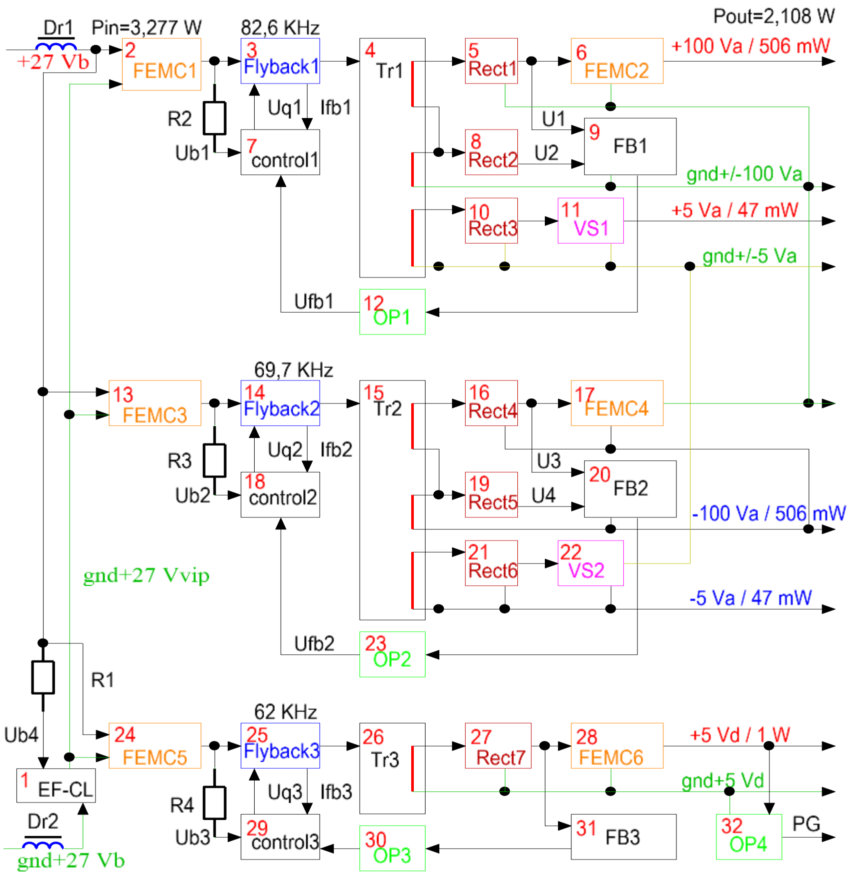

Fig. 7. Functional scheme of SPS-DP for Devices DP

The voltage of the boarding net (BN) shall be transmitted through chokes (Dr1 and Dr2) according to the technical requirements [5]. The EF-CL (1) unit has functions of electronic fuse (EF) and current limit (CL). The power supply for (1) from BN is done simply by R1. When the output current of (1) exceeds a certain limit, it is limited. If the duration of this condition exceeds a specified time (1) excludes the submission of BN to the SPS while retaining this state of „stand by” while BN does not restart.

The SPS-DP is made up of three reverse convertors. Accordingly, the first A is made up of the elements (2–12), the second B of (13–23) and the third C of (24–32).

Voltages "±100 Va" with their mass "gnd ±100 Va" serve to supply the Sensor; "±5 Va" with their mass "gnd ±5 Va" power up he ADC and DA, and the digital blocks (Fig. 6) are powered by "±5 Vd" with their mass "gnd ±5 Vd".

Converter A works as follows: 1 powered 2, R2 via Ub1 powered 7. After the start-up process of 7 appears the impulses Uq1 for normal operation of 3. The current Ifb1 of 3 is monitored dynamically for each period of Uq1. The Tr1 transformer serves as a galvanic insulation and with 5 and 6 generates the main voltage U1, the auxiliary voltage U2, and by 10 and 11 -stabilized voltage "+5 Va". For stabilization of "+100 Va", are used U1, U2, 9, 12 and Ufb1. Converter B works similarly to A. Converter C produces one output voltage "+ 5 Vd" and one galvanically bound signal PG for telemetry, for authentication of the working capacity.

For the device "DP" in BN are used circuit breakers type BSK5E5-32, for which there are electrical and time parameters for the starting current. They are shown in the Table. 2.

Table 2. Characteristics of the circuit breakers for the BN type BSK5E5-32

Io	[A]	1.5	2.5	4	6	8	10	14	18
To	[s]	0.05	0.15	0.20	0.30	0.40	0.60	0.80	1.00
Iz	[A]	0.5	0.7	1–1.5	2.5	3–3.5	4	5	7
Tpr	[s]	1	2	4	7	10	15	20	30

Time diagram the factory setting for БСК5Э5-32 are indicated on Fig. 8. Maximum values are indicated in orange and purple, and green is the permissible starting current. The following abbreviations are used:

- **Io**: the maximum value in the plus bus of the onboard network, in excess of which is unconditionally and immediately disconnected power;
- **To**: the maximum time for completion of transition processes after command switching device.
- **Iz**: the value of the overload current (for the time interval **Tpr**), in excess of which turns off the power of the device;

- **Tpr** – the time allowed for the occurrence of the overload current of the device after that triggers the switch to the positive bus of the battery.

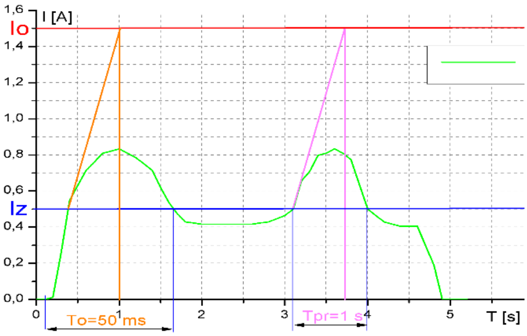

Fig. 8. Time diagram of БСК5Э5-32 from BN of ISS

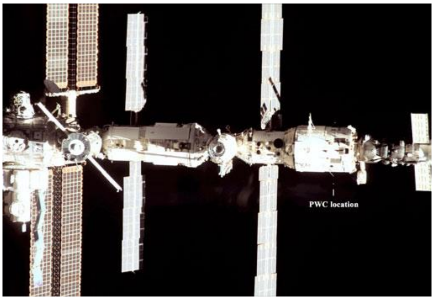

Fig. 9. PWC location of ISS [6].

3. Results

The exit of ISS cosmonauts Pavel Vladimirovich Vinogradov and Roman Romanenko in outer space took place for PWC equipment installation on the external surface of the Service Module of Russian Segment of ISS in April 2013; see Fig. 9 [6]. Power consumption of the device DP is not more than 2.2 W. Peak value of inrush current = 0.35 A for the BN in the range of 17–36 V that means that these values apply to the entire range. The time of starting the process $t_{пуск}$ = 3–5 mS. The value of the tripping current protection is **Iz** = 0.35 A. Protection of SPS-DP turns off the DP after time 70 mS [7].

4. Discussion

On April 23, 2014 was the first anniversary of functioning of scientific equipment in the frame of space experiment "Obstanovka-1 stage" on board the Russian segment of the ISS [6]. The data from experiments with DP devices allow recording of physical events and monitoring of the parameters of low temperature plasma near to the board of ISS. The DP system provides opportunities for monitoring and processing of the information obtained in the laboratory for the work period (4 years) [1].

References

1. Bachvarov, D., A. Boneva, B. Kirov, Y. Boneva, G. Stanev, and N. Baruh. Primary information pre-processing system for LP, DP devices – project "Obstanovka". In: Proceedings of the International Conference „Big Data, Knowledge and Control Systems Engineering" – BdKCSE'2014, November 5, 2014, Sofia, Bulgaria, pp. 65–74, URL: http://conference.ott-iict.bas.bg/wp-content/uploads/2014/01/BdKCSE2014_Pre.pdf
2. Klimov, S. I., V. E. Korepanov, Yu. V.Lissakov, O. V. Lapshinova, I. V. Sorokin, S. A. Belyaev, G. A. Stanev, K. Georgieva, B. Kirov, M. P. Gough, H. S. C. K. Alleyne, M. Balikhin, J. Lichtenberger, Cs. Ferencz, L. Bodnar, K. Szego, S. Szalai, J. Juchiewicz, H. Rothkaehl, and K. Stasiewicz, "Obstanovka" experiment onboard International Space Station for space weather research. URL: http://www.rmki.kfki.hu/~szalai/Publications/Klimov-ESTEC-2002-Ostanovka.pdf
3. Project OBSTANOVKA 1-st stage. URL: http://www.obstanovka.free.bg.
4. The project "Charge-ISS" is the Study of near-surface processes of polarization of a spacecraft of an extra-large size (of the International space station – ISS), 2006–2010. URL: http://www.ras.ru/presidium/documents/directionsp.aspx?ID=928925e5-f80e-462f-b441-ee25cf492833&print=1
5. International Space Station Program. NASA/RSA Joint Specifications/Standards Document for the ISS Russian Segment, SSP 50094, Revision A, NASA, Johnson Space Center Houston, Texas, March 10, 2000. 551 p. URL: http://everyspec.com/NASA/NASA-JSC/NASA-SSP-PUBS/SSP_50094A_29695/
6. Space experiment "OBSTANOVKA 1-st stage" on Russian Segment of ISS http://www.iki.rssi.ru/obstanovka/eng/participants.htm
7. Neychev, S., St. Klimov, A. Anev, N. Baruh, L. Belyakova, A. Belikova, A. Boneva, A. Buzekova, D. Buchvarov, G. Galev, V. Grushin, M. Zaharinova, I. Ivanov, O. Lapshina, R. Nedkov, D. Novikov, V. Myagkih, P. Gramatikov, V. Osipov, G. Stanev, D. Teodosiev, and L. Todorieva, Device for measurement of spacekraft potential of ISS – equipment and first results, In: Proceedings of 9th Scientific Conference with International Participation "SPACE, ECOLOGY, SAFETY" – SES 2013, 20–22 November 2013, Sofia, Bulgaria, p. 15–23. http://www.space.bas.bg/BG/magasin/SES/PROCEEDINGS%20SES%202013.pdf

ВТОРИЧНА ЗАХРАНВАЩА СИСТЕМА ЗА МОНИТОРИ НА КОСМИЧЕСКИ ПОТЕНЦИАЛ ДП-1 И ДП-2, ПРОЕКТ "ОБСТАНОВКА", МЕЖДУНАРОДНА КОСМИЧЕСКА СТАНЦИЯ

П. Граматиков, Р. Недков, Г. Станев

Резюме

В статията е представен български прибор, предназначен за изследване на зареждането на Руския сегмент на Международната космическа станция (МКС). Двата еднакви прибора (ДП1 и ДП2) са част от приборите на блоковете на Плазменовълнов комплекс 1 и 2. ДП1 и ДП2 служат и за изследване динамиката на потенциала на МКС в зависимост от: слънчевата и геомагнитната активност (част от програмата "Космическо време"); дейността на космонавтите по време на работата им на повърхността на станцията; корекциите на орбитата на МКС и скачването и разкачването на космическите кораби с МКС. ДП1 и ДП2 измерват потенциала на корпуса не повече от 3 m от повърхността на МКС. Пределните диапазони на измерване са: диапазон ±200 V; честотен диапазон от 0 до 500 Hz и праг на чувствителност на потенциала 3.125 mV. Представени са блокови и функционални схеми на прибора ДП и вторичната електрозахранваща система, предназначена за захранването на измервателната сонда, аналоговата и цифровата платки. Представени са три снимки на изработената в България апаратура.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e10

DETERMINING THE BALLISTIC CHARACTERISTICS OF SPACE PENETRATOR

Stoyko Stoykov

Aviation Faculty, National Military University, Bulgaria
e-mail: sostoykov@nvu.bg

Keywords: *Planet investigation, Penetrator, Modeling the movement, Aerodynamic tube.*

Abstract

To model the movement of an aviation penetrator, it is necessary to know the coefficient of the drag and the coefficient of the lift force. The article presents a method of calculating them using the geometric dimensions of the penetrator. The obtained values of the coefficients are compared with those obtained when blowing a penetrator in the aerodynamic tube. By the sustainability criterion is determines the degree of damping of the penetrator. The results of modeling the movement of the penetrator show, that the mathematical model of motion can be used to solve the task of targeting.

1. Introduction

The mathematical modeling of a penetrator requires information on the drag coefficient and the lift force. The article offers a method of calculating them using the geometric dimensions of the penetrator.

2. Results

The test is conducted for a penetrator with the following characteristics:

- Θ = 21.39 s, characteristic fall time;
- $m_б$ = 64 kg, mass; $d_б$ = 0.203 m; $S_б = \frac{\pi d_б^2}{4}$ = 0.0324 m^2; $L_к$ = 0.835 m;
- $H_{ст}$ = 0.397 m; $H_к$ = 0.40 m; $D_{ст}$ = 0.205 m.

Ballistic coefficient "c" is determined by form. [3, 7]:

(1) $$c = (\Theta - a)\frac{1}{к} = 1.4649,$$

where a, k are coefficients (a = 20.202; k = 0.811).

The coefficient of form *i* is determined by form [2]:

(2) $i = \frac{cm_б}{d_б^2} 10^{-3}$ = 2.275.

For the standard drag coefficient, the Siachi law C_{xe} (M = 0) = 0.255 [7] is used. The coefficient of resistance C_{xb} is determined by formula [4] and for aviation penetrator it equals:

(3) $C_{xб}$ = 0.5801.

Through the analytical formula [4]

(4) $C_{xб} = 2[C_{xб}^0 + 0,0052(l_к - 1) + Ah_{ст} + B_1(d_{ст} - 1,3)]$,

the impedance coefficient C_{xba} is determined.
The relative dimensions of the penetrator are:

$l_к = \frac{L_к}{d_б}$ = 4.1133; $h_{ст} = \frac{H_{ст}}{d_б}$ = 1.9557; $d_{ст} = \frac{D_{ст}}{d_б}$ = 1.0099;
$h_к = \frac{H_к}{d_б}$ = 1.9704,

The values $C_{xб}^0$ and A are determined by [5, 7] and the following values are taken:

$C_{xб}^0$ = 0.053; A = 0.0646.

The coefficient B_1 is determined by form. [4]:

$B_1 = -0,0274 h_{ст} + 0,0319$ = –0.0209.

The front of the penetrator has a flat shape, i.e., $h_g \approx 0$, then the calculated value of C_{xb} is increased by 0.2 [4]. Since the tailpiece of the stabilizer has feathers and two rings, calculations are made for a box stabilizer.
For the coefficient of drag impulse C_{xba} we obtain:

(5) $C_{xба} = 2[C_{xб}^0 + 0,0052(l_к - 1) + Ah_{ст} + B_1(d_{ст} - 1,3)] + 0,2 = 0,6032$.

When blowing a model of a aviation penetration at M = 0 for the coefficient of impedance C_{xb0}, the following result is obtained:

(6) C_{xbo} = 0.5701.

The values of the drag coefficients C_{xb} and C_{xba} are close to the value of C_{xbo} determined by blowing the model.

This indicates that the proposed methods using the reference drag coefficient and using the geometric dimensions of the penetrator can be used to calculate the elements of the penetrator trajectory.

As a result of blowing the aviation penetrator pattern at different angles of attack, the following results for the coefficient C_{xb} (α) of drag resistance (Table 1 and Figure 1) are obtained.

Table 1. Dependency of $C_{xб}(\alpha)$

$\alpha_б$, deg.	**0**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**
$C_{xб}$	0.5701	0.5701	0.6185	0.7653	1.1172	1.762	2.7876	4.2819	6.3326

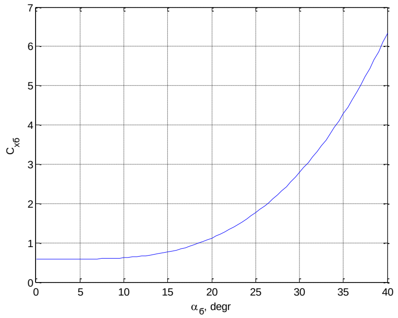

Fig. 1. Relevance of the coefficient Cxb (αb) of the impedance of the angle of attack αb

Using the Saichi law as a reference law for the change of the resistance and the results of the Table 2, the dependence of C_{xb} (M, a) (Fig. 2) is obtained. For the conditions under consideration it is assumed that the coefficient of the form is constant.

Table 2. Dependency of $C_{xб}(M, \alpha)$

$C_{xб}(M,\alpha)$	**M = 0**	**0,2**	**0,4**	**0,6**	**0,8**	**1**
$\alpha = 0^0$	0.5701	0.5701	0.5824	0.5892	0.6484	1.2422
10^0	0.6185	0.6185	0.6209	0.6282	0.6913	1.3243
20^0	1.1172	1.1172	1.1216	1.1347	1.2486	2.3921
30^0	2.7876	2.7876	2.7985	2.8313	3.1156	5.9687
40^0	6.3326	6.3326	6.3574	6.4319	7.0776	13.5592

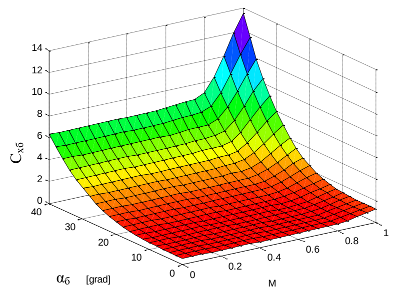

Fig. 2. Dependency of $C_{xб}(M,\alpha)$

For the coefficient Lift Force of the formula [6, 7], its values for different angles of attack were calculated (Table 3, Figure 3).

Table 3. Dependency of $C_{yб}$ от $\alpha_б$

$\alpha_б$, deg	**0**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**
$C_{yб}$	0	0.4740	0.8958	1.2656	1.5833	1.8490	2.0625	2.2240	2.3333

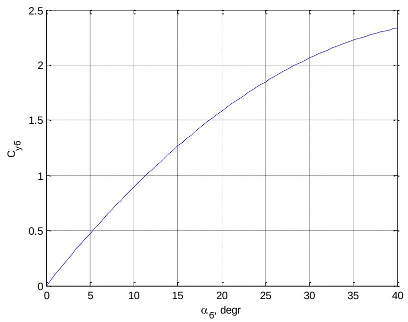

Fig. 3. Dependency of the coefficient $C_{yб}$ от $\alpha_б$

Using the sustainability criterion [4], the degree of damping of fluctuations is determined:

$$K(S) = 0.2885,$$

which satisfies the condition of sustainability.

As a result of the mathematical modeling of the aviation penetrator movement under different start conditions, the deceleration time of the penetrator attack angle αb, the coefficients of: the drag resistance C_{xb}, the lift force C_{yb} and the moment m_z (Figs. 4–10).

When solving the penetrator motion equations for conditions $\lambda = 0^o$, $H = 500$ m, $V = 180$ m/s, $\alpha_0 = 4^o$ the oscillation damping time $t = 0.82$ s ($\alpha_b = 0.01^o$), (Fig. 4). The C_{xb} coefficient of the drag impedance changes insignificantly (from 0.5895 to 0.5845), (Fig. 5).

The coefficient of Lift C_{yb} and the coefficient m_z of the moment diminish analogously, as the angle of attack (Figs. 6, 7).

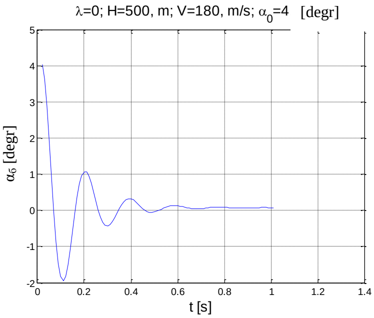

Fig. 4. Dependence of (αb) from time (t)

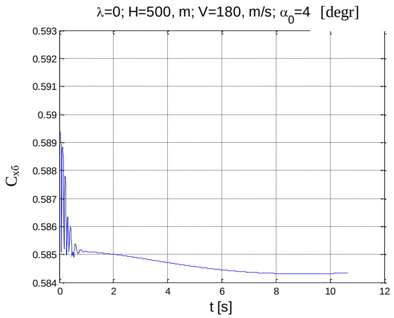

Fig. 5. Dependence of $C_{xб}$ from time (t)

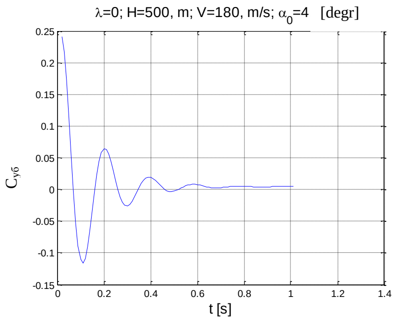

Fig. 6. Dependence of $C_{yб}$ from time (t)

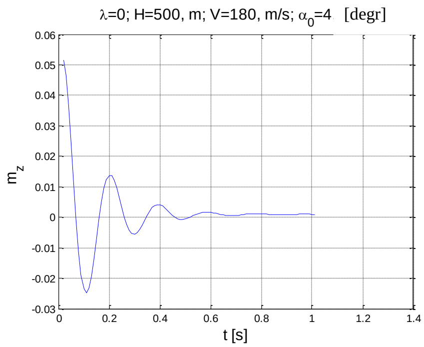

Fig. 7. Dependence of m_z from time (t)

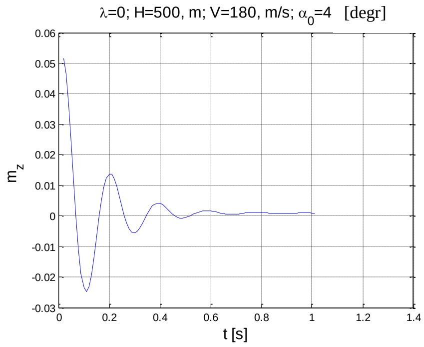

Fig. 8. Dependence of m_z from time (t)

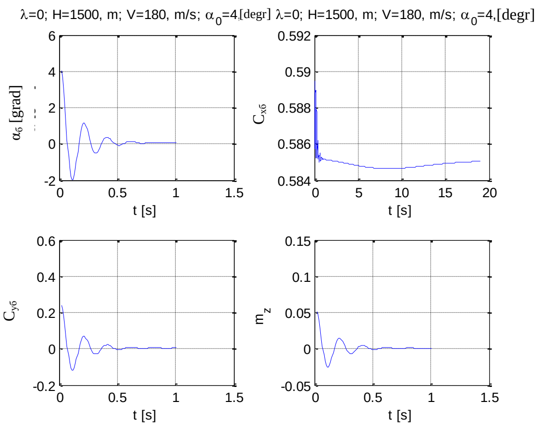

Fig. 9. Dependence of $\alpha_б$, $C_{xб}$, $C_{yб}$ u m_z from time (t)

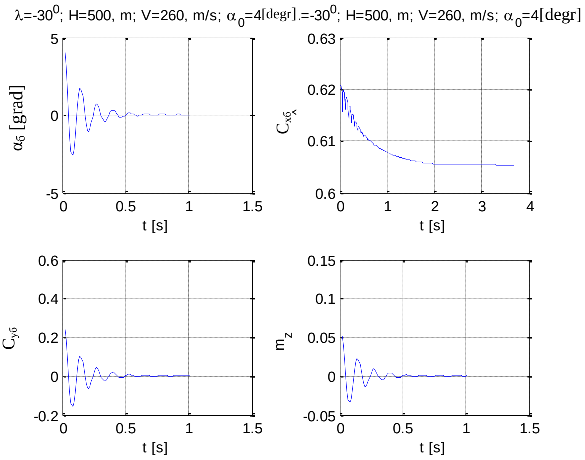

Fig. 10. Dependence of $\alpha_б$, $C_{xб}$, $C_{yб}$ и m_z from time (t)

3. Conclusions

The results of the mathematical modeling of the movement of the aviation penetrator (shown in the above figures) lead to the following conclusions:

1. As the penetrator starts up, the damping time of αb decreases and the frequency of oscillations increases;
2. By increasing the initial attack angle α0 of the bomb, the C_{xb} coefficient of the resistance of the penetrator changes insignificantly;
3. The character of the change of the coefficients C_{yb}, m_z is the same as the angle of attack α_b;
4. With an increase in the angle of latency λ, the decay time of ab decreases;
5. The damping time of αb does not depend on the height of the penetrator.

The results obtained show that the aviation penetrator pattern created can be used to solve the task of targeting.

References

1. Atanasov, M. A., V"zmozhnosti za reshavane na zadachata na pricelvane pri bombopuskane po "glova skorost, Dolna Mitropolija, PhD thesis, NVU „Vasil Levski”, 2006. (in Bulgarian)
2. Atanasov, M. A., Tochnost na strelbata i bombopuskaneto s izpolzvane na aviacionen pricelen kompleks s"s sledjashta sistema, NVU „Vasil Levski”, 2018. (in Bulgarian)
3. Baranov, V. and G. Mardirosjan. Svjazannaja zadacha optimizacii parametrov penetratorov dlJa mezhplanetnыh issledovanij. Sb. nauchnыh trudov Tulyskogo gosud. universiteta, Tula, 1996, pp. 35–39. (in Russian)

4. Pokrovskij, G. I. Sassaparely V. I. et al., Kurs aviacionnыh bomb, M., VVIA „N. E. Zhukovskogo“, 1950. (in Russian)
5. Mardirosjan, G., D. Jordanov, L. Kraleva, and D. Danov. Penetrator za ekologichni izsledvanija. In: Proceedings. "30 godini organizirani kosmicheski izsledvanija v B"lgarija", SRI-BAS, Sofia, 1999, pp. 385–389. (in Bulgarian)
6. Mardirossian, G., D. Danov. Preliminary analysis or the ballistic parameters of a penetrator for ecological studies. Aerospace Res. in Bulgaria, 2001, 16, pp. 89–96.
7. Stojkov, O. S., Metodi i tochnost na reshavane na zadachata na pricelvane pri bombopuskane, Dolna Mitropolija, 2010, ISBN 978-954-713-100-2. (in Bulgarian)

РЕЗУЛТАТИ ОТ МОДЕЛИРАНЕ НА ДВИЖЕНИЕТО НА КОСМИЧЕСКИ ПЕНЕТРАТОР

С. Стойков

Резюме

За моделиране на движението на космически пенетратор е необходимо да се знае коефициента на челно съпротивление и коефициента на подемната сила. В статията се предлага метод за тяхното изчисляване чрез геометричните размери на пенетратора. Получените стойности на коефициентите се сравняват с тези получени при обтичане на модела на пенетратора в аеродинамична тръба. Чрез критерия за устойчивост е определена степента на затихване на колебанията на пенетратора. Получените резултати от моделиране на движението на пенетратора показват, че математическият модел за движение може да се използва за решаване на задачата на прицелване.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e11

APPLYING POTENTIAL-BASED PANEL METHOD FOR STEADY FLOW ANALYSIS ACROSS A WING WITH FINITE SPAN

Konstantin Metodiev

Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: komet@space.bas.bg

Keywords: *Panel method, Iterative scheme, Finite span wing, Potential flow.*

Abstract

In the paper hereby, an incompressible irrotational steady flow across a submerged body with finite dimensions will be studied. For this purpose, it is necessary to solve Laplace's differential equation about a potential function in order to obtain the conservative velocity vector field. A general solution to the problem utilizing the Green identity implies the double layer potential function at an arbitrary point not belonging to the boundary surface. The potential is expressed by source/sink and doublet singularities distributed over the body surface and a wake attached to the trailing edge. The wake ensures that the Kutta condition is fulfilled. The submerged body geometry is approximated further by quadrilateral panels in order to compute the surface integrals for each panel exactly. To form a linear non-homogenous algebraic system, it is essentially to compute each panel influence to a collocation point of interest. The obtained coefficient matrix is diagonally dominant. The system is solved iteratively by means of the Gauss-Seidel method.

The goal is development of a non-proprietary source code in order to work out a solution to the stated problem. The developed source code is authentic. Auxiliary libraries have not been used. Validation case and numerical results are depicted and discussed in the paper.

1. Introduction

The proposed approach towards working out a solution to the stated fluid problem utilizes the so-called double layer potential method applied to the Laplace's equation. In this study case, the flow is assumed irrotational and incompressible. This is a relatively old method which has been thoroughly studied and many solution codes have been developed as well. Nevertheless, one advantage of the method provokes development of the current study case: the method is fast and applicable to complex geometries of thick bodies generating lift. What is more, by solving the Laplace's equation the velocity vector field might be found out prior to using equations of motion, such as Euler or Navier-Stokes.

The presented study emphasizes on applicability of iterative schemes for working out a solution to a non-homogenous linear algebraic system relevant to the stated fluid problem. In addition, authentic source code development in C is yet another project goal. To achieve it, Katz and Plotkin's textbook, [1], was extensively used by the author as a guide throughout the presented study case.

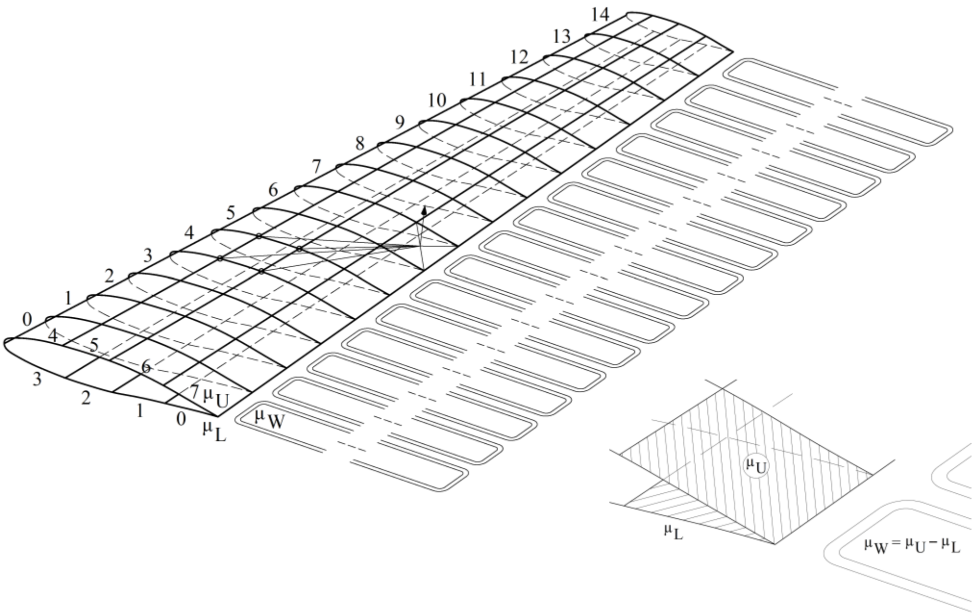

Fig. 1. Wing and a wake shed by the trailing edge. Lower right corner: the wake panel strength computation.

2. Problem statement and solution

For an incompressible irrotational flow the continuity equation

(1) $$\nabla \mathbf{u} = 0,$$

takes the form:

(2) $$\nabla^2 \Phi = 0,$$

where and Φ is potential function of a conservative velocity field **u** = gradΦ. The boundary condition at surface of a submerged body implies that normal velocity component vanishes:

(3) $$\nabla \Phi . \mathbf{n} = 0.$$

The potential vector gradΦ is measured in body frame of reference. In addition, a disturbance created by the body decays at infinity **r**→∞, i.e.

(4) $$\lim_{r\to\infty}\left(\nabla\Phi-\nabla\Phi_{\infty}\right)=0,$$

where gradΦ_∞ is a vector due to the far field potential. A general solution to the problem stated by formulae 1–4, might be worked out considering the Green identity. In this way, the potential function at an arbitrary point P not belonging to the boundary surface is computed by

(5) $$\Phi(P)=-\frac{1}{4\pi}\int_{Body}\left[\sigma\left(\frac{1}{r}\right)-\mu\mathbf{n}.\nabla\left(\frac{1}{r}\right)\right]dS+\frac{1}{4\pi}\int_{Wake}\left[\mu\mathbf{n}.\nabla\left(\frac{1}{r}\right)\right]dS+\Phi_{\infty}(P),$$

where σ (source/sink) and μ (doublet) are flow singularities strengths, **r** is distance from point P to the surface (body, wake, etc.). The surface integrals are taken over the body and a wake model, Fig. 1. The wake is assumed to be thin, so that the dot product **n**.gradΦ is continuous across it. This means that the wake cannot support fluid-dynamic loads, [1, p. 46]. In order to find the potential function Φ, a unique combination of sources and doublets distribution on the surface must be known in advance, [1, p. 47]. The solution (5) is also denoted as a double layer potential. The integral is computed over a double-sided surface and the normal vector **n** points inwards. According to Lamb, [2, p. 40], the impermeability condition (3) results in a constant inner potential

(6) $$\nabla\Phi.\mathbf{n}=0\Rightarrow\Phi_i=const,$$

which implies that the current lines are not allowed to enter or leave the inner region – nor they are contained within. If the above equality holds, there can be no fluid motion inside the body. Assuming that the inner potential can be set to Φ_i = Φ_∞ = const, equation (5) might be rewritten as follows:

(7) $$-\frac{1}{4\pi}\int_{Body}\left[\sigma\left(\frac{1}{r}\right)-\mu\mathbf{n}.\nabla\left(\frac{1}{r}\right)\right]dS+\frac{1}{4\pi}\int_{Wake}\left[\mu\mathbf{n}.\nabla\left(\frac{1}{r}\right)\right]dS=0.$$

A numerical solution to equation (7) is worked out in the current study.

The boundary conditions (BCs) might either determine the zero normal velocity component (direct, Neumann BCs, also (3)) or specify the velocity potential itself (indirect, Dirichlet BCs) at the boundary surface. Equation (7) interprets the Dirichlet's boundary condition. In addition, it is customary to assign following quantity to the source strength:

(8) $$\sigma=-\mathbf{n}\left(\mathbf{V}_{linear}+\mathbf{\Omega}\times\mathbf{r}-\mathbf{V}_{free\ stream}\right),$$

where $\mathbf{V}_{\text{linear}}$ is body linear velocity, **Ω** is body angular velocity, and $\mathbf{V}_{\text{free stream}}$ is the free stream velocity.

Further problem refinement is required to describe flow over a thick body with sharp trailing edge generating a lift. In order to hold the rear stagnation point at the trailing edge, sufficient amount of circulation must be created while the body is moving through the fluid. This statement is yet another interpretation of the Kutta condition implying that a jump in the velocity potential exists in the vicinity of the trailing edge and the velocity there is finite.

The wake strength at the trailing edge is determined by setting to zero the vortex element strength located at the trailing edge. Then, the vortex distribution might be regained by an equivalent doublet distribution, [1, p. 250]

(9) $$\gamma_{T.E.} = -\nabla\mu = 0,$$

which condition is fulfilled if

(10) $$\mu_{T.E.} = const = \mu_W, \quad \mu_W = \mu_U - \mu_L,$$

where indices W, U, L denote wake, upper, and lower surface respectively, Fig. 1.

The computational algorithm might also be seen in Fig. 1. The influence of singularities distributed onto the body and the wake is computed for each collocation point, which is placed at the panel centroid. In the example depicted in Fig. 1, the collocation point is placed at panel I = 7, j = 6 and the influence panel is I = 5, j = 4. For each collocation, point equation (7) might be written as follows:

(11) $$\sum_{k=1}^{N}\frac{1}{4\pi}\int_{Body}\mu\mathbf{n}.\nabla\left(\frac{1}{r}\right)dS + \sum_{l=1}^{N_W}\frac{1}{4\pi}\int_{Wake}\mu\mathbf{n}.\nabla\left(\frac{1}{r}\right)dS - \sum_{k=1}^{N}\frac{1}{4\pi}\int_{Body}\sigma\left(\frac{1}{r}\right)dS = 0,$$

where singularities μ and σ accept unit *constant* strength. The summation is evaluated for all panels discretizing the body and the wake. To abbreviate the amount of writing further, following symbols are adopted: for a doublet panel

(12) $$\frac{1}{4\pi}\int_{Panel}\mathbf{n}.\nabla\left(\frac{1}{r}\right)dS \equiv C$$

and for a source element

(13) $$-\frac{1}{4\pi}\int_{Panel}\left(\frac{1}{r}\right)dS \equiv B.$$

The integrals in (12) and (13) solely depend on the panel geometry. Having computed all panel influences onto all collocation points, the following non-homogenous linear algebraic system is obtained:

(14) $$\sum_{k=1}^{N}C_k\mu_k + \sum_{w=1}^{N_{Wake}}C_w\mu_w + \sum_{k=1}^{N}B_k\sigma_k = 0,$$

where the source strength σ is known at this stage, (8). Equation (14) can be simplified further by computing the wake doublet in terms of the unknown surface doublet, Fig. 1, (also see (10)):

(15) $$\mu_W = \mu_U - \mu_L \ .$$

Then, the influence coefficient takes the form:

(16) $$C_W \mu_W = C_W \left(\mu_U + \mu_L \right) \ .$$

Therefore, the first two additives in (14) might be grouped which yields the following expression:

(17) $$\sum_{k=1}^{N} A_k \mu_k = -\sum_{k=1}^{N} B_k \sigma_k \ ,$$

where the influence coefficient is computed depending on whether the influence panel is at the trailing edge or not:

(18) $$\begin{array}{ll} A_k = C_k & \textit{panel is not at trailing edge} \\ A_k = C_k \pm C_w & \textit{panel is at trailing edge} \end{array}$$

The expanded form of system (17) is as follows:

(19) $$\begin{Vmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{Vmatrix} \begin{Vmatrix} \mu_1 \\ \vdots \\ \mu_N \end{Vmatrix} = -\begin{Vmatrix} b_{11} & \cdots & b_{1N} \\ \vdots & \vdots & \vdots \\ b_{N1} & \cdots & b_{NN} \end{Vmatrix} \begin{Vmatrix} \sigma_1 \\ \vdots \\ \sigma_N \end{Vmatrix} \ ,$$

where coefficients a_{ij} and b_{ij} are computed according to generic formulae (12) and (13). For particular case of quadrilateral source and doublet with constant strength, formulae derived by Hess and Smith, [3], are used:

- Source

(20) $$b = -\frac{1}{4\pi} \sum_{edges} \left\{ \begin{array}{l} \dfrac{(x - x_a)(y_b - y_a) - (y - y_a)(x_b - x_a)}{d_{ab}} \ln \dfrac{r_a + r_b + d_{ab}}{r_a + r_b - d_{ab}} - \\ z \left[\operatorname{atan}\left(\dfrac{m_{ab} e_a - h_a}{z r_a} \right) - \operatorname{atan}\left(\dfrac{m_{ab} e_b - h_b}{z r_b} \right) \right] \end{array} \right\}$$

- Doublet

(21) $$a = \frac{1}{4\pi} \sum_{edges} \left[\operatorname{atan}\left(\frac{m_{ab} e_a - h_a}{z r_a} \right) - \operatorname{atan}\left(\frac{m_{ab} e_b - h_b}{z r_b} \right) \right] \ .$$

In the equation above, following abbreviations are used:

$$d_{ab}=\sqrt{\left(x_b-x_a\right)^2+\left(y_b-y_a\right)^2}\quad r_i=\sqrt{\left(x-x_i\right)^2+\left(y-y_i\right)^2+z^2}$$

(22) $$m_{ab}=\frac{y_b-y_a}{x_b-x_a},\quad e_i=\left(x-x_i\right)^2+z^2,\quad h_i=\left(x-x_i\right)\left(y-y_i\right)$$

$$i=a,b$$

where x, y, and z are collocation point coordinates and x_a, y_a, z_a, x_b, y_b, and z_b are panel corner points coordinates, indices a and b denote panel corner points belonging to a same panel edge. All quantities are computed in *local (panel)* coordinate system, Fig. 3. Hence, preliminary coordinate transformations must be made.

After computing the influence coefficients, a non-homogenous linear algebraic system is obtained in terms of doublet μ distribution on the wing surface. The system is said to be strictly diagonal dominant if following requirement is met:

(23) $$\left|a_{ii}\right|>\sum_{\substack{j=1\\ j\neq i}}^{n}\left|a_{ij}\right| \ .$$

In other words, the absolute value of each main diagonal element must be greater than sum of absolute values of the remaining elements in the current row respectively.

If the requirement (23) is met, then the following stationary iterative method

(24) $$x_i^k=\frac{1}{a_{ii}}\left(b_{ii}-\sum_{j=1}^{i-1}a_{ij}x_j^k-\sum_{j=i+1}^{n}a_{ij}x_j^{k-1}\right)\quad i=1,2,\ldots,n\quad k=1,2,3,\ldots$$

for solving system (17) is said to converge unconditionally. Method (24) is named after Gauss and Seidel. The formula (24) is a modification of the widely known Jacobi method:

(25) $$x_i^k=\frac{1}{a_{ii}}\left(b_{ii}-\sum_{\substack{j=1\\ j\neq i}}^{n}a_{ij}x_j^{k-1}\right)\quad i=1,2,\ldots,n\quad k=1,2,3,\ldots$$

The convergence criterion used in the algorithm is the relative difference

(26) $$\max\left|\mathbf{x}^k - \mathbf{x}^{k-1}\right| / \left|\mathbf{x}^k\right| < 10^{-3}.$$

Both iterative schemes (24) and (25) require initial guess for the vector **x**.

```
#ifndef DEFS_H
#define DEFS_H

#define PI 4. * atan(1.)
#define I 20 // number of panels
#define J 80 // number of panels
#define IX(i, j) (i) * J + j // double indexing notation for one-dim array
#define Om {0., 0., 0.} // wing angular velocity, s^-1
#define Vl {0., 0., 0.} // wing linear velocity, m/s
#define Ve {1., 0., 0.0875} // free stream velocity, m/s

typedef double real;

typedef struct _panel {
        real *x_, *y_, *z_; // corner points in panel coordinates, [4]
each
        real *x, *y, *z; // corner points in global coordinates, [4] each
        real *n, *b, *t, S, *p; // normal[3], binormal[3], tangent[3], …
        real *Va; // apparent velocity
        real dsig, dmu; // solution goes here
        real cp; // static pressure coefficient
} myPanel;

int geom(myPanel *foo, char *type);
int paraView(real *x, real *y, real *z, real *scalars0, real *vec0, real
             *vec1, real *vec2, char *fileName);
myPanel* createPanels(int N);
int deletePanels(myPanel *foo, int N);
real influenceDueToSource(myPanel *foo, real x, real y, real z);
real influenceDueToDoublet(myPanel *foo, real x, real y, real z);
real* solveLS_GS(int N, real *a, real *b);
int do_forces(myPanel *foo);

#endif // DEFS_H
```

Fig. 2. Source code header file

In Fig. 2, the source header file is shown. Each panel is represented by a structure _panel containing panel geometry, apparent velocity (8), and a few solution quantities, namely singularities strengths and static pressure coefficient. The panel geometry includes corner points coordinates expressed in both global and local (panel) coordinate system. The latter is formed by normal, tangent, and binormal unit vectors as it are shown in Fig. 3 in case of circular cylinder. In addition, the header contains following function prototypes. Function “createPanels” allocates memory for specified number of panels and pointers inside the structure. It returns a pointer to first panel inside the so formed one-dimensional array of structures. Function “deletePanels” does the opposite. Function “geom” calculates all necessary geometric parameters relevant to each panel. The remaining function prototypes are self-explanatory and easily understandable. The functions "influenceDueToSource" and "influenceDueToDoublet" compute influences due to singularities at an arbitrary collocation point. The function "solve_GS" solves a linear algebraic system (17) iteratively. It returns a pointer to a solution vector allocated within the function body. What does function "paraView" is arranging the results to meet the "vtk" file requirements [4] and storing them onto the hard drive so that the user can visualize the results by means of a third-party viewer. One dimensional arrays are solely used in the code and accessed by a two-dimensional macro IX(i, j), Fig. 2.

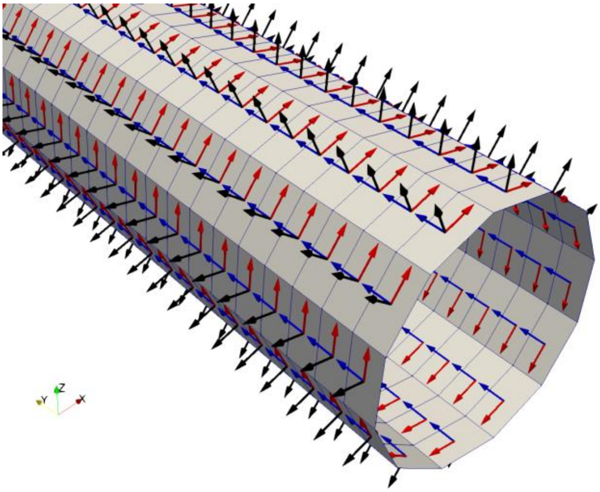

Fig. 3. Local coordinate system for each panel, black – normal, red – tangent, and blue – binormal stored at panel's centroid

3. Results

A circular cylinder has been used to validate the developed source code. The static pressure coefficient distribution is visible in Fig. 4. Both front and rear stagnation areas are clearly visible.

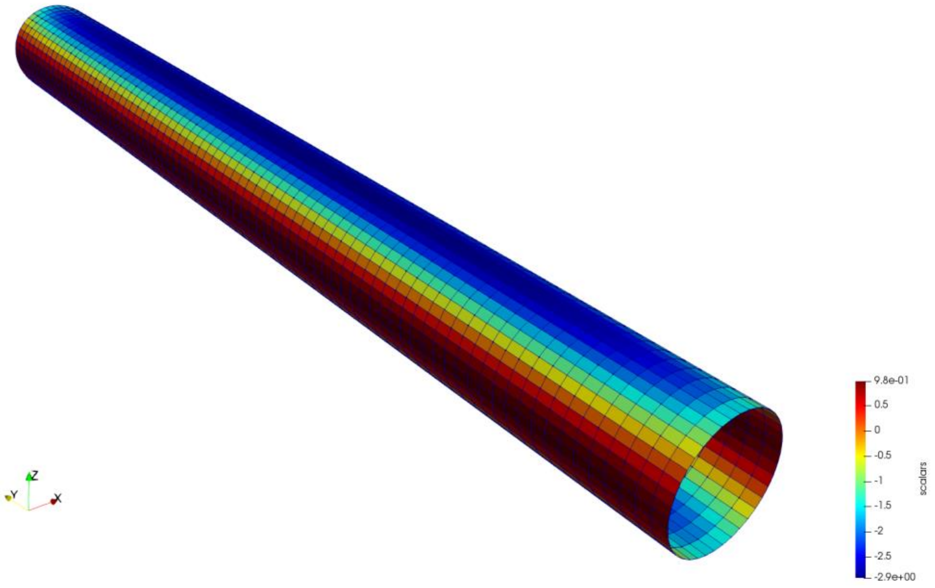

Fig. 4. Static pressure coefficient distribution over a cylinder, α = 0 deg, 40 × 80 panels

-0.498	0.001	0.000	0.000	0.000	0.000	-0.054	-0.001	-0.000	-0.000	-0.000	-0.000	-0.039
0.000	-0.498	0.001	0.000	0.000	0.000	-0.001	-0.054	-0.001	-0.000	-0.000	-0.000	-0.002
0.000	0.000	-0.498	0.001	0.000	0.000	-0.000	-0.001	-0.054	-0.001	-0.000	-0.000	-0.000
0.000	0.000	0.000	-0.498	0.001	0.000	-0.000	-0.000	-0.001	-0.054	-0.001	-0.000	-0.000
0.000	0.000	0.000	0.000	-0.498	0.001	-0.000	-0.000	-0.000	-0.001	-0.054	-0.001	-0.000
0.000	0.000	0.000	0.000	0.000	-0.498	-0.000	-0.000	-0.000	-0.000	-0.001	-0.054	-0.000
-0.051	0.002	0.000	0.000	0.000	0.000	-0.500	-0.000	-0.000	-0.000	-0.000	-0.000	-0.054
-0.000	-0.051	0.002	0.000	0.000	0.000	-0.000	-0.500	-0.000	-0.000	-0.000	-0.000	-0.001
0.000	-0.000	-0.051	0.002	0.000	0.000	-0.000	-0.000	-0.500	-0.000	-0.000	-0.000	-0.000
0.000	0.000	-0.000	-0.051	0.002	0.000	-0.000	-0.000	-0.000	-0.500	-0.000	-0.000	-0.000
0.000	0.000	0.000	-0.000	-0.051	0.002	-0.000	-0.000	-0.000	-0.000	-0.500	-0.000	-0.000
0.000	0.000	0.000	0.000	-0.000	-0.051	-0.000	-0.000	-0.000	-0.000	-0.000	-0.500	-0.000
-0.036	-0.000	0.000	0.000	0.000	0.000	-0.054	-0.001	-0.000	-0.000	-0.000	-0.000	-0.500
-0.002	-0.036	-0.000	0.000	0.000	0.000	-0.001	-0.054	-0.001	-0.000	-0.000	-0.000	-0.000
-0.000	-0.002	-0.036	-0.000	0.000	0.000	-0.000	-0.001	-0.054	-0.001	-0.000	-0.000	-0.000
0.000	-0.000	-0.002	-0.036	-0.000	0.000	-0.000	-0.000	-0.001	-0.054	-0.001	-0.000	-0.000
0.000	0.000	-0.000	-0.002	-0.036	-0.000	-0.000	-0.000	-0.000	-0.001	-0.054	-0.001	-0.000
0.000	0.000	0.000	-0.000	-0.002	-0.036	-0.000	-0.000	-0.000	-0.000	-0.001	-0.054	-0.000
-0.032	-0.002	-0.000	0.000	0.000	0.000	-0.039	-0.002	-0.000	-0.000	-0.000	-0.000	-0.054
-0.004	-0.032	-0.002	-0.000	0.000	0.000	-0.002	-0.039	-0.002	-0.000	-0.000	-0.000	-0.001
-0.000	-0.004	-0.032	-0.002	-0.000	0.000	-0.000	-0.002	-0.039	-0.002	-0.000	-0.000	-0.000
-0.000	-0.000	-0.004	-0.032	-0.002	-0.000	-0.000	-0.000	-0.002	-0.039	-0.002	-0.000	-0.000
0.000	-0.000	-0.000	-0.004	-0.032	-0.002	-0.000	-0.000	-0.000	-0.002	-0.039	-0.002	-0.000

Fig. 5. Matrix A (eq. 17) upper left corner

Serial computations were also made by means of a finite span wing with airfoil NACA2412 at different angles of attack. Static pressure coefficient distribution is shown in Fig. 6 and Fig. 7.

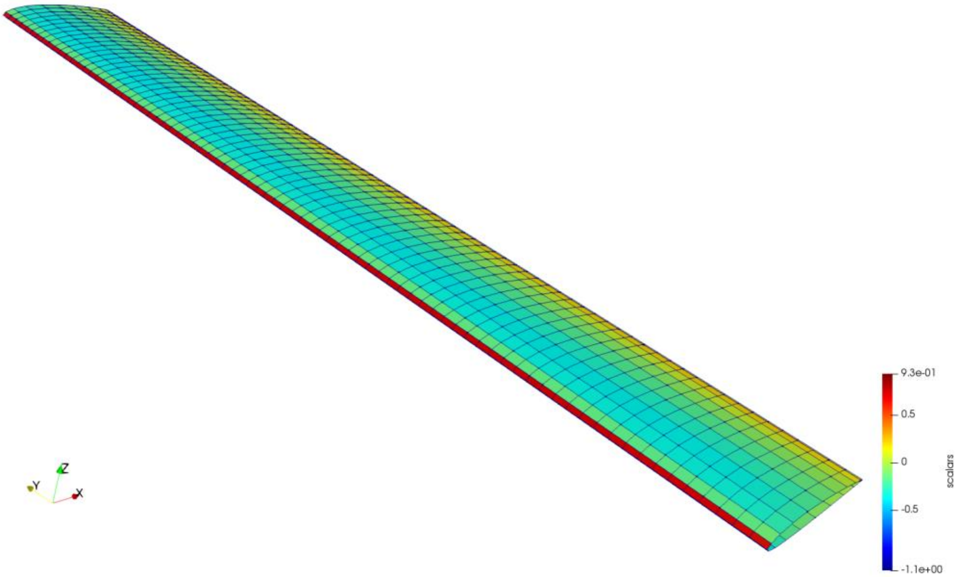

Fig. 6. Static pressure coefficient distribution over a NACA2412 wing, α = 0 deg, 20 × 80 panels

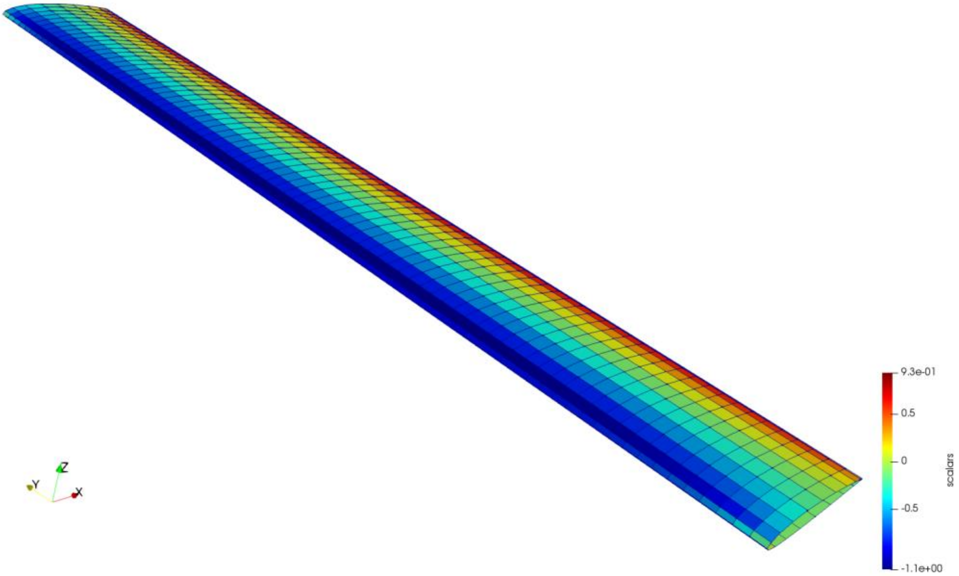

Fig. 7. Static pressure coefficient distribution over a NACA2412 wing, α = 10 deg, 20 × 80 panels

4. Discussion

The obtained static pressure coefficient distribution over a circular cylinder surface fully agrees with what Katz and Plotkin, [1, p. 69], discuss in their textbook. The static pressure coefficient varies within 1 and –3, so does the same quantity which is shown in Fig. 4. In three-dimensional case, the suction pressures are much smaller, i.e. the so-called “relieving effect” is obtained numerically.

The static pressure coefficient “suction” values are easily observable when the angle of attack is high, Fig. 7, blue region right after the leading edge. The coefficient distribution is symmetric in relation to the mid cross section because the sideslip angle is zero. Although side wing patches are absent, the static pressure coefficient distribution flattens at both wing ends. This result might be explained by the wingtip vortex phenomena.

In Fig. 5, the upper left corner of coefficient matrix A (17) is shown in case of circular cylinder serial computations. The dominating main diagonal could also be noticed. It justifies usage of the iterative scheme (24).

Secondary quantities might be computed further such as lift coefficient and induced drag. Additional demonstration of the Gauss–Seidel method rate of convergence might be seen in [5].

The source code used in the current study is developed by means of Minimalist GNU v. 5.1.0 for Windows. The visualizer used is ParaView v. 5.6.0.

References

1. Katz, J., A. Plotkin, Low Speed Aerodynamics, 2nd edition, Cambridge University Press, 2001.
2. Lamb, H., Hydrodynamics, 4th edition, Cambridge University Press, 1916.
3. Hess, J. L. and A. M. O. Smith, Calculation of Potential Flow About Arbitrary Bodies, Progress in Aeronautical Sciences, 8, 1967, pp. 1138
4. ASCII VTK Files. https://www.visitusers.org/index.php?title=ASCII_VTK_Files
5. Solution of Linear Equations by Stationary Iterative Methods, 2018. URL: http://www.space.bas.bg/ acsu/Ax_b/

ПРИЛОЖЕНИЕ НА ПАНЕЛЕН МЕТОД ЗА АНАЛИЗ НА СТАЦИОНАРНО ТЕЧЕНИЕ ОКОЛО КРИЛО С КРАЙНА РАЗПЕРЕНОСТ

К. Методиев

Резюме

В настоящата статия се изследва течение на несвиваем безвихров поток около тяло с крайни размери. За да се намери консервативния вектор на полето е необходимо да се реши диференциалното уравнение на Лаплас относно потенциална функция. Нетривиално решение на задачата с използване на втора формула на Грин дава като резултат стойност на потенциалната функция на двойния слой в произволна точка от полето, която не принадлежи на граничната повърхност. Потенциалната функция зависи от особености в полето на течението „източник/падина“ и „дипол“, разпределени по повърхността на тялото, както и от следа, прикрепена към задния ръб на крилото. Следата гарантира удовлетворяването на условието на Кута. Геометрията на тялото се апроксимира с квадратични панели, с цел да се пресметнат лицевите интеграли за всеки панел точно. За да се сведе задачата до линейна алгебрична система е необходимо да се пресметне влиянието на всеки панел в точка от полето. Получената матрица коефициенти е с преобладаващ главен диагонал. Системата се решава итеративно по метода на Гаус-Зайдел.

Целта е разработване на сорс код за решаване на поставената задача. Кодът е автентичен и в него не са използвани спомагателни библиотеки. Тестовете за валидация на кода, както и числените резултати са показани графично и обсъдени в статията.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e12

COMPARISON OF OBLIQUE SHOCK WAVE ANGLE IN ANALYTICAL AND NUMERICAL SOLUTION

Assen Marinov

Aviation Faculty, National Military University
e-mail: asen_aerodynamics@abv.bg

Keywords: *Subsonic aircraftt, Shock wave angle, Wave drag.*

Abstract

The drag of the subsonic aircraft is largely formed by the skin friction drag and lift-induced drag [2]. At transonic flight occurs shock wave. Determination of shock wave angle is important part of design of every aircraft, which working in supersonic airflow regimes. Formation of shock waves cause formation the wave drag. The wave drag could account about 35 % from total drag of aircraft. Shock wave angle is directly linked with the intensity of itself.

This work compares shock wave angle calculations using analytical and numerical solving methods.

1. Introduction

For analytical solve of shock wave angle are used conservation equations of mass, momentum, and energy [1].

Consider that flow is steady, inviscid and adiabatic flow with no body forces, ***continuity equation*** is:

(1) $$\oiint_S \rho V . dS = 0.$$

The continuity equation for an oblique shock wave is:

(2) $$\rho_1 V_{1n} = \rho_2 V_{2n}.$$

Momentum equation

The integral form of the momentum equation can be resolved into two components – tangential and normal to the shock wave.

Tangential component:

(3) $$\oiint_S (\rho V.dS)V = -\oiint_S pdS$$

(4) $$V_{1t} = V_{2t}, \qquad (S_1 = S_2).$$

The normal component:

(5) $$p_1 + \rho_1 V_{1n}^2 = p_2 + \rho_2 V_{2n}^2.$$

Energy equation

If consider, that flow is steady, inviscid, adiabatic and without body forces, then energy equation reduce to:

(6) $$\oiint_S \rho\left(e + \frac{V^2}{2}\right)V.dS = -\oiint_S pV.dS.$$

For ideal gas:

(7) $$h_1 + \frac{V_{1n}^2}{2} = h_2 + \frac{V_{2n}^2}{2}\ .$$

Deduce: changes across an oblique shock wave are governed only by the component of velocity normal to the wave.

2. Analytical solving method

Since equation (4), tangential components of the velocity, remain the same, while normal component decreases across the shock, the flow is deflected by angle θ toward the shock front after passing it [1].

Considering that: $M_{1n} = M_1 sin\beta$ and $M_{2n} = M_2 \sin(\beta - \theta)$,

(8) $$\frac{tan\beta}{\tan(\beta - \theta)} = \frac{(\gamma + 1)M_1^2 sin^2\alpha}{(\gamma - 1)M_1^2 sin^2\alpha + 2}.$$

One of possible solution of this equation is:

(9) $$tan\theta = 2cot\beta\left(\frac{M_1^2 sin^2\beta - 1}{M_1^2(\gamma + cos2\beta) + 2}\right).$$

Equation (8) determines angle θ, when set M_1 and β.
The other solution determines β, when set M_1 and θ [3].

$$(10) \quad \beta = arctan\left[\frac{b + 9a.tan\mu}{2(1-3ab)} - \frac{c(27a^2\,tan\mu + 9ab - 2)}{6a(1-3ab)}.tan\left(\frac{n}{3}\pi + \frac{1}{3}arctan\frac{1}{c}\right)\right],$$

$$a = \left(\frac{\gamma-1}{2} + \frac{\gamma-1}{2}tan^2\mu\right)tan\theta;$$

$$b = \left(\frac{\gamma+1}{2} + \frac{\gamma+3}{2}tan^2\mu\right)tan\theta;$$

$$c = \sqrt{\frac{4(1-3ab)^3}{(27a^2c + 9ab - 2)^2} - 1},$$

where n = 1 for weak shock solution,
function $\beta = f\,(M, \theta, \gamma, n) \rightarrow$ *function* $\beta = f\,(1.5, 12, 1.4, \text{and } 1)$.

3. Numerical solving method

Another form of the solution shock wave angle β is numerical solution, applying the fundamental laws of mechanics to a fluid gives the governing equations for a fluid.

A 2D geometry for the model have a deflection angle $\theta = 12^o$. Cell zone condition for the surface body is defined as ideal gas. Fig. 1 shows boundary zones in the calculation domain.

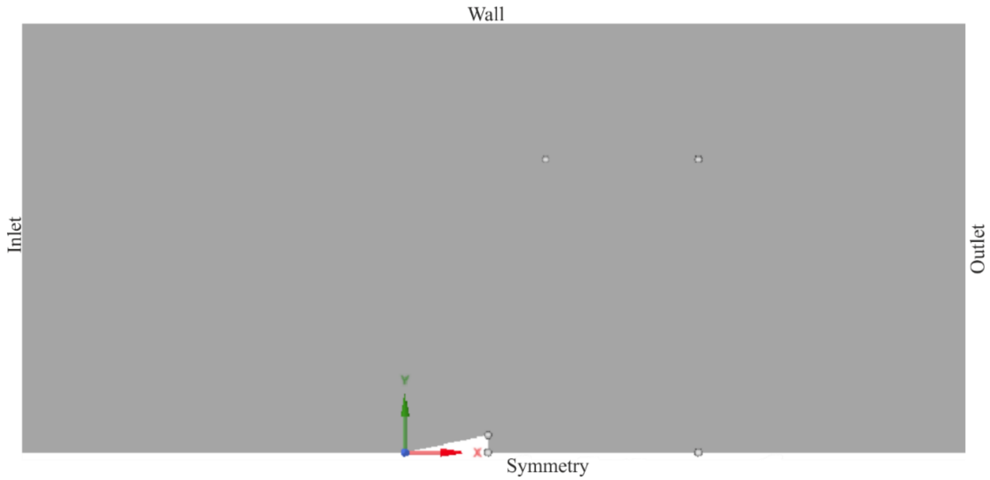

Fig. 1. Calculation domain

4. Results

The results from analytical and numerical solutions for weak shock wave angle are presented in Table 1.

Table 1

Method	Wave angle	Note
Analytical	64.35^o	Weak solution
Numerical	64.3^o	Weak solution

The color of Fig. 2 indicates the changing Mach area of the shock wave.

For a set Mach number of each value of the flow deviation angle after the shock wave θ, correspond two values of angle β. The smaller angle value corresponds to a weak shock wave (supersonic airflow after the shock wave), while at the higher angle value of the the shock wave corresponds to a strong shock wave (subsonic velocity of the airflow after the shock wave). When body is with a wedge shape, always realizes a weak shock wave (Fig. 3).

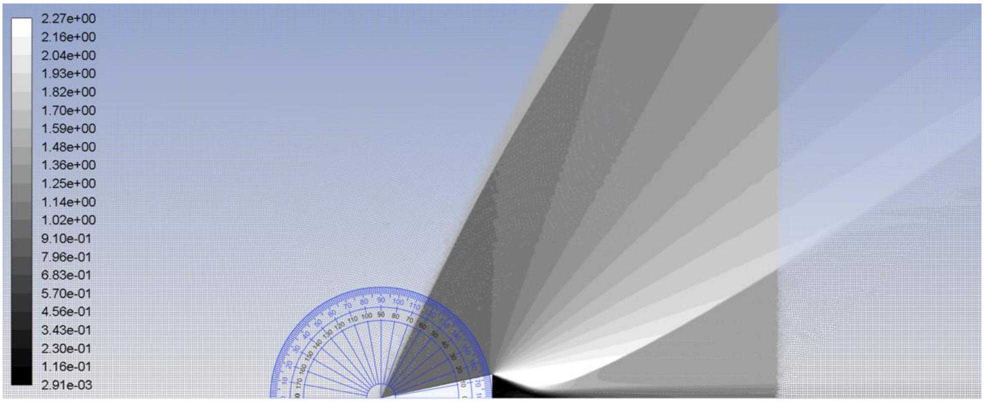

Fig. 2. Mach area of the shock wave

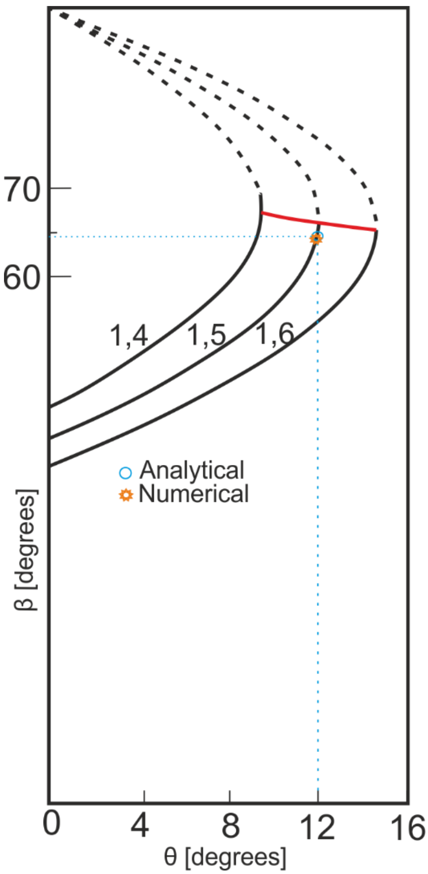

Fig. 3. $\beta = f(\theta)$

5. Conclusions

The simulation of shock wave angle has been carried out using analytical and numerical methods and carried out and the following conclusions can be drawn:

- The CFD simulation is able to predict precisely shock wave angle;
- A reasonably good agreement was obtained between analytical and numerical methods, when determining shock wave angle.

References

1. Anderson, J., Fundamentals of Aerodynamics, McGraw-Hill series in aeronautical and aerospace engineering, 2017, ISBN 978-1-259-12991-9.

2. Kambushev, M., S. Biliderov, Determination of aerodynamic coefficients necessary for the control of MAVs, International scientific journal TRANS MOTAUTO WORLD, 2017, pp. 94–97. ISSN 2367-8399
3. Lael vonEggers Rudd, Mark J. Lewist, Comparison of shock calculation methods, Journal of Aircraft, 1998, 35, 4, pp. 647–49. ISSN 0021-8669

СРАВНЯВАНЕ НА НАКЛОНА НА УДАРНА ВЪЛНА ПРИ АНАЛИТИЧНО И ЧИСЛЕНО ПРЕСМЯТАНЕ

А. Маринов

Резюме

В полет самолетите предизвикват във въздушния поток силни смущения. Тласкайки намиращия се пред тях въздух, те повишават налягането на въздушния поток толкова повече, колкото по-голяма е скоростта на полета. Ъгълът между допирателната към фронта на ударната вълна и вектора на скоростта се изменя по фронта. Като с отдалечаване от обтичаното тяло се намалява този ъгъл, което води и до намаляване на интензивността на скока на уплътнение.

Фронтът на скока на уплътнение разделя въздушния поток на смутена и несмутена част, като в смутената част настъпва съществено изменение на параметрите на въздушния поток.

Колкото е по-голям ъгълът на наклон на скока на уплътнение, толкова по-съществено нарастват плътността и налягането зад него. Това води до съществено изменение на силите, които действат на обтичаното тяло, което от своя страна налага да се търсят способи за точно определяне на съответния ъгъл на наклон.

Bulgarian Academy of Sciences. Space Research and Technology Institute.
Aerospace Research in Bulgaria. 31, 2019, Sofia

DOI: https://doi.org/10.3897/arb.v31.e13

MODEL OF THE OPERATOR FOR CONTROLLING THE AIRCRAFT WITH MATHEMATICAL MODELLING OF THE MOVEMENT

Ognyan Stoykov

Aviation Faculty, National Military University, Bulgaria
e-mail: oss_10@abv.bg

Keywords: *Aircraft, Mathematical modelling, Linear model, Transmitting function of the operator.*

Abstract

For controlling the aircraft with mathematical modelling of its movement in case of aiming at surface targets is proposed the use of the transmitting function of the operator, which is in itself a linear model in a tracking mode.

Introduction

For mathematical modelling of the movement of the aircraft aiming at surface targets is used the transmitting function of the operator.

The operator monitors the target (angles $\beta_{1т}$, $\varepsilon_{1т}$) and the moving grid on a screen and observing the position of the target relative to the targeting grid, i.e. the differences $\beta_1 - \beta_{1т}$ and $\varepsilon_1 - \varepsilon_{1т}$, attempts to eliminate them utilizing the controls of the aircraft. The aircraft reacting to the deflection of the controls changes the parameters of its position relative to the target [1, 3, and 5].

Main body

In order to obtain the model of the aforementioned process, it is necessary to model the reaction of the operator to the mismatching, i.e. to obtain the calculated values of the aircraft controls deflection depending on the angles of mismatching:

$$(1) \qquad \begin{aligned} \delta_{ви} &= \delta_{ви}(\varepsilon_1 - \varepsilon_{1т}); \\ \delta_{еи} &= \delta_{еи}(\beta_1 - \beta_{1т}), \end{aligned}$$

where $\delta_{ви}$, $\delta_{еи}$ are the calculated values of deflection of the elevators and ailerons (the angle of the rudder $\delta_н = 0$)

The role of the operator as a control unit of the closed system pilot-aircraft consists in:

- obtaining information from the onboard instruments and the environment;
- processing the obtained information and determining the amount of force that needs to be applied on the aircraft controls;
- applying the necessary amount of force on the controls of the aircraft.

The model of control of the aircraft, with participation of the operator, works based on the principle “tracking with pursuit” or tracking with compensation. In the system pilot-operator, the operator observing the quantity of the input and output signal of the system attempts to reduce to minimum the mismatching between the target and the tracking grid [2, 4].

To adequately describe the work of the operator in the control system, their transmitting function (or another mathematical definition) has to reflect the main characteristics of the operator as a

- ability to part of the control system. The main characteristics of the operator are:

- temporary delay of their reaction by the input signal;

- ability to adapt to the dynamic characteristics of the object of control and the character of the input signal impact;

- ability to react to deviations of the parameter from the predefined quantity, to its derivative and the integral of the deviation from parameter;

- intensify the quantity of the force applied on the controls;

- non-linearity of the operator’s characteristics;

- functioning of the operator within the control system as a multi-channel unit;

- dependency of the quality of operator’s control upon their psycho-energetic potential.

There are various models (linear, non-linear) in existence that describe the work of the operator, including the abovementioned characteristics. The linear models, which are the most convenient from an engineering point of view, have been developed in details and have been widely applied. These, however, possess certain disadvantages:

- they do not consider to a sufficient degree the ability of the operator to forestall the process;

- they are not capable of explaining the experimental data which indicate that in some occasions the operator manifests discreet behavior.

With a linear model of the operator in tracking mode, they are regarded as a unit of the tracking system and can be described with transmitting function. It is assumed that the transmitting functions of all the operators possess identical structures and their individual characteristics are given an account of through the values of the coefficients of the transmitting functions. The model of the operator provides not an accurate but an approximate account of the characteristics of the real operator.

To obtain the calculated values of aircraft’s controls deflection depending on the angles of mismatching

$$\delta_{\text{ви}} = \delta_{\text{ви}}(\varepsilon_1 - \varepsilon_{1\text{т}}),\ \ \delta_{\text{еи}} = \delta_{\text{еи}}(\beta_1 - \beta_{1\text{т}}),$$

the following transmitting function of the operator is proposed:

(2) $$W_{\text{л}}(p) = \frac{K_0 e^{-\tau p}(T_1 p + 1)K_1}{(T_2 p + 1)(T_3 p + 1)},$$

where τ is the time characterizing the delay of the reaction to the input signal;

- K_0 – coefficient of intensification of the operator;
- T_1 – constant coefficient, characterizing the ability of the operator to differentiate, i.e. to react to the speed of change of the input signal;
- T_2 – constant coefficient of the inertial (aperiodic) unit of the operator;
- T_3 – constant coefficient defining the neuro-shoulder reaction;
- K_1 coefficient of intensification of the neuro-shoulder unit.

The structural scheme of the preliminary function of the operator is illustrated on Fig. 1.

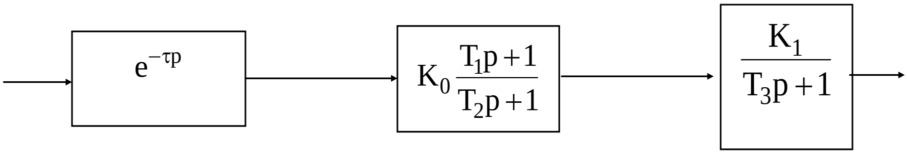

Fig. 1. Structural scheme of the transmitting function of the operator

The first unit responsible for receiving data from the instruments and processing the signals, in its dynamic characteristics is an intensifying unit with a delay function.

The second unit is a calculating element, performing intensification and differentiation of the received signals – it possesses the characteristics of intensifying, inertial and boosting units.

The third unit according to its dynamic characteristics is inertial and reflects the neuro-muscular influence on the object of control. Experiments show that the operator changes their transmitting characteristics depending on the quantities of the controlled object and the type of the functions of the interference. Experience, training and fatigue influence the type of the transmitting characteristics, i.e. the operator does not possess a certain, defined transmitting function but they can “tune” for work in accordance with any given function of a certain class.

For a certain type of tasks and specific aircraft the values of the coefficient of intensification K_0 and K_1 and the constant coefficients T_1, T_2, and T_3 and τ vary in small range.

The transmitting function of the operator is used to calculate the angles of deflection of the controls of the aircraft when mismatching signals are fed to it.

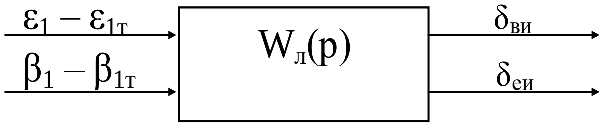

Fig. 2. Transmitting function of the operator

Using the Taylor series and restricting it to the use of only the first two members

(3) $$e^{-\tau p} = 1 - \tau p,$$

and introducing the symbols:

(4) $$\varepsilon_1 - \varepsilon_{1т} = \varepsilon_{11}; \quad \beta_1 - \beta_{1т} = \beta_{11},$$

we obtain:

(5) $$\frac{\delta_{ви}}{\varepsilon_{11}} = -\frac{K_{с.в}(T_1 p + 1)(1 - \tau p)}{(T_2 p + 1)(T_3 p + 1)};$$

(6) $$\frac{\delta_{ве}}{\beta_{11}} = \frac{K_{с.е}(T_1 p + 1)(1 - \tau p)}{(T_2 p + 1)(T_3 p + 1)},$$

where $K_c = K_0 K_1$ is the summary coefficient of intensification.

The higher the value of T_1 is, the more difficult the control process becomes for the operator. Moreover, the necessary intensification of T_1 requires greater precision in determining the speed of change of the input signal of the operator.

With a proficient enough operator the transient process fades faster, if the following values of the quantitates τ, T_1, T_2, and T_3 are adopted:

(7) $\tau = 0.1$ s, $T_1 = 0.1$ s, $T_2 = 0.1$ s, $T_3 = 0.1$ s.

From formulas (5), (6) and taking into consideration (7) are obtained the following differential equations for the longitudinal and lateral channel of control:

(8) $$\ddot{\delta}_{\text{ви}} + 20\dot{\delta}_{\text{ви}} + 100\delta_{\text{ви}} = -\ddot{\varepsilon}_{11}K_{\text{св}} + 100\varepsilon_{11}K_{\text{св}}.$$

(9) $$\ddot{\delta}_{\text{еи}} + 20\dot{\delta}_{\text{еи}} + 100\delta_{\text{еи}} = \ddot{\beta}_{11}K_{\text{се}} - 100\beta_{11}K_{\text{се}}.$$

Introducing the symbol:

(10) $$z_1 = \dot{\delta}_{\text{ви}} + 20\delta_{\text{ви}} + K_{\text{св}}\dot{\varepsilon}_{11},$$

after differentiation we obtain:

(11) $$\dot{z}_1 = \ddot{\delta}_{\text{ви}} + 20\dot{\delta}_{\text{ви}} + K_{\text{св}}\ddot{\varepsilon}_{11}.$$

Then from equations (8) is obtained:

(12) $$\dot{z}_1 = 100K_{\text{св}}\varepsilon_{11} - 100\delta_{\text{ви}}.$$

Introducing the symbol:

(13) $$\dot{z}_2 = \dot{\delta}_{\text{ви}} + K_{\text{св}}\dot{\varepsilon}_{11},$$

from formula (10) is obtained:

(14) $$\dot{z}_2 = z_1 - 20\delta_{\text{ви}}.$$

From formula (13) it can be written:

(15) $$\dot{\delta}_{\text{ви}} = \dot{z}_2 - K_{\text{св}}\dot{\varepsilon}_{11}.$$

After integrating with zero input conditions:

(16) $$\delta_{\text{ви}} = z_2 - K_{\text{св}}\varepsilon_{11}.$$

And, finally, from equation (8) is obtained the system:

(17) $$\begin{aligned} &\delta_{\text{ви}} = z_2 - K_{\text{св}}\varepsilon_{11}; \\ &\dot{z}_2 = z_1 - 20\delta_{\text{ви}}; \\ &\dot{z}_1 = 100K_{\text{св}}\varepsilon_{11} - 100\delta_{\text{ви}}. \end{aligned}$$

Analogically, from equation (9) is obtained:

(18) $$\begin{aligned} &\delta_{еи} = z_4 + К_{се}\beta_{11}; \\ &\dot{z}_4 = z_3 - 20\delta_{еи}; \\ &\dot{z}_3 = -100К_{се}\beta_{11} - 100\delta_{еи}. \end{aligned}$$

As a result of the integration of these equations are obtained the angles, required for control of the aircraft:

(19) $$\begin{aligned} &\delta_{в} = \delta_{ви} + \delta_{в.бал}; \\ &\delta_{е} = \delta_{еи}\ , \end{aligned}$$

where: $\delta_{ви}$, $\delta_{еи}$ are the values of the angles of the elevator and the ailerons, calculated using the transmitting function of the operator;

$\delta_{в.бал}$ – the balanced value of the angle of deflection of the rudder.

Conclusion

The problem of mathematical modeling of the movement of an aircraft with the use of the transmitting function of the operator consist in determining the summary coefficients of intensification $К_{св}$, $К_{се}$ based on the minimal time of fading of the preceding process using the accepted values of the quantities T_1, T_2, T_3, τ and known symbols in the transmitting function for the longitudinal and lateral channels of control.

References

1. Atanasov, M. A. and S. Mitev. Tochnost na opredeljane na koordinatite na v"zduhoplavatelno sredstvo po sposoba na doplerovoto opredeljane na p"tnata skorost, In: Nauchna konferencija „100 g. ot rozhdenieto na k-n Dimit"r Spisarevski“, NVU, fakultet “Aviacionen”, 2016, pp. 41–45, ISBN 978-954-713-109-5, 2018. (in Bulgarian)
2. Atanasov, M. A. Algorit"m na rabota na aviacionen pricelen kompleks s"s sledjashta sistema pri bombopuskane, In: Sb. ot dokladi na NK „S"vremenni tendencii v aviacionnoto obuchenie“, NVU, fakultet „Aviacionen“ NVU, 2017, pp. 114–121, ISBN 978-954-713-110-1 (in Bulgarian)
3. Atanasov, M. A., Modelirane na procesa na pricelvane s aviacionen pricelen kompleks s"s sledjashta sistema pri bombopuskane. In: Sb. ot dokladi na NK „S"vremenni tendencii v aviacionnoto obuchenie“, NVU, fakultet „Aviacionen“, 2017, pp. 121–127, ISBN 978-954-713-110-1 (in Bulgarian)
4. Stojkov, S. O., Analiz na v"zmozhnostite na metodite za prinuditelno otdeljane na aviacionni bombi. In: Sb. ot dokladi na NK „S"vremenni tendencii v aviacionnoto obuchenie“, NVU, fakultet “Aviacionen”, 2017, 56–60, ISBN 978-954-713-710-1
5. Stojkov, S. O., Analiz na bojnite v"zmozhnosti na iztrebitelite, JuNK, Tom I. NVU “V. Levski”, fakultet “Aviacionen”, D.M., 2013, pp. 85–89, ISBN 978-954-753-177-2 (Vol. 1). (in Bulgarian)

МОДЕЛ НА ОПЕРАТОРА ЗА УПРАВЛЕНИЕ НА ЛЕТАТЕЛНИЯ АПАРАТ ПРИ МАТЕМАТИЧЕСКО МОДЕЛИРАНЕ НА ДВИЖЕНИЕТО

О. Стойков

Резюме

За управление на летателния апарат при математическо моделиране на неговото движение с прицелване по земни цели се предлага използване на предавателната функция на оператора, представляваща линеен модел в режим на следене.