# ГОДИШНИК
# НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

## *Физически факултет*

# ANNUAL
# OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”

## *Faculty of Physics*

Том/Volume 111

УНИВЕРСИТЕТСКО ИЗДАТЕЛСТВО „СВ. КЛИМЕНТ ОХРИДСКИ“
ST. KLIMENT OHRIDSKI UNIVERSITY PRESS
СОФИЯ • 2018 • SOFIA

*Address for correspondence:*
Faculty of Physics
St. Kliment Ohridski University of Sofia
5, J. Bourchier, Blvd, 1164 Sofia, Bulgaria
Fax: (+359 2) 96 25 276
E-mail: annual@phys.uni-sofia.bg

ISSN 0584-0279

# СЪДЪРЖАНИЕ

# CONTENTS

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY "ST. KLIMENT OHRIDSKI"
FACULTY OF PHYSICS
Volume 111

# OIL SPILL DETECTION IN SAR SATELLITE IMAGES: A REVIEW

IRINA GANCHEVA

*Department of Meteorology and Geophysics*

*Ирина Ганчева.* РАЗПОЗНАВАНЕ НА НЕФТЕНИ РАЗЛИВИ НА SAR-СПЪТНИКОВИ ИЗОБРАЖЕНИЯ: ОБЗОРНА СТАТИЯ

Акцентът на тази обзорна статия е преглед на различни техники за автоматично откриване на нефтени разливи посредством анализ на SAR-(радар със синтетична апертура) спътникови изображения. Разглежда се физическото поведение на нефт върху водна повърхност и неговият ефект върху отразения или излъчения сигнал, като се представят и различни методи за широкомащабно наблюдение на океана. Обърнато е внимание на възможностите за оперативно внедряване. SAR-сензорите са разгледани подробно. Обобщени са последователните стъпки в разработването на автоматичен алгоритъм за откриване на нефтени разливи.

*Irina Gancheva.* OIL SPILL DETECTION IN SAR SATELLITE IMAGES: A REVIEW

This article examines different techniques for automatic detection of oil spills using Synthetic Aperture Radar satellite images. The physical behavior of oil on water surface and its effect on the reflected or emitted signal is reviewed and different methods for large scale oil spill ocean monitoring are presented, discussing in particular their operational use. The SAR sensors are reviewed in details. The different steps in the development of an automatic algorithm for the detection of oil slick are described in details.

***Keywords:*** oil spill, automatic algorithms, SAR satellite images
***PACS numbers:*** 92.20 Ny, 92.20 Wx

---

*For contact*: Irina Gancheva, Department of Meteorology and Geophysics, Faculty of Physics, Sofia University "St. Kliment Ohridski", 5 James Bourchier Blvd., Sofia 1164, Bulgaria, Phone: +359 2 8161289, E-mail: irina.gancheva@phys.uni-sofia.bg

# 1. INTRODUCTION

The environmental hazard of oil for the marine system has been studied and known for long time [1] and still oil spills are observed relatively often due to accidental or deliberate oil discharges. Some of the big oil pollution accidents such as the Deepwater Horizon spill in the Gulf of Mexico in 2010 or the Prestige tanker near the coast of Galicia in 2002 have attracted significant media interest and social outrage. Even though according to ESA accidental discharges hold only 7% of the oil pollution and operational discharges from tankers in rivers and oceans sum up to 80% of the oil spills [2]. Unfortunately in most cases they remain not traced.

There are various methods for ocean surface monitoring however the spaceborne remote sensing has proven to be one of the most efficient because of the large range of provided spectral information, great accuracy and high frequency of provided data. The identification of an oil spill, its extent and the type of the spill – if crude or natural oil – can be gained by this type of monitoring. The combination of this information with airborne monitoring can provide additional evidences about the pollutant. Very often it is necessary to analyse data from different sensors and wavelengths in order to discriminate between alga blooms, look-alikes and man-made slicks. The calculation of a set of features, such as contrast, shape, homogeneity and slick surroundings can contribute with additional clues for the detected object and thus support to distinguish actual oil from look-alikes. These procedures could significantly reduce the false alarm ratio for the automatic oil spill detection.

Synthetic Aperture Radar (SAR) compared to other methods remains the most efficient satellite sensor for oil spill detection exploring large areas regardless daytime and weather conditions. These type sensors provide regularly data and can detect even small amounts of oil without the need for good image resolution. The satellites equipped with SAR deliver data from different frequency bands, making it useful for various environmental purposes. This method presents also some limitations: it does not deliver information for the oil thickness and type. Some problems appear also due to the wind and water interaction.

Other space- and airborne methods for oil spill detection include visual, infrared and ultraviolet remote sensing, laser fluorosensors, microwave radiometers and scatterometers. All of them are useful and partly advantageous compared to SAR, however so far they have had only limited success for accurate, regular and spacious ocean monitoring.

Some good reviews on oil slick detection making use of different monitoring methods have been already published. Fingas and Brown [3] have done an excessive study on the different methods for detection of oil and on discussing their advantages and disadvantages. Another major focus in their review is the description of the techniques for slick thickness measurements and for detection of oil in the whole water column and on the sea bottom.

Solberg's review [4] is a very valuable read with a main focus on SAR and space borne-based detection methods, discussing also manual and automatic detection approaches.

Alpers et al. [5] reviews the behaviour of different oil types on sea surface and presents some discrimination methods such as statistical approaches, using differences in the dielectric constant and polarimetric parameters.

Topouzelis [6] presents a study on oil slick monitoring using SAR images with a focus on the semi- and fully automatic methodologies for their detection. A special advantage of this review is the comparison between different dark object classification methods and their operational advantages and disadvantages.

Gens' review [7] covers different applications of SAR images for oceanographic purposes such as ocean wave imaging, ocean currents, sea-floor topography, oil spill and ship detection, wind speed, rainfall and others.

## 2. PHYSICAL BACKGROUND

There are different methods for oil spill monitoring depending on the sensor type being used. In order to estimate their theoretical feasibility the optical properties of oil and water should be considered first.

Measurements of the oil/water reflectivity and relative absorbance for different wavelengths have been done, however not across the whole spectrum and with limited accuracy [8–11]. In the review of Fingas and Brown [3] is presented a generalized summary of different measurements, coming to the conclusion that the oil/water reflectivity and relative absorbance have very similar behaviour for different wavelengths ranging from 200 to 1100 nm. The similarity in the properties of both media leads to the conclusion that considering only the visible spectrum would not be sufficient to firmly distinguish oil.

The visible remote sensing uses radiation with wavelength of 400–700nm and in this range oil has slightly higher surface reflectance than water. Nevertheless the absorption/reflection differences are rather small and usually not sufficient to clearly distinguish oil from water [12]. The use of polarized lenses has the potential of enhancing the visibility of oil and some authors investigate this effect [13], however a major difficulty for the visible remote sensing remains the issue to reduce the sun glitter, which might be confused for oil sheens [14, 15].

The biggest limitation for the visible remote sensing is the need for clear skies during the overpass, which is rarely the case. For the implementation of oil spill detection algorithms long time series of data are to be analysed, which is hard to be obtained considering the overpass period and the problems with the collection of usable data.

The use of infrared signal for the detection of oil slicks relies on the thermal properties of oil. It heats up faster than water and thick oil films absorb the solar radiation. Afterwards a portion is re-emitted as thermal energy and can be distinguished from water by its higher temperature. Thus thick oil films will appear hot, those with intermediate thickness are cool and thin films cannot be detected by this method. Some studies have investigated the effect of thickness on the temperature and conclude that the minimum detectable layer is between 10 and 70 μm [3]. The exact reason why intermediate layers appear as cool is not completely understood, however the authors make a hypothesis about the destructive interference of thermal radiation waves.

An important limitation of the infrared detection technique is the fact that the longer oil stays on sea surface, the more it mixes with it and in general the emulsion cannot be detected with infrared sensors [16]. The reason is that the thermal conductivity of the emulsion is quite similar to the one of the ambient. It should be also mentioned that infrared scanning at night-time presents lower contrast of the images compared to daytime results [17], thus the operational use of that method is difficult. In addition, the sediments, seaweeds or other organic matter, shoreline and oceanic fronts interfere with the infrared signal. In conclusion, there are several major disadvantages of the infrared remote sensing for oil spill monitoring, which make it an unfavourable choice for the operational use.

Another technique for oil spill detection is analysing the images in the ultraviolet spectral band. In this range of the electromagnetic spectrum oil is highly reflective even for layers thinner than 0.1 μm. Earlier studies combine ultraviolet and infrared images to produce relative thickness map of oil slicks. However, ultraviolet signals are also affected by sun glints, biogenic material or wind slicks [18], which is a major limitation. This method is not widely used today due to the unimportance of the oil thickness when taking counteractions.

Another recent technique makes profit of laser fluorosensors. It is based on the excitation of some electronic states in the aromatic compounds in petroleum oils after absorption of ultraviolet light. These states release energy through fluorescence emission in the visible spectrum range [19–21]. The emitted radiation is of specific wavelength and can distinguish oil from other biologically active materials. Another benefit is that the detected signal provides information about the oil types. On the other hand, the major limitation is that this type of sampling is done with aircrafts and thus is rather inconvenient for large ocean areas.

Using microwave radiation for ocean monitoring can be done in two ways – with active and passive microwave sensors. For the purpose of oil spill detection the active remote sensing has proven to be more effective, as the passive sensing is dependent on weather conditions and daytime.

The passive microwave sensor measures the reflected space radiation, detecting the difference in the emissivity factor for water (0.4) and for oil (0.8) [22].

**Fig. 1.** A SAR image from Baltic sea. The dark objects marked with the white arrows are oil spills and the other dark formations are look-alikes; image from [4]

Additionally, the change of the signal with the oil film thickness provides a tool for measuring the spill thickness. A serious disadvantage is the unsatisfying spatial resolution, in the range of tens of meters for a radiometer. Some extensive studies on the usage of passive microwave sensors have been presented in [23–25].

The remote sensing in the microwave spectrum range in active mode is well studied and widely spread method, providing good results. The active microwave sensor, better known as radar, is widely used for ocean and land monitoring. It sends signals in the microwave range and by the detection of their reflectance amplitude and phase, one can examine the objects or landscape on the way. The resulting radar image of ocean surface is known as sea clutter, because the capillary waves reflect the incident radar signal and produce a bright image. On the other hand oil on the sea surface dampens the capillary waves and an oil slick appears as a dark area in the bright ocean [26]. Similar effect of damping the waves and receiving a dark spot on the ocean surface is gained by fresh water and wind slicks, glacial flour, biogenic oils and sea weeds, calming the water above them [7]. Extensive studies and classification of these and other look-alikes have been accomplished, considering that they are the key for the dislocation of oil spills and false alarms. Still the percentage of false detected dark spots is around 20% [27].

Fig. 1 is an example of how oil spills and look-alikes are visible on a satellite images and it is a SAR image from the Baltic sea. The dark objects indicated with white arrows are actual oil spills and the other dark formations correspond to look-alikes. With this the challenges for the correct classification become obvious.

An important advantage of the active remote sensing technique is its implementation regardless weather conditions and daytime since its operation is not based on visible light and the water molecules in the air do not reflect the signal.

Radars used for military purposes are not useful for oil spill detection as they remove the clutter signal, essential for oil monitoring. The radars, which deliver the best information for environmental remote sensing, are Synthetic Aperture Radar (SAR) and Side-Looking Airborne Radar (SLAR). The images, produced by SAR are with a better resolution and greater range, making it the better option for ocean monitoring. An extensive comparison between SAR and SLAR is done by [28].

SAR radars send signals at different wavelengths, making them useful for various purposes. The frequency and the polarization of the SAR sensor have a significant impact on the detectability of oil slicks. The different SAR bands correspond to the wavelength of the transmitted signal – X-band (2.4 – 3.75cm), C-band (3.75–7.5cm), S-band (7.5–15cm), L-band (15–30cm) and P-band (30–100cm).

The accurate measurement of the surface roughness is the key for detecting anomaly patterns on sea surface with SAR. An adequate choice for the wavelength of the transmitted signal $\lambda_t$ is essential, because the backscattered signal will be modified by the surface roughness at the same scale as $\lambda_t$. Kim et al. [29] has examined the expected dampening of the return signal as a function of the wavelength and showed that the dampening ratio of the X-band is higher than that of the C-band for fixed incidence angle and wind speed. The high ratio between the transmitted and the dampened signal is an indication for the presence of oil. For oil spill detection different studies have shown that the X-band delivers best results, followed by C-band and L-band [29–31], as the first two have approximately the same scale as the Bragg waves. For large incidence angles $\theta = 20° – 60°$ the scattering mechanism at the ocean surface can be described with the Bragg scattering theory due to the wavelength of the capillary waves which is about a couple of centimetres [32]. Using the L-band has proven to be the most inefficient one as the Bragg waves don't resonate with the incident signal causing much smaller dampening of the backscattered wave [33].

The polarization of the SAR signal plays an important role as well. There are horizontal (H) and vertical (V) polarization and the possible combinations of transmitted and received signal are HH, HV, VV and VH. There are studies analysing the quality and usability of the different polarization combinations [4, 34, 35]. For airborne radars the VV polarization provides the best results, however all combinations are reasonable to use. The HH polarization is suitable for ship detection, meaning that combined observations usually deliver better results. In the last decade research on polarimetric SAR data has proven to deliver good results for oil spill detection [4, 36, 37]. In this case the polarimetric SAR sensor collects data of all combinations – co-polarized (VV and HH) and cross-polarized (VH and HV).

## 3. SATELLITES

The Synthetic Aperture Radar (SAR) is an active method for Earth observation and topography monitoring and more specifically their variations in time. A high resolution image of the examined surface can be obtained by analysing the intensity and phase of the received signal compared to the transmitted signal, as long as the scanned objects remain stationary over the scanning period of time. Alternatively the principle of SAR functioning can be explained by considering the Doppler shift of the echo signal. The vertical position of the objects can be determined comparing the upshift and downshift of the echo signal.

Tremendous reviews with information about the construction principles and function of SAR are presented in [38, 39].

For the in-depth understanding of SAR some geometrical definitions should be considered, as presented in figure 2. The SAR antenna is moving parallel to the Earth surface and vertical to the radar beam and its exact position is known at any time. The along track direction of the antenna is referred as azimuth or cross range and the perpendicular one – range or cross track. The footprint is the land piece scanned at the present moment and the swath is the land strip along which the antenna is moving. The swath width can vary from few kilometres up to 20 km for an airborne SAR and from 30 to 500 km for a spaceborne SAR.

The azimuth resolution $\delta_a$ of a SAR is given by the construction specification of the synthetic aperture or the path length during the echo signals form the target is received by the radar. It can be calculated using

$$\delta_a = r_0 \Theta_a = r_0 \frac{\lambda}{2L_{\mathrm{SA}}} = \frac{d_a}{2}$$

with the factor two in the denominator because of the two-way path of the signal, $r_0$– the slant range distance, $\Theta_a = \lambda / d_a$ – the virtual beam width, λ the radar signal-wavelength, $L_{\mathrm{SA}}$ – the synthetic aperture length and $d_{\mathrm{a}}$ – the antenna’s length [38].

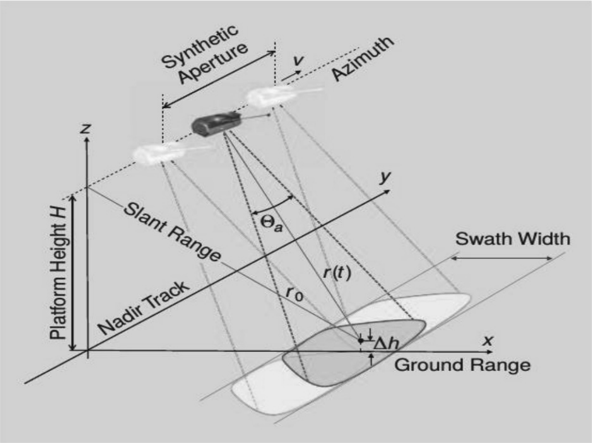


**Fig.2.** SAR geometry with illustration of some of the basic SAR terms; $r_0$ is the slant range, $\Theta_a$ – the azimuth beam width, and $v$ is the sensor velocity; from [39]

The first approaches for ocean monitoring for oil spills emerged with the data, delivered by ERS, ENVISAT ASAR and RADARSAT-1. Later the launch of RADARSAT-2, TerraSAR-X, COSMO SkyMed and SENTINEL-1 brought new imaging modes and high resolution data on regular basis for the ocean monitoring [4].

Earlier the ENVISAT data was mostly used for the operational oil spill detection till the end of the mission in 2012. Afterwards its place have been taken by the SENTINEL-1A mission launched in April 2014 and 1B launched in April 2016. There is a lot of information about the SENTINEL missions on the ESA official website [40].

The two SENTINEL-1 satellites are identically constructed and have near-polar, sun-synchronous orbit with 12 days repeat cycle for each satellite, making a 6-day revisit time for the European land and sea area. They share the same orbit plane with 180° orbital phasing difference. The satellites use HH–HV or HH polarization for monitoring the polar and sea-ice areas and VV–VH or VV polarization for the other land and water observations. The planning of European Space Agency is to launch SENTINEL-1C and SENTINEL-1D in 2021 and 2023 respectively.

The SENTINEL-1 mission has a single C-band synthetic aperture radar instrument working at a central frequency of 5.405 GHz and programmable bandwidth of 0–100MHz. It operates in four exclusive acquisition modes – Stripmap (SM – used only on request for extraordinary events), Interferometric Wide swath (IW – main operational mode for most service requirements), Extra-Wide swath (EW) and Wave mode (WV). Table 1 summarizes the important information for the SENTINEL-1 acquisition modes and figure 3 pictures the SENTINEL-1 product modes visually.

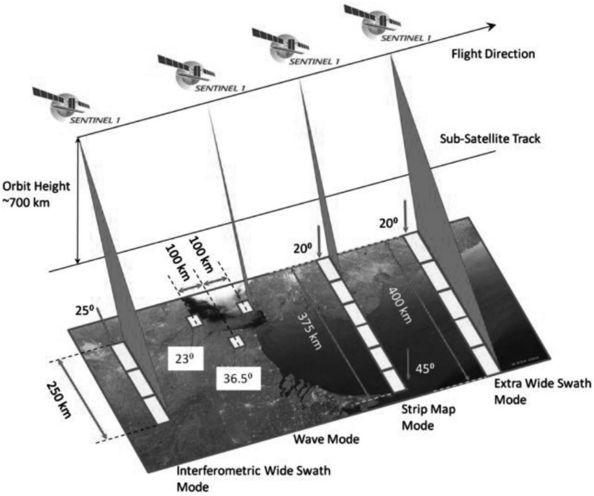


**Fig. 3.** Operation modes for SENTINEL-1A and B satellites, from [40]

**Table 1.** Operation modes of the SENTINEL 1 satellite mission, from [40]

| Mode | Incidence Angle [deg] | Resolution [m] | Swath Width [km] | Polarization |
|---|---|---|---|---|
| Stripmap | 20–45 | 5×5 | 80 | HH+HV, VH+VV, HH, VV |
| Interferometric Wide swath | 29–46 | 5×20 | 250 | HH+HV, VH+VV, HH, VV |
| Extra Wide swath | 19–47 | 20×40 | 400 | HH+HV, VH+VV, HH, VV |
| Wave | 22–35<br>35–38 | 5×5 | 20×20 | HH, VV |

The SENTINEL missions are of a particular interest for the research community due to the free access to the data base. The website of the Copernicus Open Access Hub [41] provides data from the SENTINEL-1, 2 and 3 missions freely after a short registration. Satellite images gained on different bands, with great accuracy and spatial resolution are provided on weekly basis, opening new opportunities for non-commercial research.

The TerraSAR-X mission is a German satellite mission launched by DLR in June 2007 with scientific and commercial purposes. Its lifespan was planned for 5 years, but more than 10 years after the launch, it is still operational. The high resolution images of the mission are used for hydrology, geology, climatology, oceanography, cartography, environmental and disaster monitoring. The SAR antenna in flight attitude points at 33.8° off nadir. It runs on a sun-synchronous circular down-dusk orbit with a revisit period of 11 days. It is equipped with a

X-band SAR antenna with 31 mm wavelength or a 9.6 GHz frequency. The spatial resolution is 40 m and the coverage is up to 270×200 km$^2$ making it suitable for observation of even small oil spills. It is also particularly useful for ship detection.

TanDEM-X is the follow-up twin mission of TerraSar-X, launched 3 years later, in June 2010. It has a high vertical accuracy better than two meters and together with TerraSAR-X it is from the first configurable synthetic aperture radar interferometers. TanDEM-X flies in close formation to TerraSAR, only couple of hundred meters apart, monitoring the Earth from different angles. More information about the missions can be found on the official DLR-website [42].

The satellite RADARSAT-2 was launched by the Canadian Space Agency in December 2007 and is still actively monitoring the Earth. It is a commercial mission, enhancing marine surveillance, ice monitoring, disaster management, environmental monitoring, resource management and mapping. RADARSAT-2 is equipped with a C-Band active antenna with centre frequency of 5.405 GHz and bandwidth of 100 MHz. On the website of CSA there is some more valuable information about the mission [43].

## 4. ALGORITHMS

With the growing amount of satellite data over different ocean and sea regions worldwide the possibilities for marine monitoring have increased dramatically. The manual image processing is time consuming and requires the availability of skilled operators which is significant limitation for the process optimization. Considering that a fast reaction is crucial for undertaking adequate counteractions, the importance of an automatized process for oil spill detection arises.

For the automatic detection of anomaly sea patterns, or dark objects as they appear on SAR images, there are certain procedures during the image processing which should be done in order to detect spills.

The first step when working with raw SAR data is always the pre-processing of the image. This includes image calibration, land masking and speckle reduction. These steps enhance the visibility of dark areas and make the distinction of their borders easier. Secondly, the dark spots are detected and isolated in a segmentation process. There are different techniques however mostly this is done using thresholding. In the third part of the detection algorithm a set of features is extracted and calculated for the detected dark objects. The classification process is the last step and it strongly depends on the information gained by the previous two steps. An extensive archive of correctly classified dark objects for training the algorithms and information about the local meteorological conditions is a helpful contribution for distinguishing between actual oil spills and look-alikes.

The image segmentation and feature extraction are the most important steps for the automatic oil spill detection. If here a dark area is not detected or the features lead to misleading conclusions an oil spill can remain undetected.

The dark formation identification is the first crucial step in oil spill detection. For the automatic approach a threshold algorithm, adaptive or not, can be applied. An early attempt of this technique is presented in [44], where the bimodal histograms in a set window is analysed. It was proven as a good procedure for thick spills; however the thin spills remain undetected. A similar approach was presented in [45], where a comparison between bimodal histogram method and adaptive thresholding is presented. Later it was shown that bimodal histogram provides decent results [46]. This algorithm was firstly developed for RADARSAT-1 SAR data and it spatially averages the image before applying an adaptive thresholding.

Simple thresholding algorithms use one value for the whole image and all pixels are compared to it. This value is usually one half of the average Normalized Radar Cross Section (NRCS) of the image or NRSC minus the standard deviation [47].

The adaptive thresholding uses a threshold value, which is calculated locally for areas covered by a moving window. Solberg et al. [48, 49] uses a threshold value of k dB below the mean value of the moving window. The value of k is calculated using a multi-scale pyramid approach and a clustering step. Karathanassi et al. [50] uses a fully adaptive value to local contrast variations. The technique of hysteresis thresholding was firstly described by [51]. It detects the edges of Gaussian-smoothed image and linear dark formations are successfully detected [52].

Another method based on the Laplace of Gaussian (LoG) and Difference of Gaussian (DoG) operators are presented in [53]. Firstly the original image is reduced by a unit of 2 by 2 pixels and multilayer images with decreased effect of noise and sea clutter are created. For the detection of dark areas the LoG derivative operator is applied. The sharpness of the oil spill shape is measured with a first order derivative operator – DoG. This effect is notable because water and oil are electromagnetically different. This makes the gradient of the image grey value on the water-oil boundary different. The wavelet packet transformation is a segmentation technique proposed by Liu et al. [54] and its linear feature detection scheme with LoG-operator as analysing wavelet. It is analogous to Fourier transformations, but localized in frequency and time. For the transformation are selected areas with multiple histogram peaks and for the extraction of small scale features is taken an edge detector wavelet transformation [54, 55]. The 2D wavelet transformation is highly efficient bandpass filter and can separate various scale processes delivering phase/location information in SAR image. It is good for near real-time screening of satellite data, data reduction and image enhancement.

The fussy clustering is a method where for each pixel a function is selected, which measures how much the pixel belongs to a certain value. Afterwards the Fuzzy C means (FCM) algorithm is applied and a pyramid structure is used to find membership values. The uncertain pixels are arranged in the lower pyramid level. A Sobel operator is used to enhance main edges of the original filtered image [56]. Firstly a fuzzy clustering is taken for preliminary partition of the pixels on the basis of grey level intensities and then simple cluster validity criterion is done to determine the optimal number of clusters present in the data. Another method for dark object segmentation is based on using mathematical morphology for the image segmentation [57]. The aim is to detect spills from moving tankers which is implemented in the selected features (elongatedness and spill dampening). An advantage of this method is that there is no need for prior knowledge about ocean conditions and it is also good for features extraction used in the decision process.

## 5. FEATURES

The different methods described in the previous part are used for the detection of suspicious structures on sea surface. Hereby it is important that during the detection algorithm the shape of the object is preserved as it is the one of the keys for distinguishing between actual oil slicks and look-alikes.

For creating a working oil spill detection algorithm the dark objects should be characterized following a certain characterization criteria. Every algorithm has a different set of features, but generally they can be organized in four classes.

The spatial geometry and shape of the dark object is analysed in almost every oil spill detection algorithm. Considering this information together with of some regional specifications of winds and currents, the differentiation between oil slicks and other surface patterns might be enhanced. Very often oil is discharged from moving tankers. In that case calculating the elongatedness, or the ratio between width and length of the dark object, is a very useful feature [57].

The backscattering level of the dark object and the surrounding is another informational class of features. For a neural networks classification backscattering ratios of different regions are considered as crucial [58]. The backscattering of the background surrounding is important due to wind speed dependence as well [47].

In the set of features the contextual positioning of the dark object is of particular interest. An additional database containing the locations of oil pipelines, platforms or the vessel traffic information might contribute significantly to the correct classification. Wind history is also useful for slick classification and age estimation [59].

The texture of a SAR image provides information about the spatial correlation between neighbouring pixels. The comparison between pixels of different regions on the SAR image puts the dark object in correlation with other regions.

In the following paragraphs the set of features for three different algorithms – adaptive thresholding, statistical classification method and neural networking – are observed.

Solberg et al. [60] has analysed Radarsat and ENVISAT SAR images with an adaptive thresholding algorithm. The observed features are:

- Slick complexity
- Slick power-to-mean ratio
- Slick local contrast
- Slick width
- Slick local neighbours
- Slick global neighbours
- Border gradient
- Slick area
- Distance to detected ship
- Slick planar moment
- Number of regions in the image
- Slick smoothness contrast

Fiscella et al. [47] describes an oil spill detection method where low resolution images are inspected for the presence of suspicious structures manually or automatically. Later the detected dark formations are classified using a Mahalanobis classifier and a compound probability classifier – both statistical classification methods.

For the automatic analysis of a SAR image the dark areas with Normalized Radar Cross-Section (NRCS) lower than one half of the average NRCS of sea area are selected. The actual oil spill detection is done after identification of the border of the dark object and evaluation of the following direct features:

- Perimeter
- Area
- Average NRCS inside the dark area
- Average NRCS in a limited area outside the dark area
- NRCS dark area standard deviation
- NRCS outside dark area standard deviation
- Gradient of the NRCS across the dark area perimeter
- Form factor: the dispersion of dark area pixels from its longitudinal axis.

Afterwards from this set of features the following quantities are derived/ calculated:

- Perimeter to area ratio
- Intensity ratio between average NRCS inside and outside the dark area
- NRCS standard deviations ratio inside and outside the dark area
- Ratio between NRCS intensity and its standard deviation inside the dark area
- Ratio between NRCS intensity and its standard deviation outside the dark area
- The ratio of the last two ratios.

Then the classification procedure is undertaken.

In Del Frate et al. [58] a neural network algorithm is used for the dark object classification. A long time series of over 600 images from the years 1997 and 1998 over different areas in the Mediterranean Sea is analysed. A histogram for every image is generated and the borders of the dark object are determined. The features needed for the dark object classification are the following:

- Area of the object $A$
- Perimeter $P$
- Complexity $C$ defined as $C = \frac{P}{2\sqrt{\pi A}}$
- Spreading $S$ – low value for long and thin objects and high for objects closer to circular shape
- Object standard deviation
- Background standard deviation
- Max contrast – difference between background mean value and the lowest value inside the object
- Mean contrast
- Max gradient
- Mean gradient
- Gradient standard deviation.

## 6. CLASSIFICATION METHODS

The last step of the oil slick detection is the classification procedure, undertaken in order to distinguish the actual oil slicks from look-alikes (for example anomaly alga blooms, sewage water discharges, surface water currents or capillary wave damping caused by local winds).

There are several classification methods published in the literature. An easy and common way for dark object classification is the use of statistical classifiers, where the decision is based on probability calculations. They are simple and reliable and the output can be easily reproduced.

Fiscella et al. [47] is testing a Mahalanobis [61] and a compound probability classifier where the probability of a dark object being an oil slick is calculated. The Mahalanobis classifier is comparing the set of characteristic features (written in the input vector **x**) with a template, composed from previous measurements. For this technique the Mahalanobis distances between the calculated set and the class's oil spill $m_1$ or look-alike $m_2$ is computed. The Mahalanobis distance $r_j^2$ is given by

$$r_j^2 = (x - m_j)' C^{-1} (x - m_j)$$

in the matrix form with $j = 1,2$ and $C$ the covariant matrix of $x$. For the classic compound method the probability $p$ for a dark object being an oil spill is calculated using

$$p = \frac{1}{1 + \prod q_i(x_i)/p_i(x_i)},$$

where $p_i(x_i)$ are the probability distribution functions for oil spill and $q_i(x_i)$ for look-alike classes.

A data set of 123 images was tested [47] and the correctly classified data with the Mahalanobis classifier for $p_{\mathrm{Mah}} > 2/3$ is 78% and for $p_{\mathrm{Mah}} > 1/2$ is 83%. The compound probability classifier delivers correctly classified data with $p_{\mathrm{Com}} > 2/3$ is 79%, and $p_{\mathrm{Com}} > 1/2$ is 82%.

A similar method based on statistical modelling with a rule based approach is using the Gaussian density function and is presented by Solberg et al. [48]. Similar to Fiscella et al. [47] a template set is used, but here it is derived from a signature database of 7.051 dark objects containing 71 oil spills and 6.980 look-alikes. The method classified correctly 94% of the oil spills and 99% of the look-alikes.

Another widely used method is the neural network classifier. It is considered effective because it operates well with nonlinear mapping of multidimensional input on single-dimensional output and complex statistics. Different from other statistics based classifier; the neural network approach doesn't need well defined relationship between input and output vectors, as it determines it after analysing a set of training data.

The neural network algorithm is a mathematical tool for calculating the probability of occurrence of a certain event. This is done by creating an input term which is then mathematically manipulated through multiple neurons where each calculates the sum of the inputs adds a bias term and then provides the result to the up-following neurons. The model topology is specific for each neural network and gives how the input, output and the hidden units are interconnected. In the feedforward network, which is also applied for the oil spill case, the input flows only forward to the next-level neurons and cannot return to the previous layers.

Del Frate et al. [58] used in their extensive study a neural network classification algorithm. They analysed 600 ERS images from which they extracted 139 images with dark objects, 71 of which were oil spills and 68 – look-alikes. More details about the exact functioning principles can be found in the publication. Before analysing the actual images the system is trained in order to get optimized results for the given issue. Once trained the network has examined the given pictures. It has misclassified 18% of real oil spills on the images as look-alikes and 10% of the look-alikes were wrongly classified as oil. The overall rate of misclassified pixels is 14%.

A more recent study based on article neural network is presented in [62] with 91.6% correctly classified oil spills and 98.3% look-alikes.

Another classification method based on fuzzy classification rules is presented by Karathanassi et al. [50]. Firstly homogeneous dark objects are extracted in any given resolution using a threshold technique, adaptive to any local contrast and later they are classified using a fuzzy logic. Each feature of the calculated set is considered as a separate class and each class consists of a set of fuzzy expressions. This makes the logical operation and the estimation of each specific value more accurate. The method processed 12 SAR images and classified successfully high percentage of the oil spills and the look-alikes.

It is difficult to compare the skill of the different classification methods, because they use different data sets, the dark object algorithms function differently and the set of features vary. Therefore the reported classification accuracy cannot be compared directly.

Analogously it is hard to compare the computational time and make statements which technique delivers fastest the most accurate results. Analysing the same data set using different segmentation techniques, features and classification methods could deliver valuable information for possible advantages and disadvantages.

## 7. EXAMPLES

Regional studies focusing on the local specifications of the different water basins have been done for most of the European seas. Several initiatives exist, monitoring European waters and delivering real-time information about oil spills. This is very important considering that a fast reaction is crucial for identifying the polluter.

The platform Oceanides [63] is a database with information collected from observations via aircraft and satellite radar images, all available from a single source. The project is funded by the European Commission as part of CORDIS (Community Research and Development Information Service). Oceanides is focused on marine monitoring and can be used for different purposes. The tool organizes the data depending on the interest of the operator and assembles the knowledge required to establish a more effective monitoring of oil pollution and identification of possible polluters. It is operational for European waters and is already implemented on regional scale for oil pollution monitoring.

Another on-line platform created for oil spill detection is CleanSeaNet [64]. This platform is a project of the European Maritime Safety Agency which is a decentralized EU agency. CleanSeaNet is a service focused on identification of oil spills, combined with vessel detection in European waters and it works by analysing radar satellite images. The images are processed within 30 minutes after the satellite passes overhead and an alarm for potential pollution is issued directly after. Correlating satellite data of detected vessels with vessel traffic reports increases the possibility of correctly identifying the polluter.

CleanSeaNet collects supplementary data such as optical marine images and oceanographic and meteorological information, which significantly increases the correct detection ratio.

For Bulgaria the Black sea is of a particular interest considering the geographical location and the importance to the region.

The extensive study of Malinovsky et al. [65] deals with the SAR analysis for oil spill detection particularly for the Black sea region. Using the polarization ratio in the ENVISAT Alternating Polarization Images they have detected 424 oil spill events in proximity to the major ship routes and oil platforms.

Ivanov and Kucheiko [66] have studied SAR images of the Eastern Black sea (2011–2013) and the Northern & Middle Caspian Sea (2009–2013) and compared the extend and source of the oil pollutions, finding very different results for the both regions. The oil spills in the Black sea are caused mainly by ships (tank washings and deliberate illegal discharges) and have a great extend up to 320 km$^2$. Those in the Caspian sea are of a much smaller surface area not exceeding 70 km$^2$ coverage [66].

Other studies of oil-spills in the Black sea are presented by [67–69], where different observation methods for oil slicks detection are applied.

In the past there were attempts to establish a real time monitoring system for tracking oil spills in Black sea as well; however as to the present moment they are not operational.

The online platform of JRC (Joint Research Centre) published information on the oil-spills discovered in the Black Sea in the period 2000–2004 (http://publications.jrc.ec.europa.eu/repository/handle/JRC55159).

The results were used in the project of the Commission on the Protection of the Black Sea Against Pollution (http://www.blacksea-commission.org/_projects_MONINFO.asp) "Monitoring and Information Systems for Reducing Oil Pollution" implemented in the period 2009–2010 with the main objective to prevent and take measures against operational/accidental/illegal oil pollution.

## 8. CONCLUSIONS

Spaceborne SAR sensors have proven to be most efficient among others for oil spill detection and their capacity for long-term, large-scale ocean monitoring has been demonstrated. Their good spatial resolution and feasibility at all-weather, all-time makes them a reliable source for long time series data. The latter is determining for creating a fully automatized method for oil spill detection, since every algorithm needs a data base for training, so that the dark object classification can be performed with a minimum false alarm ratio.

In this article various algorithms for oil spill detection have been presented, all of which deliver reasonable results. It is important to stress the fact that their success ratio cannot be compared directly, since all use different data sets with different quality. Moreover a serious limitation is the use of unverified data. For the determination of the success ratio the dark objects should be classified as oil spill or look-alike, based on the optical inspection of the SAR images, done by an operator, which is not always an available approach.

It is not trivial to apply a working detection algorithm to a new sea area. In order to create an operational procedure for a particular geographic region some specifications of the water basin should be considered. Those include local topography, water density, colour, seasonal variations in alga blooms, winds and currents and the coastal borders.

In general, creating a fully automatized detection method is a challenging task, considering the long list of limitations. A semi-automatic method might be more beneficial considering the success ratio.

In this review different methods for the image segmentation procedure for extraction of dark objects are presented – adaptive and hysteresis thresholding algorithms, a method using the Laplace of Gaussian and Difference of Gaussian operators, a wavelet packet transformation, fuzzy clustering method and one, based on mathematical morphology.

For the correct classification of dark formations the extraction and computation of a set of features, different for every algorithm, is crucial. The major feature classes are presented together with some concrete examples.

Different classification procedures are presented – one based on probability methods, another using the Gaussian density function, a neural network algorithm and a fuzzy classifier. The success ratios for the different approaches are listed.

## REFERENCES

[1] Antonucci, R. *ARA& A*, 1993, **31**, 473.
[2] Antonucci, R., J.S. Miller. *ApJ*, 1985, **297**, 621.
[3] Tran, H. D. *ApJ*, 2001, **554**, L19.
[4] Tran, H. D. *ApJ*, 2003, **583**, 632.
[5] Lumsden, S. L., D. M. Alexander. *MNRAS*, 2001, **328**, L32.
[6] Shu, X.W., J.X. Wang, P. Jiang, et al. *ApJ*, 2007, **657**, 167.
[7] Nicastro, F., A. Martocchia, G. Matt. *ApJ*, 2003, **589**, L13.
[8] Petrov, G. P., I. M. Yankulova. *NewA*, 2012, **17,** 137.
[9] Marinucci, A., S. Bianchi, F. Nicastro, et al. *ApJ*, 2012, **748**, 130.
[10] Ho, L. C., J. E. Greene, A.V. Filippenko, et al. *ApJS*, 2009, **183**, 1.
[11] Tremaine, S., K. Gebhardt, R. Bender. *ApJ*, 2002, **574**, 740.
[12] Ho, L. C., A.V. Filippenko, W. L.W. Sargent. *ApJS*, 1997, **112**, 315.
[13] Panessa, F., L. Bassani. *A&A*, 2002, **394**, 435.
[14] Panessa, F., L. Bassani, M. Cappi, et al. *A&A*, 2006, **455**, 173.
[15] Bianchi, S., A. Corral, F. Panessa, et al. *MNRAS*, 2008, **385**, 195.
[16] Bassani, L., M. Dadina. *ApJS*, 1999, **121**, 473.
[17] Lamastra, A., S. Bianchi, G. Matt, et al. *A&A*, 2009, **504**, 73.
[18] Moran, E. C., A. J. Barth, L. E. Kay, et al. *ApJ*, 2000, **540**, L73.
[19] González-Martín, O., J. Masegosa, I. Márquez, et al. *ApJ*, 2009, **704**, 1570.
[20] Vasudevan, R. V., A. C. Fabian. *MNRAS*, 2009, **392**, 1124.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ

Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS

Volume 111

# TSUNAMI RADIATION PATTERN IN THE SOUTHERN AEGEAN SEA

LYUBA DIMOVA

*Department of Meteorology and Geophysics*

*Люба Димова.* РАЗПРОСТРАНЕНИЕ НА ВЪЛНИ ЦУНАМИ В ЮЖНАТА ЧАСТ НА ЕГЕЙСКО МОРЕ

В това изследване са представени различни сценарии на вълни цунами, генерирани от земетресения в южната част на Егейско море. Реконструирани са две от най-големите исторически събития на вълни цунами в цялото Средиземноморие. Допълнително са избрани и моделирани шест различни цунамигенни зони близо до островите Крит и Родос, за да се оценят максималните амплитуди и времето на пристигане на вълните цунами. За всички сценарии е моделиран правоъгълен сеизмичен източник, способен да генерира земетресение с магнитуд, равен или по-голям на регистрирания в исторически план магнитуд. Числено се моделират разпространението на вълните цунами, очакваните височини и ефектите по крайбрежните райони.

*Lyuba Dimova.* TSUNAMI RADIATION PATTERN IN THE SOUTHERN AEGEAN SEA

This paper presents several scenarios of earthquake-generated tsunami in the southern Aegean Sea. We reconstructed two of the largest historically known tsunami events in the area of the whole Mediterranean. Additionally we identified and modeled six different tsunamigenic zones near the islands of Crete and Rhodes to estimate the maximum water elevations and the arrival time of the tsunami. For all scenarios, we took into account rectangle seismic faults, capable to generate an earthquake with magnitude equal or slightly larger than the highest magnitude registered in historical times. Then we modeled numerically the tsunami wave propagation, the expected heights and some effects on the coastal areas.

***Keywords:*** tsunami, numerical methods, earthquakes
***PACS numbers:*** 91.30.Nw, 92.10.hl, 02.60.-x, 91.30.Px

---

*For contact*: Lyuba Dimova, Department of Meteorology and Geophysics, Faculty of Physics, Sofia University St. Kliment Ohridski,5, J. Bourchier Blvd., 1164 Sofia, Phone: +359 2 81 61 629, E-mail: lyuba_dimova@phys.uni-sofia.bg

# 1. INTRODUCTION

Mediterranean region suffered by numerous tsunamis in the past and many of them had catastrophic impact. Although tsunami catalogues are not complete enough regarding the early history events, the most of the tsunamis in the Mediterranean region are due to earthquakes. Some of the events are still with uncertain origin. Reconstruction of past historical events as well as hypothetical events in active seismic zones is important for the evaluation of the tsunami hazard in the region. In this study we focused on 8 tsunamigenic events in the southern Aegean Sea. Six of the tsunamigenic sources are located close to coast which is typical feature of the most sources in the Mediterranean. We took into account two more historical events that are quite important since the magnitude of both earthquakes is greater than 8. The impact of these strong events is comparable with the December 26, 2004 earthquake in Sumatra offshore. According to several catalogues and databases we built the hypothetical focal mechanism of 6 earthquakes, we calculated the geometry of the seismic faults and we reconstructed the tsunami waves referred to AD 21 July 365 event and AD 8 August 1303 event respectively for western Hellenic arc (WHA) and eastern Hellenic arc (EHA).

Aegean microplate is located in the eastern Mediterranean. Its southern boundary is the subduction zone south of Crete where the African plate submerged beneath the Aegean Sea plate. This area is among the most stressed part of the collision between the Eurasia and African plates therefore the seismicity there is the highest in the whole Mediterranean. The Hellenic arc parallels the subduction zone at a distance of about 120 km to the north. The Hellenic arc consists of two major transform faults, known as Cephalonia transform fault to the north-western ends and Rhodes transform fault to the eastern ends and several active volcanic points (Methana, Santorini, Nisyros, the Bodrum Peninsula). The seismicity along the Hellenic arc is extremely high with its shallow and intermediate depth of the earthquakes. The tsunamigenic potential of the earthquake sources is higher when the depth of the event is shallower. More than 4500 earthquakes (magnitude greater than 4.5) occurred in the area since the beginning of 20th century, while more than 100 events of tsunami waves present in the historical records since BC 1610 up to present. Fig. 1 shows earthquakes with magnitude greater than 4.5 and focal depth shallower than 100 km as well as historical tsunamigenic sources [1, 2].

# 2. TSUNAMIGENIC SOURCES

Eight potentially tsunamigenic seismic sources are selected in the region of southern Aegean Sea. As a major geotectonic structure, the Hellenic arc produces large earthquakes and tsunamis. Two of the key events in the region of Crete are discussed in the next sections. The location of both historical events is uncertain.

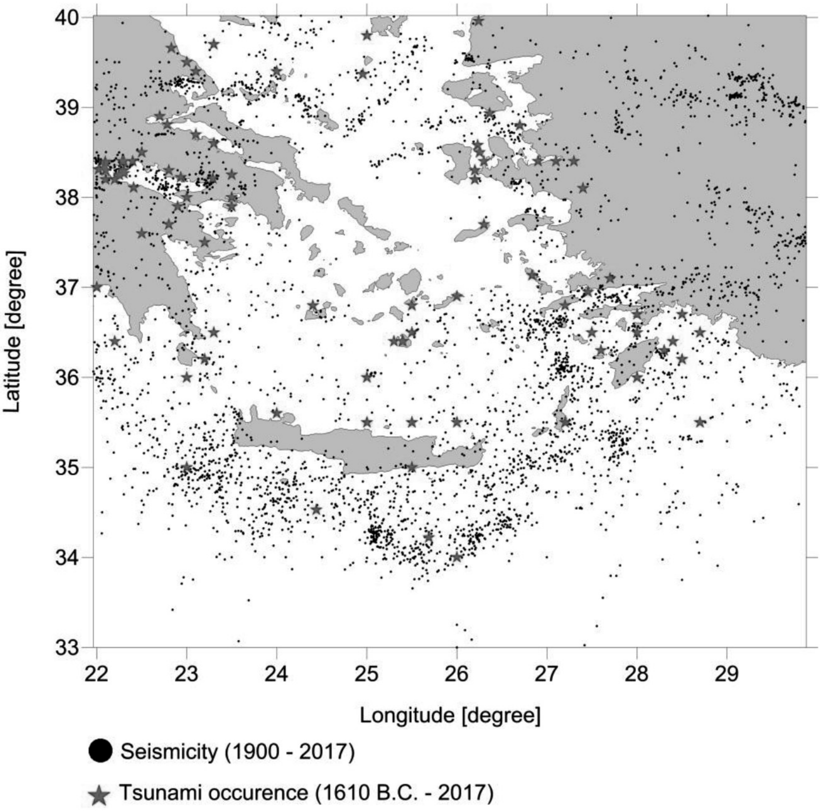


**Fig. 1.** Earthquake and tsunami occurrence in the southern Aegean Sea [1, 2]

It is considered that the large tsunamigenic earthquake occurred in AD 365 during the rupture of the western segment of the Hellenic subduction zone on a NE-dipping fault within the overriding plate [3]. This event destroyed cities and drowned thousands of people in the coastal regions from Alexandria (Egypt) to the Adriatic and Sicily. The moment magnitude $M_w$ was estimated around ~8.3. The historical, geological and archaeological evidence leave little doubt that the tsunami in AD 365 was generated by 6–9 m co-seismic uplift in the area of West Hellenic Arc (WHA) [4].

The event from AD 1303 very likely occurred during the rupturing of the arc between Crete and Rhodes islands and it is one of the largest historically discoursed tsunamigenic earthquakes in the Mediterranean. The tsunami induced by the earthquake propagated long distances from Acre (Israel) through Alexandria (Egypt) and Libya. The most affected area was the north-eastern part of the coast of Crete and the capital Heraklion located to the north. The magnitude $M_w$ is assumed to be around ~8.0. The tsunami intensity was reported to be *X* degree [4].

Many authors [e.g., 3, 5–8] simulated the large tsunamis of AD 365 and AD 1303 and the results show that the wave amplitudes are underestimated as compared to the ones expected from the historical and geological record of the tsunami. This may be due to rough bathymetry and inappropriate seismic source geometry.

Data for focal mechanisms and depth of the additional five tsunamigenic zones in the area of the southern Aegean Sea are taken into account utilizing the fault databases compiled in SHARE (The European Database of Seismogenic Faults), GreDaSS and DISS (The Database of Individual Seismogenic Sources) [9]. We simulated hypothetical earthquakes to estimate the tsunami impact for Crete and other small islands. Recent earthquake in the region of Gramvousa fault zone occurred in October 12, 2013 with $MW = 6.8$ without expecting any serious tsunami treat because of the depth of the earthquake. Yolsal and Taymaz [7] modelled several scenarios of tsunami waves generated by the earthquakes in the region of Ptolemy Trench. Karpathos Island was hit by tsunami, the eyewitness accounts the maximum tsunami run-up heights, associated with February 9, 1948 earthquake ($M \sim 7.0$) on the island of Karpathos, between 3.6– 7.0 m [10]. There are two main hypotheses on the origin of the tsunami- underwater earthquake or underwater landslide.

One more tsunamigenic zone that we studied in this work is located in the region of Rhodes. Since the seismicity there is high we chose one of the two earthquakes in 1957 with $M_w = 7.3$ and the corresponding focal mechanism [10], to explore possibility for tsunami and sequences on the nearest coasts. Fethiye town is one of the attractive tourist destinations in Turkey and was affected by earthquakes and tsunamis in the past [11].

The position of all eight tsunamigenic seismic faults is shown in Fig. 2. We took into account rectangle sources placed in different depths that respond to the upper border depth of the fault (UBDF). The proposed names of the seismic faults correspond to the databases cited before (SHARE, GreDaSS and DISS) and we named them as follow: WHA (AD 365), EHA (AD 1303), Gramvousa, Palaeochora-Tympaki, Ptolemy Trench, South Cretan Sea, Karpathos I and Rhodes.

The fault parameters of the selected sources are presented in Table 1. For each tsunamigenic zone we took into account an earthquake with magnitude equal to the largest magnitude registered in the past. The relation between the moment magnitude and the seismic moment, proposed by Hanks and Kanamori [12] is used to obtain the released energy. The seismic moment is related to the geometry of the faults and the slip on the faults by the regressions proposed in Mai and Beroza [13] and Wells and Coppersmith [14]. The longitude and latitude are converted in Universal Transverse Mercator (UTM) coordinate system since the calculations are made in Cartesian coordinates. The focal mechanisms for five of the tsunamigenic faults are selected to be in accordance with the database SHARE. Focal mechanisms for both of the reconstructed events and for Rhodes fault are taken in accordance with published data [10, 15–19].

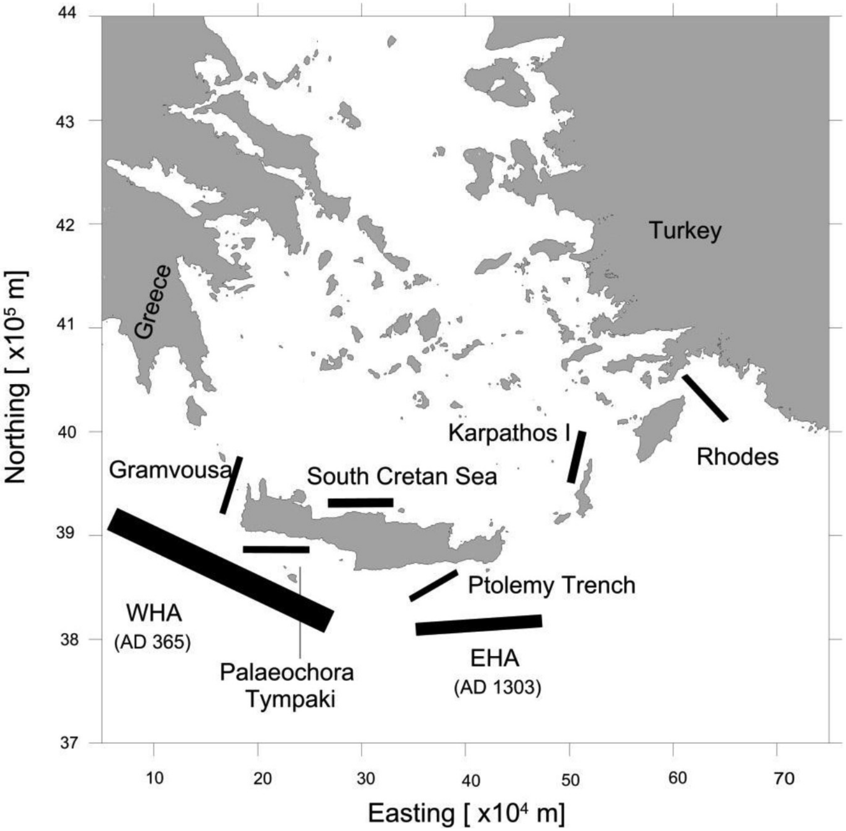


**Fig. 2.** Position of the tsunamigenic seismic sources

## 3. NUMERICAL MODELLING

The theory of numerical modelling of tsunami waves started its history during the last century but it is widely used after the 2004 tsunami in Sumatra where at least 230 000 people were killed. Numerical models are used to evaluate and predict the physical characteristics of tsunami. They play an important role in tsunami hazard mitigation and are especially useful for preparing maps of inundation for coastlines vulnerable to tsunami flooding. There are number of mathematical methods based on the resolution of bathymetry data and parameters of the seismic fault, developed to model and identify tsunami wave characteristics. Tsunami source models are well developed for submarine earthquakes, where the seafloor deformation is caused by a finite dislocation (i.e., slip on the fault) [20]. Simple implementation of this type of source model usually assumes a finite rupture interface with uniform dislocation (e.g., slip movement along a fault). We computed the initial conditions of studied events implementing analytical formulas of Okada [20, 21] and appropriate bathymetry data. The basic parameters for each event described in Table 1 are essential for calculation of deformation of the sea bottom which specifies the initial tsunami waveform. Numerical code UBO-TSUFD is applied to model the propagation and the interaction of tsunami with the coastal areas. The code is developed in University of Bologna (Italy) and extensively used to study the propagation of tsunamis generated by earthquakes. We utilized this method since it was tested and validated for different benchmark problems based on analytical solutions: one on a plane wave propagating on a flat channel with a constant slope beach, and one on a laboratory experiment. Additionally it was proved for several realistic tsunami cases [22]. UBO-TSUFD reproduced quite well the theoretical and experimental data. The code solves the governing equations of Navier-Stocks in the shallow water theory approximation by using the finite-difference technique with an explicit leap-frog scheme on a staggered grid. Tsunami waves belong to the long-wave theory, therefore the vertical acceleration of water particles are negligible compared to the gravitational acceleration except for an oceanic propagation of tsunami [23]. A good approximation is that the pressure is hydrostatic and therefore the vertical motion of water particles has no effect on the pressure distribution. In addition the horizontal velocity of water particles are vertically uniform. There are several conditions that need to be fulfilled in order to have good results- initial conditions, boundary conditions, stability conditions, preparation of a grid and bathymetry data. Linear and nonlinear theories are used according to the degree of nonlinearity of the phenomena. Usually coarse grids in the deep ocean and fine grids in the nearshore zone are used. An expanded explanation of UBO-TSUFD can be found in [22, 23]. The code is extensively applied to simulate tsunami waves for seas like Mediterranean and other small basins, where the Coriolis force is neglected.

The code computes water surface oscillations, velocity field, arrival time of the waves and water inundation when nonlinearity is applied. The grid used in our calculations has cells with 500×500 m resolution. The utilized time step is 1.5 s and it is adjusted to satisfy the CFL (Courant-Friedrichs-Lewy) stability condition (1):

$$\frac{\Delta x}{\Delta t} = \sqrt{2gH}\ , \tag{1}$$

where $\Delta x$ is the cell size, $\Delta t$ is the time step for every cell, $g$ is the gravity acceleration and $H$ is the maximum depth of the ocean, in our case for the Mediterranean region it is 4500 m.

The total number of nodes of the grid is 2803401. The resolution of the bathymetry and topography data is 30 s [24]. Bathymetry map of the studied region is shown in Fig. 3.

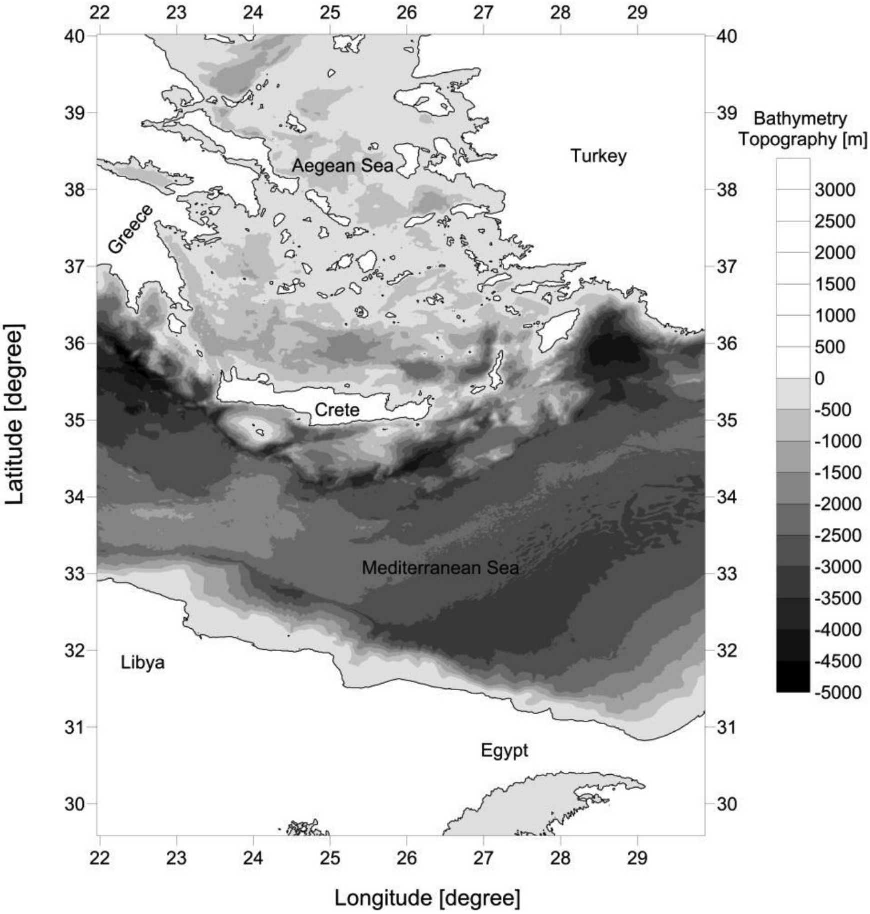


**Fig. 3.** Bathymetry map of the region

**Table 1.** Seismic fault parameters

| Tsunamigenic zone | Position [lon, lat] | Magnitude Mw | UBDF* [km] | Length [km] | Width [km] | Slip [m] | Strike [degree] | Dip [degree] | Rake [degree] |
|---|---|---|---|---|---|---|---|---|---|
| WHA AD 365 | 23.0, 35.0 | 8.3 | 5 | 200 | 80 | 9 | 295 | 15 | 90 |
| EHA AD 1303 | 26.1, 34.5 | 8.0 | 4 | 125 | 50 | 6 | 268 | 48 | 71 |
| Gramvousa | 23.5, 35.5 | 6.8 | 2 | 25 | 14 | 2 | 180 | 80 | 270 |
| Palaeochora-Tympaki | 24.0, 35.2 | 7.0 | 2 | 33 | 15 | 2.5 | 100 | 75 | 270 |
| Ptolemy Trench | 25.3, 34.8 | 7.0 | 2 | 33 | 17 | 2.5 | 75 | 65 | 270 |
| South Cretan Sea | 24.8, 35.5 | 7.1 | 2 | 40 | 19 | 2.7 | 270 | 50 | 270 |
| Karpathos I | 27.1, 35.9 | 7.0 | 2 | 33 | 17 | 2.5 | 200 | 80 | 240 |
| Rhodes | 28.7, 36.4 | 7.3 | 3 | 50 | 23 | 3.2 | 328 | 73 | 71 |

*UBDF – Upper Border Depth of the Fault

## 4. RESULTS OF THE TSUNAMI SIMULATIONS

The tsunami initial condition for scenario WHA is presented in the upper panel of Fig. 4. The source is considered as thrust fault and it is placed almost parallel to the western Hellenic Subduction zone. This fault generated a sea bottom deformation characterized by high uplift near southwestern Crete and a subsidence near the northwestern coasts of Crete. The maximum positive and negative vertical deformations are 4.1 m and –1.6 m, respectively. The initial condition of scenario regarding the EHA thrust faulting source is presented in the upper panel of Fig. 4. The source, whose parameters are listed in Table 1, generated initial maximum and minimum sea water elevations of 3.3 m and –0.4 m, respectively. This scenario is associated with $M_w \sim 8.0$ and it is located offshore the southeastern coastline of Crete. The vertical displacements of the sea bottom for other modeled tsunamigenic sources are plotted in Fig. 4 (lower panel). South Cretan Sea tsunamigenic fault induced minimum water elevations –1.5 m, while Rhodes fault generated the maximum displacement of 1.5 m near the southwestern Turkish coast.

The propagation of the tsunami waves for all scenarios is presented as plots at 5, 20, 40 and 60 minutes intervals after the tsunami onset. The tsunami radiation pattern simulated for the two key events AD 365 and AD 1303 is illustrated in Fig. 5. The upper panel describing the tsunami propagation indicates that after only 5 minutes the waves have already attacked the southeastern coast of Crete. In 20 minutes the tsunami propagates northeast toward Karpathos and southwest toward the coasts of Crete. The Turkish and the Libyan coasts are reached by initially positive tsunami waves 40 minutes after the earthquake. One hour after the tsunami onset peninsula Peloponnese and the whole southern Aegean Sea is affected. The lower panel of Fig. 5 presents the tsunami radiation pattern simulated for AD 1303 event. The southwestern part of Crete, the islands Kythira and Gavdos are affected by the strong positive waves in 5 minutes after the earthquake. The propagation spreads in southwestern direction toward Libya and Ionian Sea. Most of the small islands north of Crete are attacked in 40 minutes. In one hour the eastern part of Mediterranean, including Alexandria and Cyprus is inundated. Both of the events are considered to be the largest historical tsunamis of tectonic origin in the region of Mediterranean. Tsunami radiation pattern modelled by other tsunamigenic sources is presented in Fig. 6 and Fig.7. Since the magnitude and the geometry of the six sources are smaller than the reconstructed events above, as expected, the tsunami waves are not so violent. The morphology beneath the southern Aegean Sea is very complex and therefore the propagation is slower.

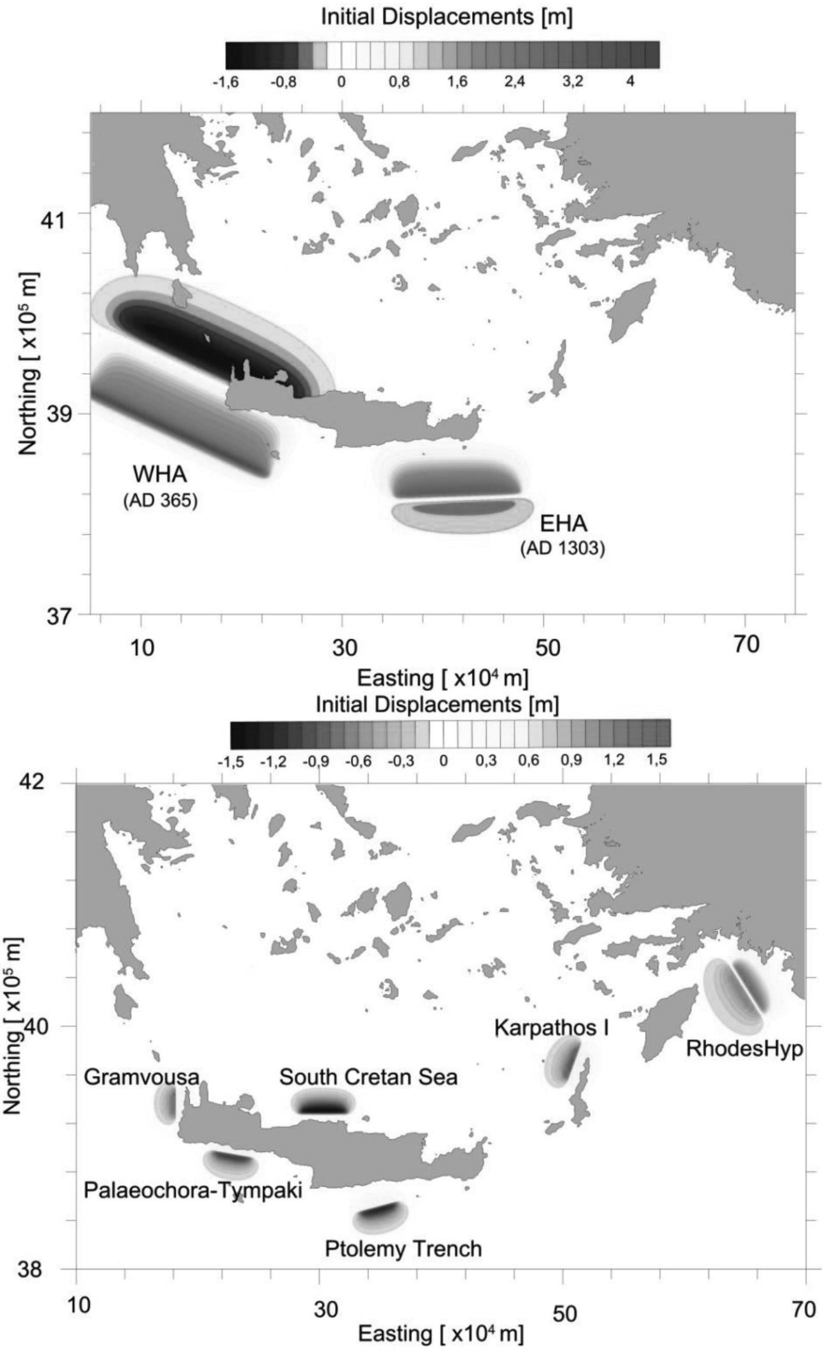


**Fig. 4.** Initial displacements (upper panel – WHA and EHA, lower panel – Gramvousa, Palaeochora-Tympaki, Ptolemy Trench, South Cretan Sea, Karpathos I and Rhodes)

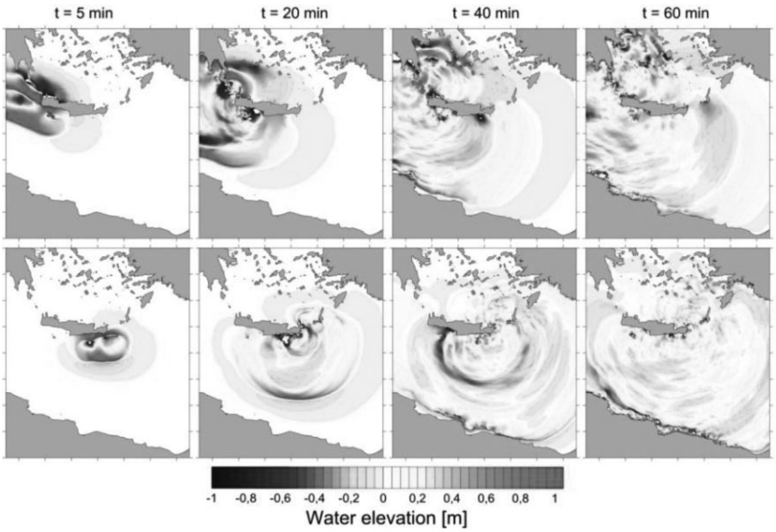


**Fig. 5.** Radiation pattern of the computed tsunami elevations for WHA and EHA

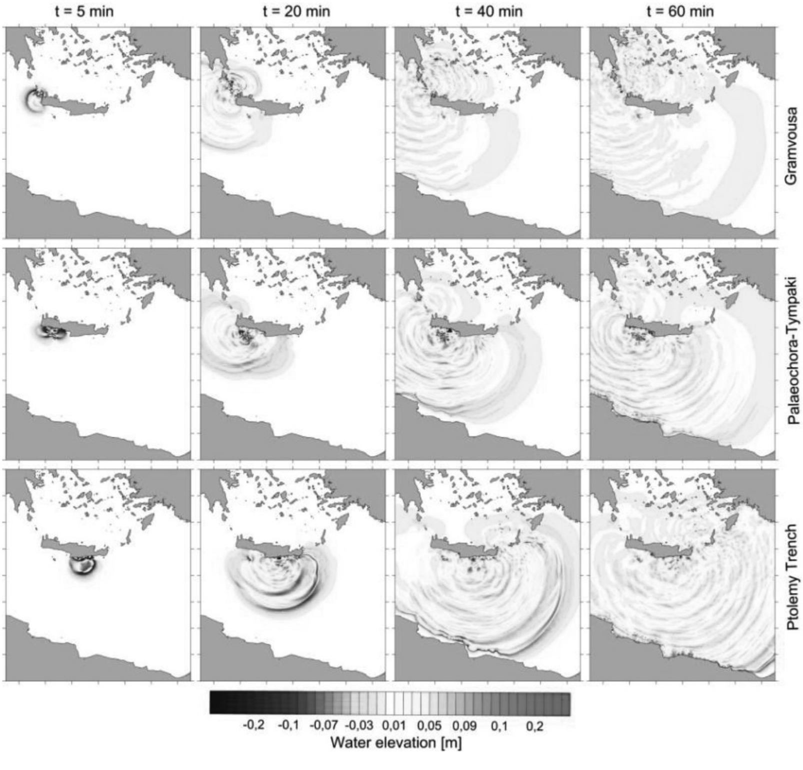


**Fig. 6.** Radiation pattern of the computed tsunami elevations for Gramvousa, Palaeochora-Tympaki and Ptolemy Trench

The effects can be considered as local, near the sources. Due to the large number of islands in the Aegean Sea, the tsunami waves diffract and this is clearly illustrated in both of the figures, and especially in the case of South Cretan Sea simulation (Fig. 7, upper panel). The propagation of the tsunami due to Rhodes source affects mostly the southwestern coast of Turkey, in particular the gulf of Fethiye and the whole Rhodes in less than 20 minutes after the earthquake.

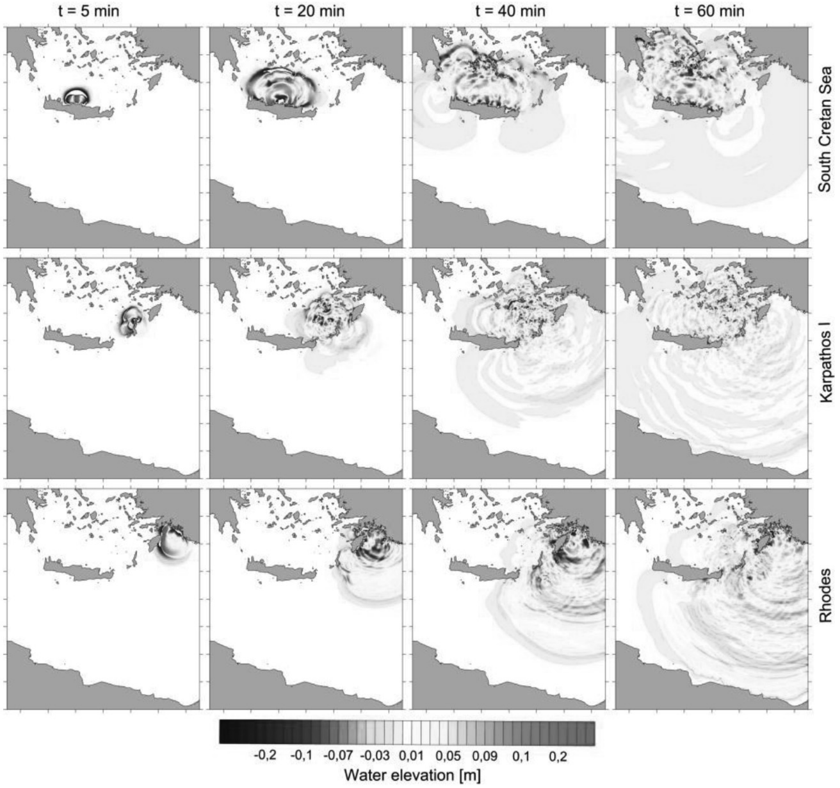


**Fig. 7.** Radiation pattern of the computed tsunami elevations for South Cretan Sea, Karpathos I and Rhodes

Fig. 8 and Fig. 9 show the maximum positive water elevation due to the events. The tsunami heights field can be interpreted as an indication of the released tsunami energy. Figures clearly illustrate the most affected coastal areas. Severe tsunami effects are expected on the west, south and east coast of Crete due to AD 365 and AD 1303 events. Large part of the tsunami energy computed for WHA propagates toward the western part of the Mediterranean, while EHA tsunamigenic source extend its energy to the south, reaching the coasts of Libya and Egypt and to the north attacking the island of Karpathos (Fig. 8).

As expected the maximum positive tsunami heights computed for the rest of the scenarios are smaller. The direction of the tsunami energy propagation strongly depends on the geometry of sources. Tsunami effects are expected in the region of whole Crete, southern Peloponnese, the islands north of Crete, Karpathos, Rhodes and southwestern part of the Turkish coast, including the cities Marmaris and Fethiye (Fig. 9).

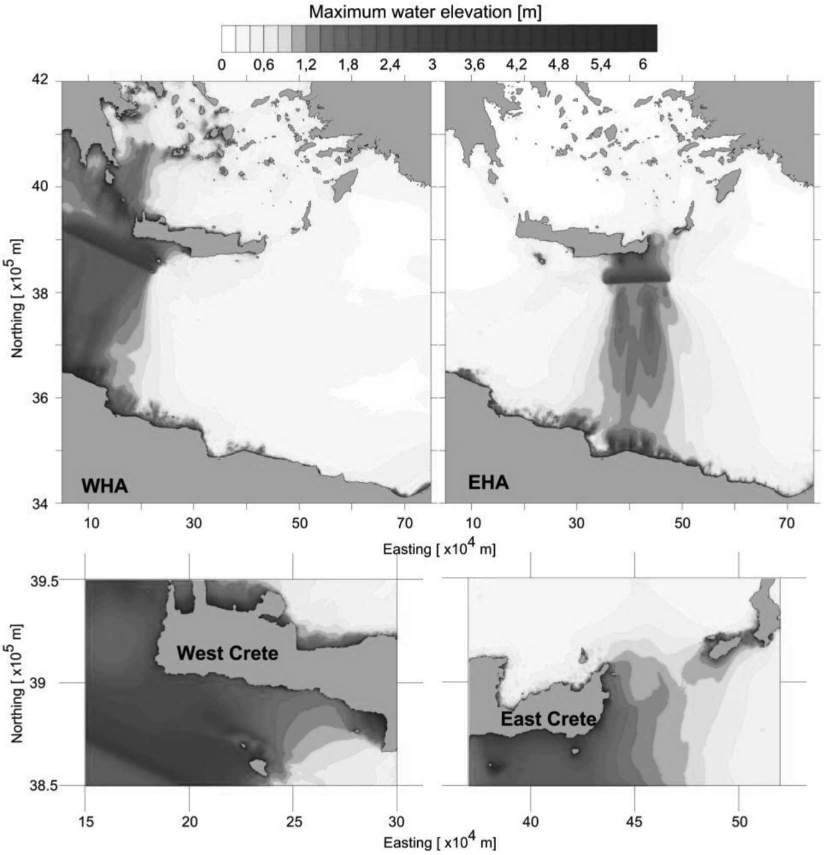


**Fig. 8.** Maximum positive tsunami wave heights computed for the WHA and EHA scenarios

The contribution of all modeled scenarios in southern Aegean Sea is presented in Fig 10. The colored areas represent the maximum tsunami wave heights. The figure shows the geographical distribution of the tsunami impact from studied events. It is evident that events like AD 365 (WHA) and AD 1303 (EHA) give the highest impact, but also Rhodes tsunamigenic zone must not be neglected, since the position of the source is very close to the coast and therefore the time for cities' evacuation like Fethiye is less than 10 minutes. South Cretan Sea source is the scenario contributing most to the impact of the islands in southern Aegean Sea.

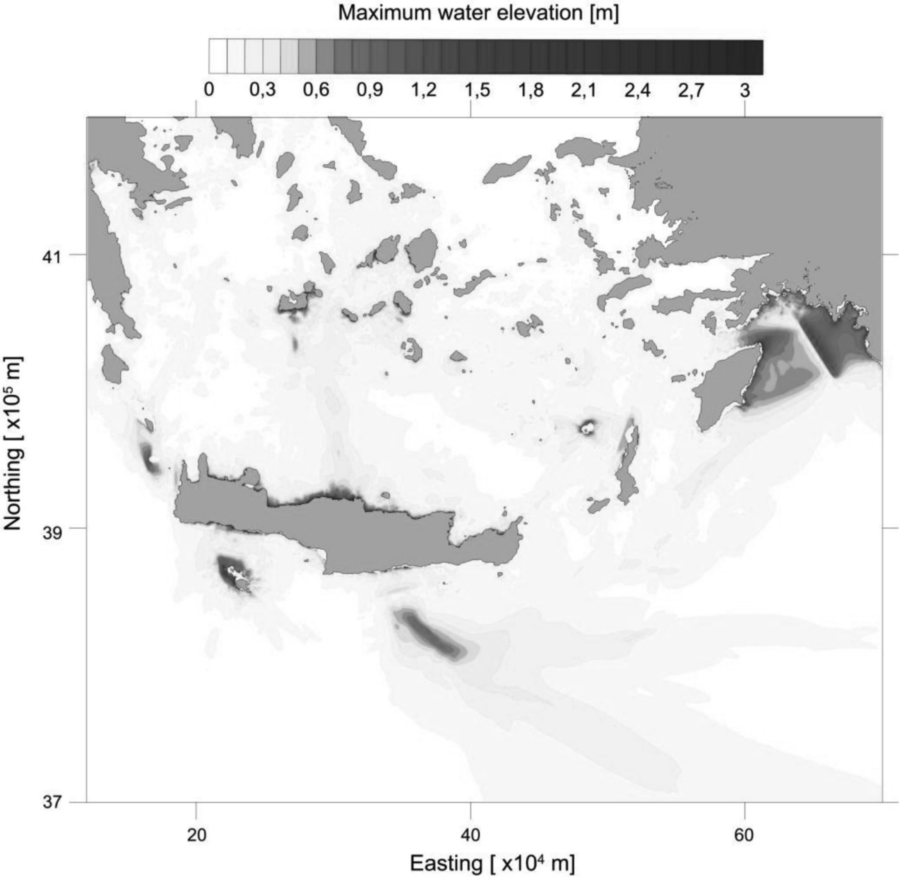


**Fig. 9.** Maximum positive tsunami wave heights computed for Gramvousa, Palaeochora-Tympaki, Ptolemy Trench, South Cretan Sea, Karpathos I and Rhodes scenarios

Synthetic mareograms in different locations are computed by the code. The mareograms reproduce the variations in the sea water elevation depending on the time elapsed from the earthquake onset. The position of considered mareograms is plotted in Fig. 11. Mareograms modeled for the scenario of South Cretan Sea tsunamigenic source are located in Santorini (Thera), Panormos and Heraklion and are presented in Fig. 12. The figure shows water level for 1 hour and 40 minutes after the earthquake occurrence.

Maximum modeled tsunami wave heights near Santorini are about 1.2 m (Fig. 12a), in Panormos are about 2.3 m (Fig. 12b) and in Heraklion the maximum water elevation is about 1.8 m (Fig. 12c).

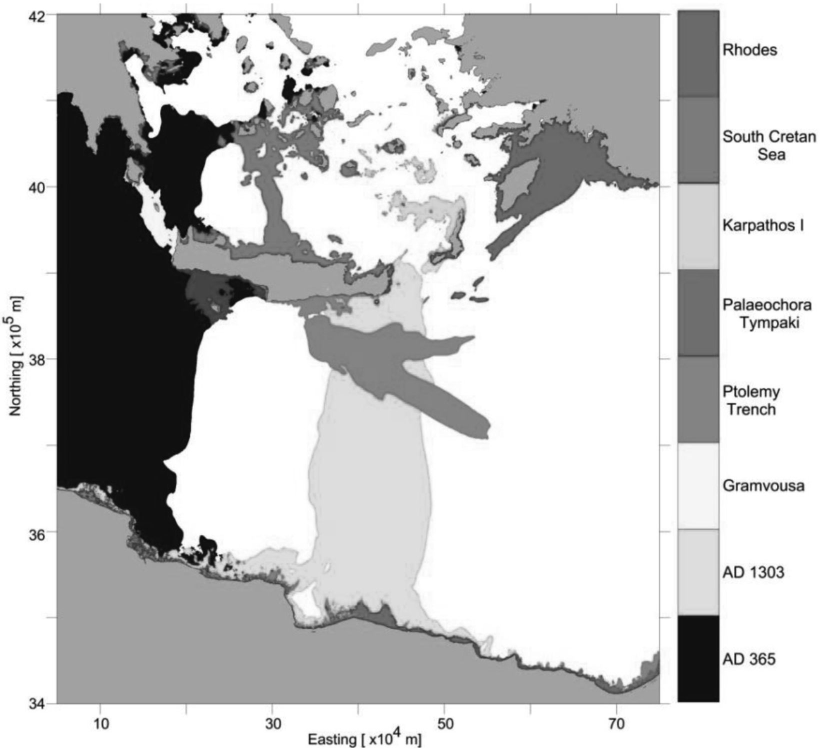


**Fig. 10.** Contribution given by all scenarios according to the computed tsunami wave heights

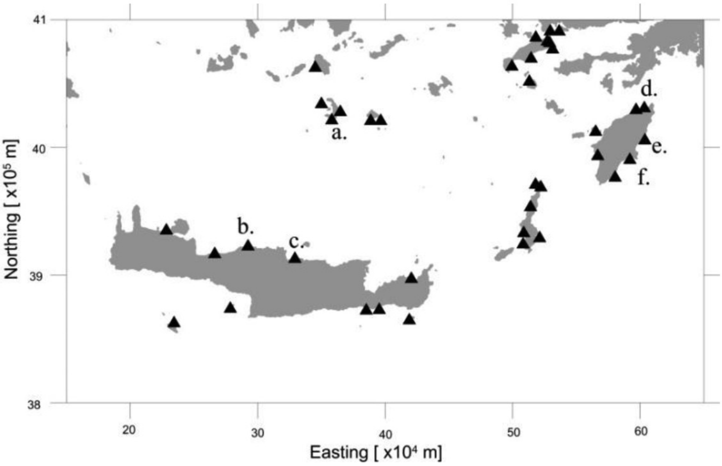


**Fig. 11.** Position of synthetic mareograms. The mareograms in labeled locations are shown in Figures 12 and 13

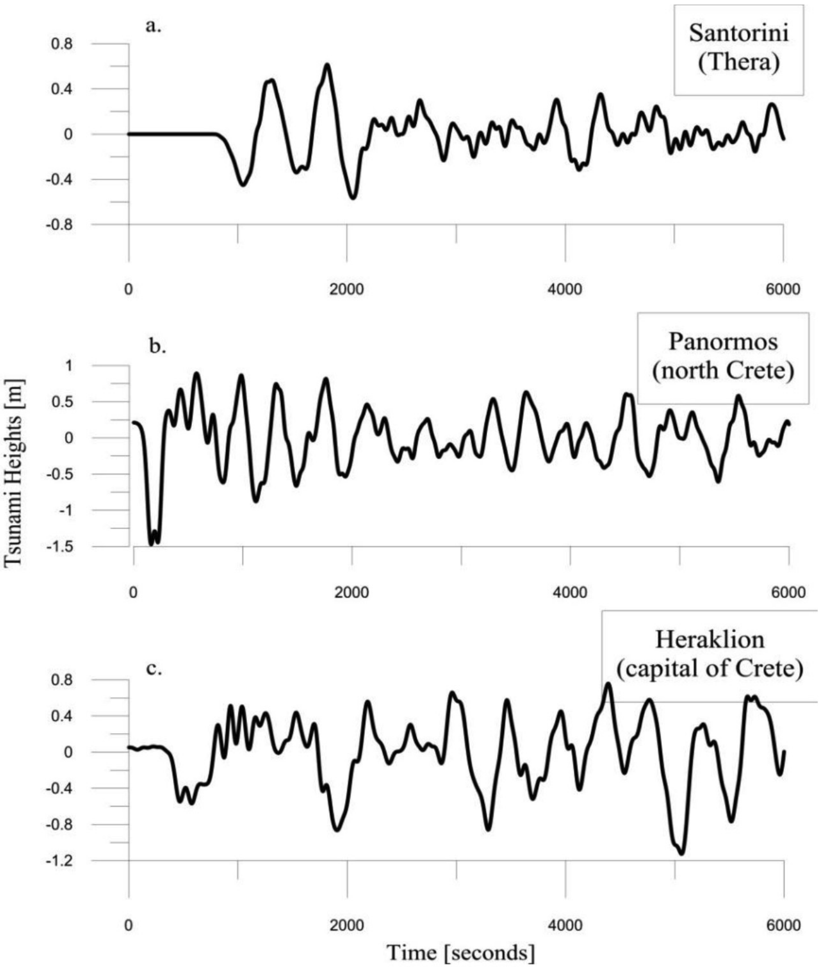


**Fig. 12.** Synthetic mareograms computed for scenario South Cretan Sea. Locations of mareograms are shown in Fig. 11

Synthetic mareograms calculated for the scenario of Rhodes tsunamigenic source are located in Rhodes, Stegna and Lardos and are illustrated in Fig. 13. Maximum modeled tsunami wave heights near Rhodes are about 2.5 m (Fig. 13d), in Stegna the water elevation is about 2.0 m (Fig. 13e) and in Lardos the maximum water amplitude is about 2.8 m (Fig. 13f). The first arrived wave is negative which corresponds to the initial displacements of the source. The eastern part of the island is the most affected due to Rhodes tsunamigenic fault. The period of the waves is much shorter than the period of the oscillations for scenario South Cretan Sea.

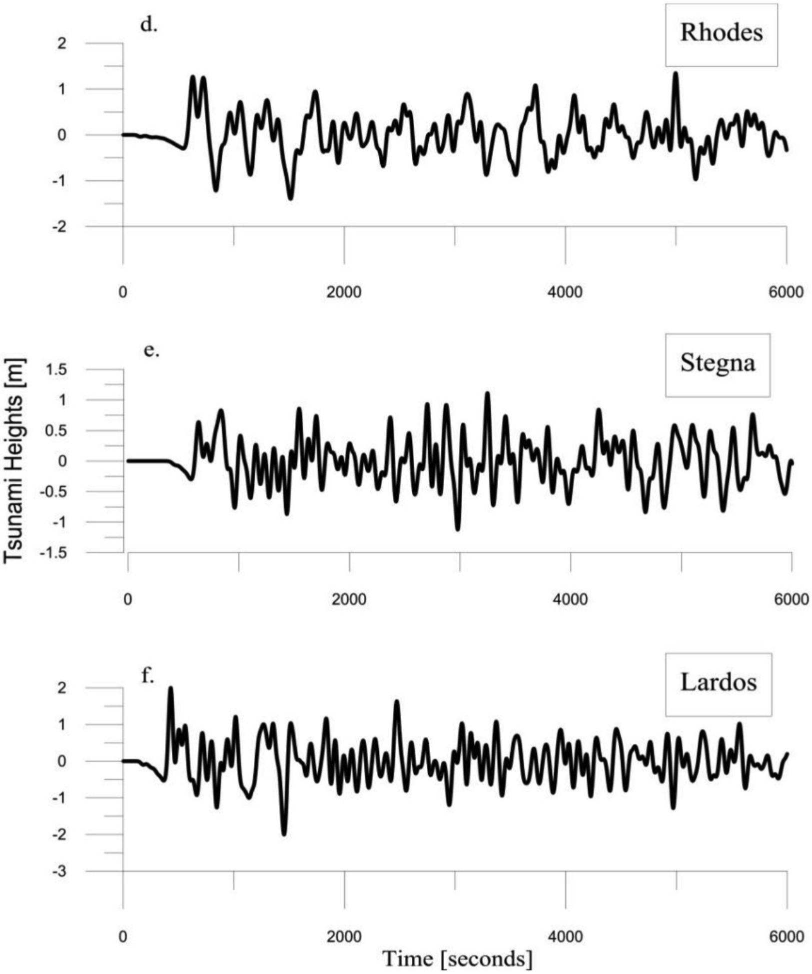


**Fig. 13.** Synthetic mareograms computed for scenario Rhodes. Locations of mareograms are shown in Fig. 11

## 5. CONCLUSIONS

We examined scenarios of tsunami radiation pattern from eight seismic sources in the southern Aegean Sea. Based on numerical simulations the largest historically known tsunamis in the Mediterranean AD 365 and AD 1303 are well reconstructed. Most of the studied sources are situated close to the coasts which results in less time for evacuation. Crete, Rhodes, Karpathos and the southwestern part of the Turkish coast are at high risk of inundation due to tsunami waves.

**Acknowledgements.** The author thanks the Tsunami Research Team from University of Bologna, Department of Physics and Astronomy, Sector Geophysics, for suggestions and possibility to work with UBO-TSUFD.

## REFERENCES

[1] International Seismological Centre. *On-line Bulletin*, http://www.isc.ac.uk, Internatl. Seismol. Cent., Thatcham, United Kingdom, 2015.

[2] National Geophysical Data Center / World Data Service (NGDC/WDS): Global Historical Tsunami Database. National Geophysical Data Center, NOAA. doi:10.7289/V5PN93H7.

[3] Shaw, B., N. Ambraseys, P. England, M. Floyd, G. Gorman, T. Higham, J. Jackson, J. Nocquet, C. Pain, M. Piggott. *Nature Geocsience*, 2008, **1**, 268.

[4] Papadopoulos, G. A., E. Gràcia, R. Urgeles, V. Sallares, P. M. De Martini, D. Pantosti, M. González, A. Yalciner, J. Mascle, D. Sakellariou, A. Salamon, S. Tinti, V. Karastathis, A. Fokaefs, A. Camerlenghi, T. Novikova, A. Papageorgiou. *Marine Geology*, 2014, **354**, 81.

[5] Tinti, S., A. Armigliato, G. Pagnoni, F. Zaniboni. *ISET Journal of Earthquake Technology*, 2005, **42**, 4, 171.

[6] Lorito, S., M. Tiberti, R. Basili, A. Piatanesi, G. Valensise. *Journal of Geophysical Research*, 2008, **113**, B01301.

[7] Yolsal-Cevikbilen, S., T. Taymaz. *Tectonophysics*, 2012, **536-537**, 61.

[8] Hamouda, A. *Journal of African Earth Sciences*, 2006, **44**, 37.

[9] Basili, R., G. Valensise, P. Vannoli, P. Burrato, U. Fracassi, S. Mariano, M.M. Tiberti, E. Boschi. *Tectonophysics*, 2008, **453**, 1, 20.

[10] Ebeling C., E. Okal, N. Kalligeris, C. Synolakis. *Tectonophysics*, 2012, **530-531**, 225.

[11] Yalciner, A. C., C. Ozer, H. Karakus, A. Zaytsev, I. Guler. In: *Proceedings of the 32nd International Conference Coastal Engineering, China,* 2010, https://doi.org/10.9753/icce.v32.management.10.

[12] Hanks, T., H. Kanamori. *Journal of Geophysical Research*, 1979, **84**, B5, 2348.

[13] Mai, M., G. Beroza. *Bull. of the Seis. Soc. of America*, 2000, **90**, 3, 604.

[14] Wells, D., K. Coppersmith. *Bull. of the Seis. Soc. of America*, 1994, **84**, 4, 974.

[15] Papadopoulos, G.A., E. Daskalaki, A. Fokaefs, N. Giraleas. *Natural Hazards and Earth System Sciences*, 2007, **7**, 57.

[16] Papadopoulos, G.A., A. Fokaefs. *ISET Journal of Earthquake Technology*, 2005, **42**, 4, 159.

[17] Yolsal, S. Source Mechanism Parameters and Slip Distributions of Crete-Cyprus Arcs, Dead Sea Transform Fault Earthquakes and Historical Tsunami Simulations. PhD Thesis, 523 pages, İstanbul Technical University, İstanbul, Turkey, November 2008.

[18] Guidoboni, E., A. Comastri. Catalogue of Earthquakes and Tsunamis in the Mediterranean area from the 11th to the 15th Century. INGV-SGA, p. 1037, Bologna, 2005.

[19] Ambraseys, N., R. Adam. The Seismicity of Egypt, Arabia and the Red Sea: A Historical Review, Cambridge University Press, Cambridge, UK, 1994.

[20] Okada, Y. *Bull. of the Seis. Soc. of America*, 1985, **75**, 4, 1135.

[21] Okada, Y. *Bull. of the Seis. Soc. of America*, 1992, **82**, 2, 1018.

[22] Tinti, S., R. Tonini. *Natural Hazards and Earth System Sciences*, 2013, **13**, 1759.

[23] Dimova, L., R. Raykova. *Ann. of Sofia Univ. "St. Kliment Ohridski", Faculty of Physics,* 2016, **109**, 24.

[24] British Oceanographic Data Centre. Gridded Bathymetry Data Sets. https://www.bodc.ac.uk/data/hosted_data_systems/gebco_gridded_bathymetry_data/.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# COMPUTER AIDED INVESTIGATION OF THE DOMINANT MODE ON THE HOMOGENEOUSLY MAGNETIZED FERRITE IMAGE GUIDE

ILIYANA ARESTOVA

*Department of Radiophysics and Electronics*

*Илияна Арестова.* ЧИСЛЕНО ИЗСЛЕДВАНЕ НА ОСНОВНИЯ МОД В ХОМОГЕННО НАМАГНИТЕН ФЕРИТЕН ОГЛЕДАЛЕН ВЪЛНОВОД

Феритният огледален вълновод (ФОВ) е неотменима съставна част на ферито-диелектричните огледални структури, предназначени за приложение в милиметровия диапазон. Тези структури се разглеждат като основа за проектиране на управляващи устройства – изолатори, циркулатори, превключватели и др., при различни намагнитвания на феритния елемент. Тук е изследвано числено по метода на крайните елементи (МКЕ) разпределението на компонентите на полето на основния мод при три различни посоки на хомогенно намагнитване на феритния елемент: 1) намагнитване, перпендикулярно на посоката на разпространение и успоредно на огледалната равнина; 2) намагнитване, перпендикулярно на посоката на разпространение и на огледалната равнина; 3) намагнитване, успоредно на посоката на разпространение. Изследването показва, че във втория случай на намагнитване се наблюдава асиметрия на електричното поле на основния мод, което може да доведе до невзаимен ефект.

*Iliyana Arestova.* COMPUTER AIDED INVESTIGATION OF THE DOMINANT MODE ON THE HOMOGENEOUSLY MAGNETIZED FERRITE IMAGE GUIDE

The ferrite image guide (FIG) is an essential element of the ferrite-dielectric structures intended for use in the millimetre wavelength range. These structures are considered a basis for designing control devices – isolators, circulators, switches, etc. using various magnetizations of the ferrite

---

*For contact:* Iliyana Arestova, Department of Radiophysics and Electronics, Faculty of Physics, Sofia University „St. Kliment Ohridski”, 5 James Bourchier Blvd., Sofia 1164, Phone: +359 2 8161 724, E-mail: ilar@phys.uni-sofia.bg

element. Here the distribution of the electromagnetic field components of the dominant mode is numerically studied by finite element method (FEM). Three different directions of homogeneous magnetization of the ferrite element have been considered: 1) magnetization perpendicular to the direction of propagation and parallel to the ground plane (Case 1); 2) magnetization perpendicular to the direction of propagation and to the ground plane (Case 2); 3) magnetization parallel to the direction of propagation (Case 3). The investigation has shown that the most promising for nonreciprocal behaviour is Case 2 due to the asymmetry of the electric field distribution

**Keywords:** millimeter waves, ferrite devices, image guide, finite element method.
**PACS number:** 84.40.Az

## 1. INTRODUCTION

The ferrite image guide (FIG) together with the dielectric image guide structures represents a possible basis for the design of control devices for millimetre waves [1–7]. Different homogeneous magnetizations of the ferrite element have been investigated [1–6], as well as the inhomogeneous (mixed) magnetization [7]. The mixed magnetization is produced by a disk shaped permanent magnet and it contains two opposite longitudinal components at both ends of the ferrite element and a transverse component perpendicular to the ground plane in the middle of it. In order to clarify the processes taking place at this mixed magnetization, it is reasonable to study separately all three mutually perpendicular homogeneous magnetizations of the FIG.

A detailed study of the modes on the dielectric image guide has been done earlier in [8]. Here the FIG is modelled at three different directions of homogeneous magnetization. The field components of the dominant mode have been obtained numerically by finite element method (FEM) at each magnetization. The purpose of this research is to study the impact of each direction of magnetization on the electric and magnetic field components of the dominant mode on the magnetized FIG. This could improve understanding of the operating mechanism of the nonreciprocal ferrite-dielectric structures with inhomogeneous magnetization.

## 2. MODEL AND MAGNETIZATION

The investigated rectangular FIG has cross-sectional dimensions $a \times b$ = 1.82×0.99 mm$^2$ (Fig. 1) and is made of ferrite mark 1СЧ4 (Russia). The ferrite has the following electromagnetic parameters: relative permittivity $\varepsilon_r = 11.1$, tangent of the angle of dielectric losses tg$\delta_\varepsilon = 0.01$ and saturation magnetization $4\pi M_s = 0.463$ T. The homogeneous magnetization in the three mutually perpendicular directions with a field strength $H = 40$ kA/m has been set successively.

Some of results obtained numerically by FEM, namely the dispersion characteristic and the distribution of the electric field magnitude, have been reported in [9]. Here we have first defined points inside the model volume, situated at those transverse planes, at which maxima and minima of the electric field magnitude have been observed. After that we have calculated the values of all six electric and magnetic field components at each defined point. The resulting distributions of field components of the dominant mode along transverse axes $x$ and $y$ are shown in Figs. 2–7. The values of the electric and magnetic field components in Figs. 2–7 correspond to the transmitted power of 1 W.

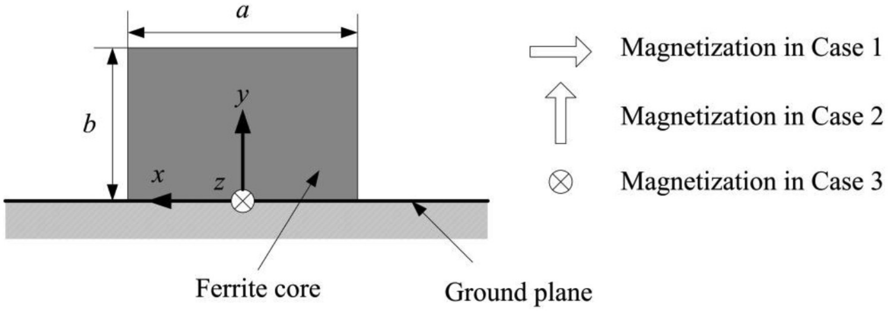


**Fig. 1.** The cross-sectional view of the FIG

## 3. RESULTS AND DISCUSSION

### 3.1. FIELD COMPONENTS AT MAGNETIZATION PERPENDICULAR TO THE DIRECTION OF PROPAGATION AND PARALLEL TO THE GROUND PLANE (CASE 1)

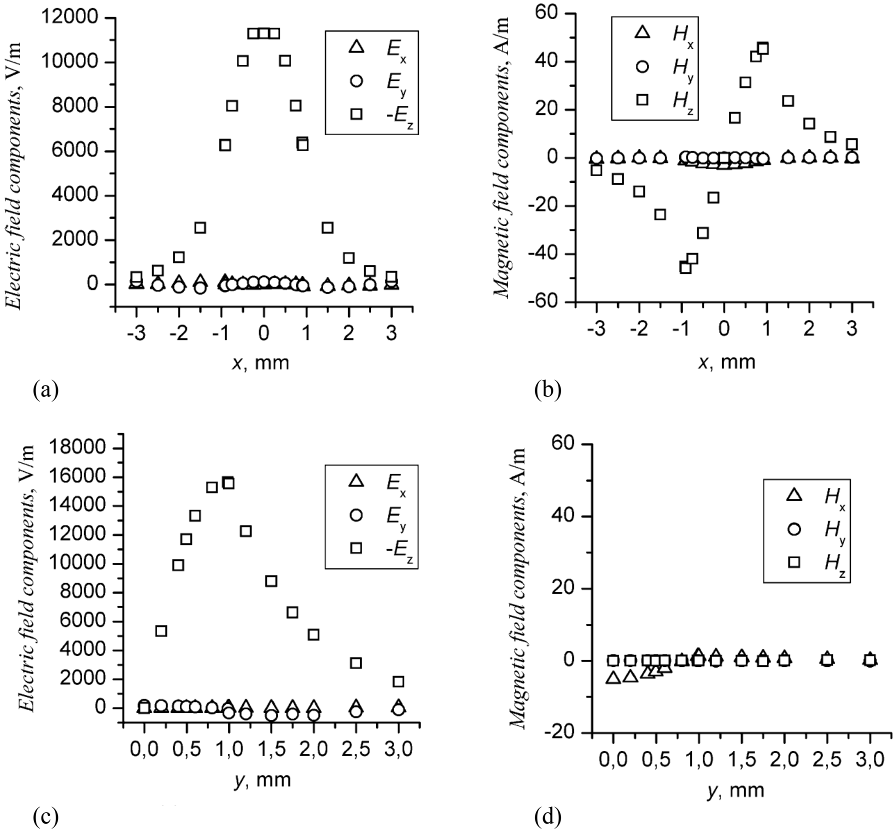
A maximum of the electric field magnitude distribution has been observed at $z = z_1 = 1.66$ mm, and a minimum at $z = z_2 = 3.34$ mm. The distributions of the six components of the dominant mode at these two transverse planes are given in Fig. 2 and Fig. 3 respectively. It can be seen that the main components of the dominant mode are $E_y$, $E_z$ and $H_x$. There is a well-defined symmetry in the distributions about $x = 0$ (Fig. 2a, Fig. 3a,b).

In accordance with the boundary conditions the component $E_y$ at $y = b = 0.99$ mm (Fig. 3c) and the component $E_x$ at $x = \pm a/2 = \pm 0.91$ mm (Fig. 3a) have interruptions. The component $H_z$ (Fig. 2b) have relatively weak extrema at $x = \pm a/2 = \pm 0.91$ mm and a maximum at $x = 0$ (Fig. 3a).

The resulting distributions show that the dominant mode on the FIG at magnetization perpendicular to the direction of propagation and parallel to the ground plane represents $E^y_{11}$ according to the Markatili's classification [10].

## 3.2. FIELD COMPONENTS AT MAGNETIZATION PERPENDICULAR TO THE DIRECTION OF PROPAGATION AND THE GROUND PLANE (CASE 2)

At this magnetization a minimum of the electric field magnitude has been observed at $z = z_1 = 3.28$ mm, and a maximum at $z = z_2 = 5.03$ mm. The distributions of the six components of the dominant mode at these two transverse planes are given in Fig. 4 and Fig. 5 respectively.



(a) (b)

(c) (d)

**Fig. 2.** Simulated distributions of the field components of the dominant mode at $z = 1.66$ mm: (a, b) $y = 0.495$ mm; (c, d) $x = 0$ mm

As can be seen from Fig. 4a, there is a shift of the maximum of the component $E_y$ from the centre of the dielectric core ($x = 0$) to the left side wall located at $x = -a/2 = -0.91$ mm. The same figure shows the presence of a significant component $E_x$, which is symmetrical about $x = 0$. The comparison of Fig. 4c with Fig. 3c shows also a greater value of the component $E_x$ in this case of magnetization. As in the previous case of magnetization, the component $E_z$ has large values in a cross-section corresponding to a maximum of the electric field magnitude (Fig. 5a,c). It has a distribution which is symmetrical about $x = 0$ (Fig. 5a).

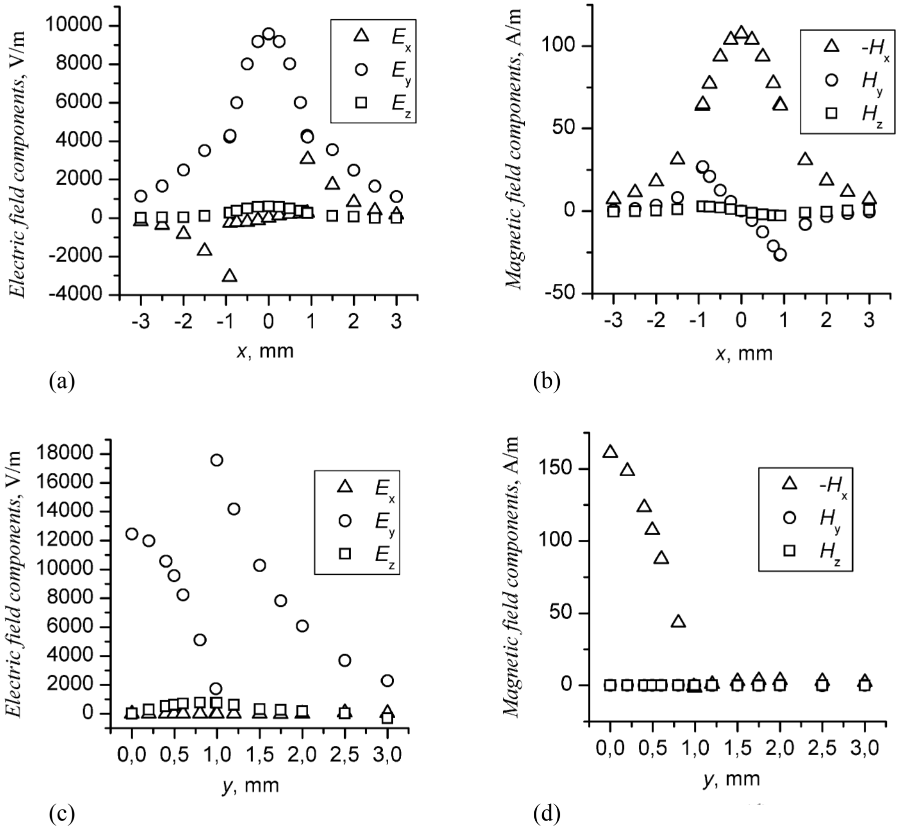
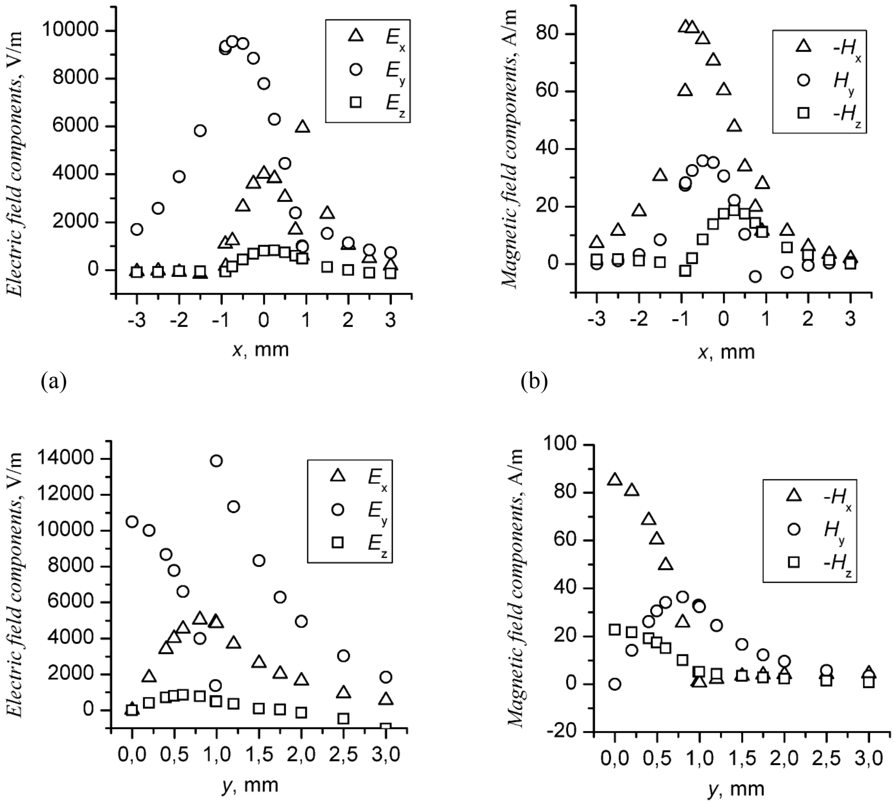


**Fig. 3.** Simulated distributions of the field components of the dominant mode at $z = 3.34$ mm: (a, b) $y = 0.495$ mm; (c, d) $x = 0$ mm

From Fig. 4b,d and Fig. 5b,d it can be seen that among the magnetic components the largest and similar in value are the components $H_x$ and $H_z$. Thus, we can conclude that the main components of the dominant mode are $E_y$, $E_z$, $H_x$ and $H_z$. The presence of an asymmetry in the transverse plane about the dielectric core centre $x = 0$, the occurrence of an additional main magnetic component $H_z$, as well as significant levels of component $E_{x,}$ require differentiation from the dominant mode in the dielectric image guide $E^y_{11}$. The dominant mode at this magnetization could be called a modified mode $E^y_{11}$, characterized by asymmetry and increased components $H_z$ and $E_x$.



(a) (b)

(c) (d)

**Fig. 4.** Simulated distributions of the field components of the dominant mode at $z = 3.28$ mm: (a, b) $y = 0.495$ mm; (c, d) $x = 0$ mm

### 3.3. FIELD COMPONENTS AT MAGNETIZATION PARALLEL TO THE DIRECTION OF PROPAGATION (CASE 3)

The distributions of the six components of the dominant mode in the case of longitudinal magnetization at two transverse planes, where a maximum and a minimum of the electric field magnitude occur, are given in Fig. 6 and Fig. 7 respectively. As can be seen from Fig. 6a and Fig. 7a, the largest electrical components are respectively $E_z$ and $E_y$, which both have symmetry about $x = 0$.

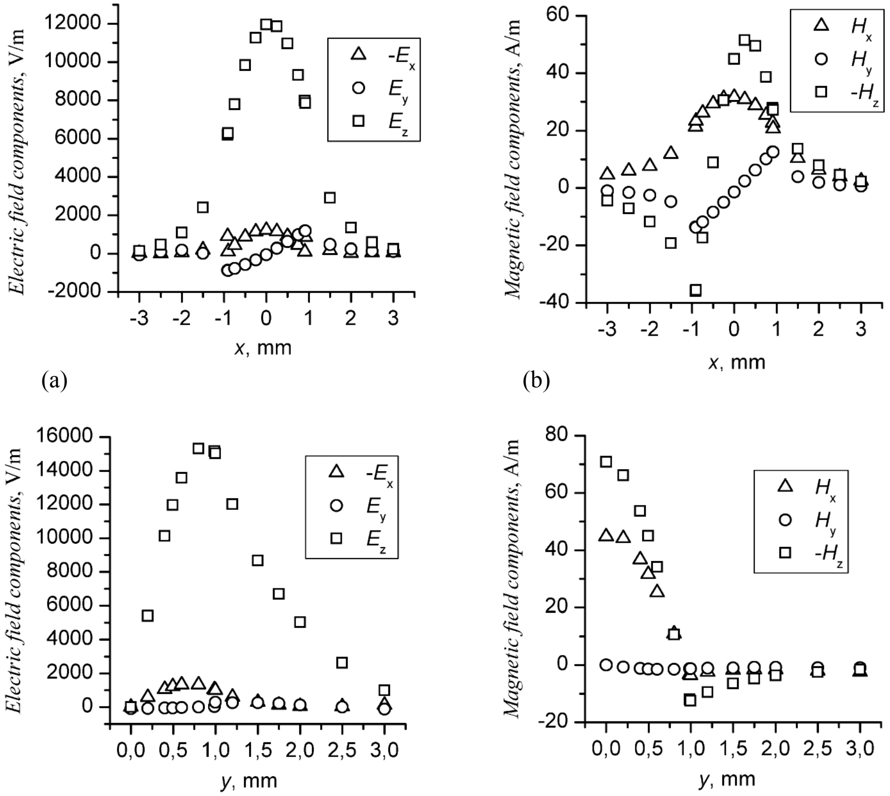
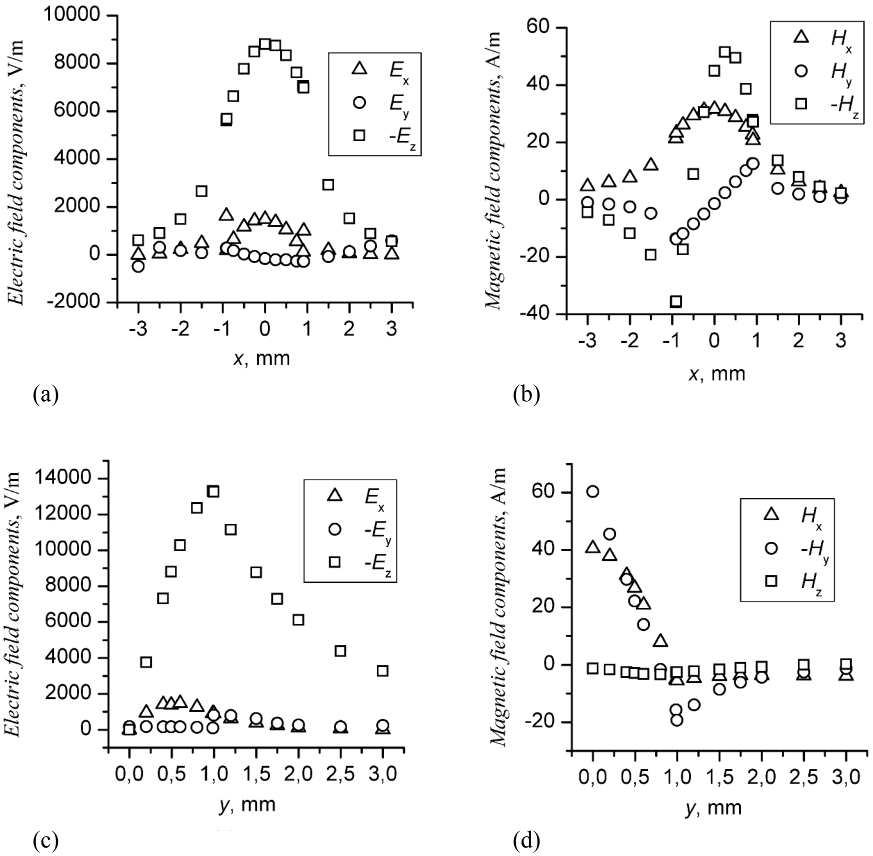


**Fig. 5.** Simulated distributions of the field components of the dominant mode at $z = 5.03$ mm: (a, b) $y = 0.495$ mm; (c, d) $x = 0$ mm

The three magnetic components shown in Fig. 6b,d and Fig. 7b,d are comparable to each other. This gives grounds to conclude that the dominant mode on the longitudinally magnetized FIG represents a modified $E^{y}_{11}$ mode, in which all three magnetic components $H_x$, $H_y$ and $H_z$ should be taken into account.

### 3.4. DISCUSSION

The transverse distributions of field components of the dominant mode corresponding to a maximum of the electric field magnitude are shown in Figs. 2, 5 and 6 for Case 1, Case 2 and Case 3, respectively. These figures show that the $E_z$ component is the largest among the electric field components and reveal that the maxima of the electric field magnitude coincide with the maxima of the $E_z$ component.



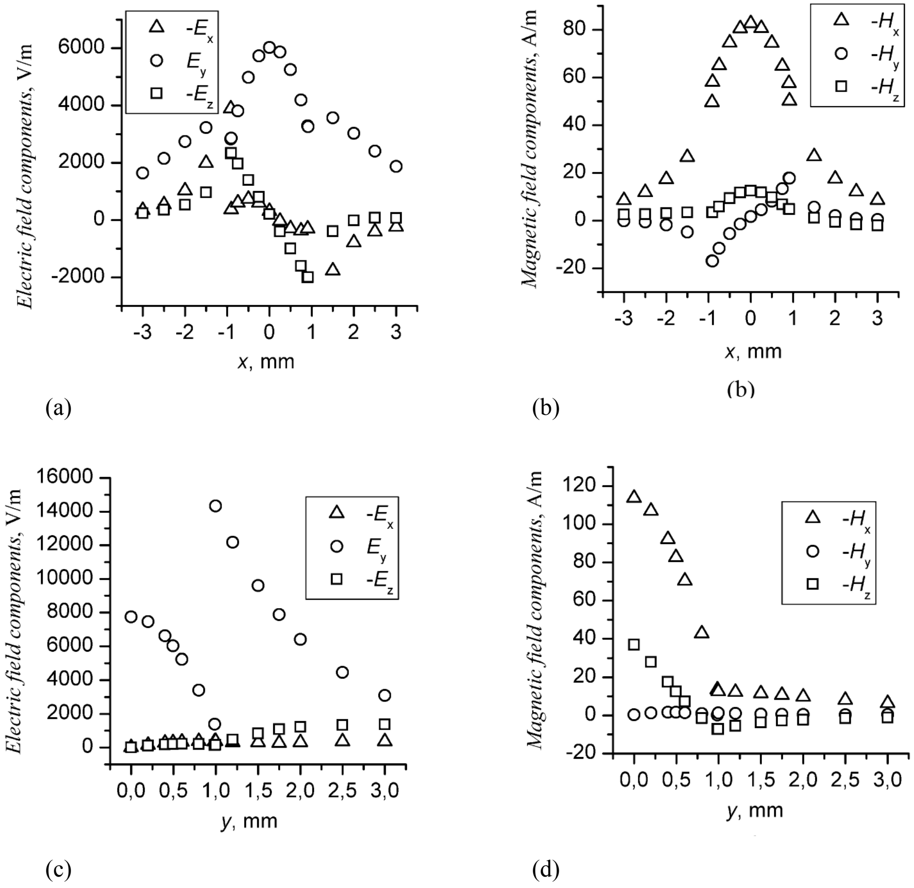
(a) (b)

(c) (d)

**Fig. 6.** Simulated distributions of the field components of the dominant mode at $z = 1.74$ mm: (a, b) $y = 0.495$ mm; (c, d) $x = 0$ mm

The obtained distributions of field components of the dominant mode corresponding to a minimum of the electric field magnitude are shown in Figs. 3, 4 and 7 for Case 1, Case 2 and Case 3, respectively. It is evident that the minima of the electric field magnitude coincide with the maxima of the $E_y$ component.

In Case 2 an electric field asymmetry has been obtained. This asymmetry can produce a nonreciprocal field distribution in coupled ferrite-dielectric image guide structures at transverse magnetization perpendicular to the ground plane. It gives reason to suppose that it is possible to expect a nonreciprocal behaviour in the coupled ferrite-dielectric image guide structures at this type of magnetization.



(a) (b)

(c) (d)

**Fig. 7.** Simulated distributions of the field components of the dominant mode at $z = 3.91$ mm: (a, b) $y = 0.495$ mm; (c, d) $x = 0$ mm

## 4. CONCLUSION

The transverse distributions of all six components of the dominant mode on the magnetized FIG have been obtained numerically. The results showed that only at magnetization perpendicular to the direction of propagation and parallel to the ground plane the dominant mode is pure $E^y_{11}$ with main components $E_y$, $E_z$ and $H_x$, according to the Markatili's classification An asymmetry about the middle of the ferrite core appears at transverse magnetization perpendicular to the ground plane, as well as a growth of the components $H_z$ и $E_x$. All three components of the magnetic field prove to be significant at longitudinal magnetization. The next step of the investigation could be the simulation of the coupled ferrite-dielectric IG structures.

# REFERENCES

[1] Mrozowski, M., J. Mazur. In: *32 Intern. Wiss. Koll.*, 1987, 181.

[2] Xia, Jiqing, Peter P. Toulius, Carmine Vittoria. *IEEE Trans. Microwave Theory Tech*., 1989, **37**, 1547.

[3] Kwan, P., C. Vittoria. *J. Appl. Phys.*, 1993, **73**, 6466.

[4] Akyol, Ahmet Soydan, Lionel Edward Davis. In: *IEEE International Microwave Symposium*, 2001, 1179.

[5] Akyol, Ahmet Soydan, Lionel Edward Davis. *IEEE Trans. Microwave Theory Tech*., 2003, **51**, 1476.

[6] Sacli, H., H. Benzina, T. Aguili, J. W. Tao. In: *Progress in Electromagnetics Research Symposium*, 2009, 1469.

[7] Arestova, I. I., S. A. Ivanov. In: *XII ICMF (International Conference on Microwave Ferrites*) *Proceedings*, 1994, 188.

[8] Arestova, Iliyana, Rositza Tomova, Gergina Angelova. *Annuaire de l'Universite de Sofia "St. Kliment Ohridski", Faculte de Physique*, 2010, **103**, 5.

[9] Arestova, Iliyana. *Comptes Rendus de l'Academie bulgare des Sciences*, 2017, in processing.

[10] Marcatili, E. A. J. *Bell System Technical Journal*, 1969, **48**, 2071.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ

Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS

Volume 111

# NUMERICAL INVESTIGATION OF COUPLED FERRITE-DIELECTRIC IMAGE GUIDE STRUCTURE FOR Ka-BAND

ILIYANA ARESTOVA

*Department of Radiophysics and Electronics*

*Илияна Арестова*. ЧИСЛЕНО ИЗСЛЕДВАНЕ НА СВЪРЗАНА ФЕРИТО-ДИЕЛЕКТРИЧНА ОГЛЕДАЛНА СТРУКТУРА ЗА Ka-ОБХВАТА

Свързани феритен и диелектричен вълноводи са изследвани числено по метода на крайните елементи (МКЕ) в честотния диапазон 26–40 GHz. Моделирана е свързана ферито-диелектрична огледална структура със същата геометрия, както в предходно експериментално изследване при нехомогенно намагнитване. Феритният елемент в настоящото числено изследване е намагнитен хомогенно перпендикулярно на огледалната равнина и на посоката на разпространение. Това хомогенно намагнитване представлява първо приближение на реалното нехомогенно намагнитване. Получената честотна зависимост на загубите в права и обратна посока на разпространение в структура с дължина на свързване *l,* равна на 17.6 mm, показва силно изразена невзаимност в широк честотен диапазон: от 36.2 до 38 GHz. С цел да се установи зависимостта на невзаимното поведение от дължината на свързване, *l* беше променяно от 12.5 до 19.5 mm. Резултатите показват, че при по-големи дължини на свързване се постига по-добра невзаимност, което означава по-голяма абсолютна стойност на изолацията и по-широка работна честотна лента.

*Iliyana Arestova.* NUMERICAL INVESTIGATION OF COUPLED FERRITE-DIELECTRIC IMAGE GUIDE STRUCTURE FOR Ka-BAND

Coupled ferrite and dielectric image guides have been investigated numerically by finite element method (FEM) in the frequency range 26–40 GHz. We have modelled the coupled ferrite-dielectric

---

*For contact*: Iliyana Arestova, Department of Radiophysics and Electronics, Faculty of Physics, Sofia University „St. Kliment Ohridski”, 5 James Bourchier Blvd., Sofia 1164, Phone: 359 2  8161 724, E-mail: ilar@phys.uni-sofia.bg

image guide (CFDIG) structure with the same geometry as in our recent experimental investigation of nonreciprocal inhomogeneously magnetized CFDIG structure. The ferrite element in this numerical investigation has been homogeneously magnetized perpendicularly to the ground plane and the direction of propagation. The ferrite element in this numerical investigation has been homogeneously magnetized perpendicularly to the ground plane and the direction of propagation. This homogeneous magnetization represents first approximation of the real inhomogeneous magnetization. The frequency dependences of losses in forward and backward direction of propagation for CFDIG structure with coupling length $l$ equal to 17.6 mm have shown strong nonreciprocity of the structure in a wide frequency band from 36.2 to 38 GHz.

In order to investigate the dependence of nonreciprocity on the coupling length, $l$ has been successively varied from 12.5 to 19.5 mm. Results have revealed that at longer coupling lengths better nonreciprocal behaviour can be achieved, which means greater absolute value of isolation and wider frequency band of operation.

***Keywords:*** nonreciprocal devices, image guide, finite element method, millimeter waves, coupled image guide structures.

***PACS number:*** 84.40.Az

## 1. INTRODUCTION

Nonreciprocal devices have been traditionally invented on the basis of magnetized ferrite elements. The ferrite materials unfortunately decrease their gyrotropy as the frequency is increased. That is the reason the conventional ferrite devices developed for centimetre-wave range cannot be used with the same success in millimetre-wave range. Alternative approaches are sought to invent nonreciprocal devices for millimetre-wave range that require only weak gyrotropy [1, 2]. One of these alternative approaches is the achievement of nonreciprocal behaviour of CFDIG structures [3–6].

Very good nonreciprocal parameters have been reported in [4, 5] for a CFDIG structure with inhomogeneously magnetized ferrite element (bar). The ferrite element has been magnetized by using a disk-shaped permanent magnet, whose diameter is comparable with the length of the ferrite bar. The experimental investigation of such type CFDIG structure has been completed recently with the help of electric probes [6]. The distributions of all three electric field components in forward and backward direction of propagation have been measured. These distributions have revealed a nonreciprocal coupling between dielectric and ferrite image guides. In forward direction of propagation the period of power transfer L equals to a half coupling length l/2, while in backward direction it is two times greater and equals to the coupling length l. Insertion losses equal to –2.5 dB and an isolation equal to –16 dB have been registered at a frequency of 34 GHz. Isolation better than –10 dB has been measured in the frequency range (33.6–34.5) GHz.

The measurement of the permanent inhomogeneous magnetic field [6] has shown that the transverse component, which is perpendicular to the ground plane, predominates over the longitudinal one in almost the whole coupled region. That is the reason to replace as first approximation in our numerical investigation the inhomogeneous magnetic field with a homogeneous one. Here we have modelled the homogeneously magnetized CFDIG structure with the same geometrical parameters as in the experimental investigation [6]. After that, we have investigated numerically by FEM the behaviour of this structure in Ka-band (26–40 GHz). Strong nonreciprocity at about 37 GHz has been registered.

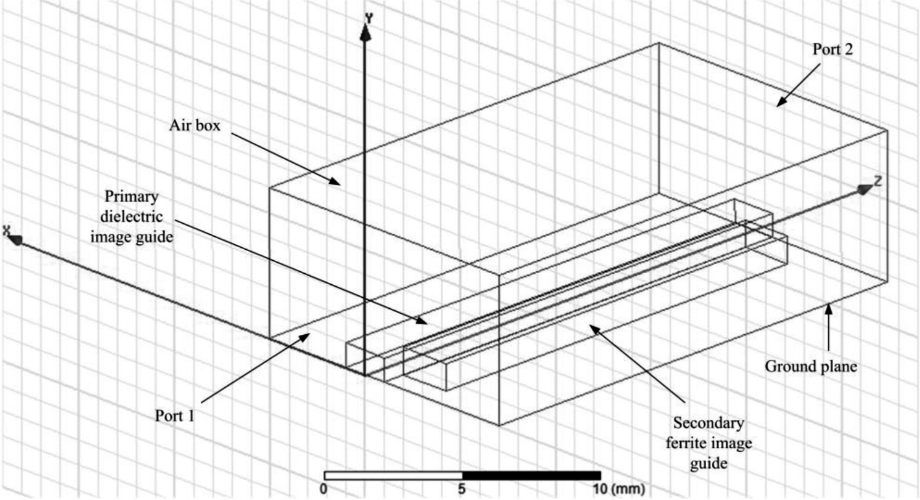


**Fig. 1.** The model of the coupled ferrite-dielectric image guide structure

## 2. MODEL

The geometry of the model of the CFDIG structure is shown in Fig. 1. The model volume coincides with the volume of the so-called air box (radiation box) with dimensions 12×6×20 mm$^3$. The primary dielectric image guide (IG) represents a parallelepiped (bar) with dimensions 2×0.97×20 mm$^3$. It is made by alumina with a relative permittivity $\varepsilon_r = 9.6$ and a dielectric loss tangent $tg\delta_\varepsilon = 10^{-4}$. The dimensions of the primary IG ensure single mode operation in Ka-band with the mode $E^y_{11}$ according to the Marcatili's mode classification [7, 8]. The secondary ferrite IG represents a bar with dimensions 2.2×1.1×17.6 mm$^3$ and is made by nickel ferrite (1СЧ4, Russia). The ferrite has a relative permittivity $\varepsilon_r = 11.1$, a dielectric loss tangent $tg\delta_\varepsilon = 10^{-2}$ and a saturation magnetization $4\pi M_s = 4.63$ kG. The coupling length $l$ coincides with length of the ferrite bar and is equal to 17.6 mm. The ferrite bar has been modelled as homogeneously magnetized along O$y$ axis with a field strength $H = 80$ kA/m.

The upper wall of the air box in the $xz$ plane and both side walls in the $yz$ planes have been defined as radiating surfaces. They have been chosen to be far enough from the corresponding walls of the primary and secondary IGs. The bottom wall of the air box represents an image plane and it has been defined as a perfect conductor. Both side walls of the air box in the *xy* planes have been defined as ports.

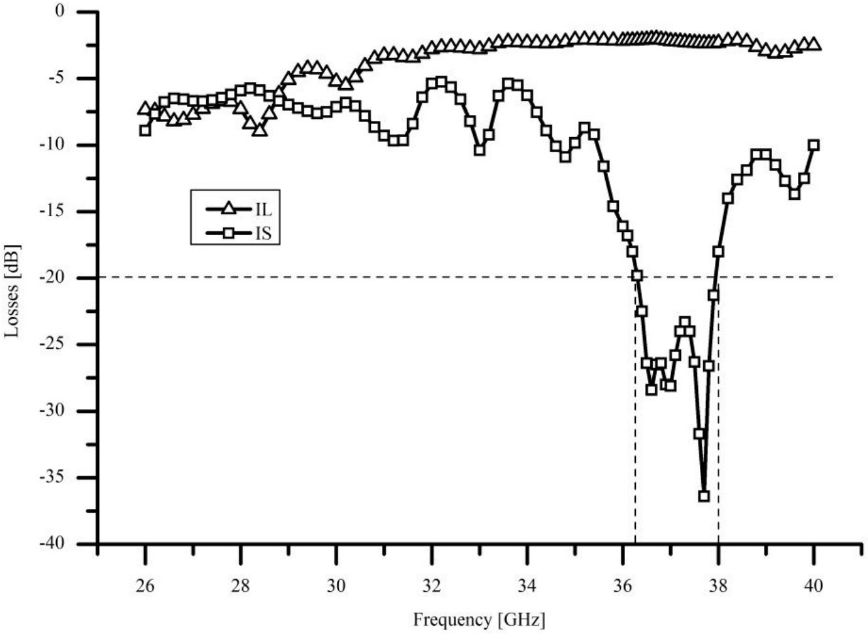


**Fig. 2.** Frequency dependences of losses for CFDIG structure with $l$ = 17.6 mm

## 3. RESULTS

The frequency dependences of losses in forward and backward direction of propagation for CFDIG structure with coupling length $l$ equal to 17.6 mm are shown in Fig. 2. The losses in forward direction known as insertion losses *IL* coincide with the scattering matrix element $S_{21}$, and those in backward direction named isolation *IS* – with the scattering matrix element $S_{12}$. The curves in Fig. 2 have shown well-expressed nonreciprocal behaviour of the CFDIG structure. The insertion losses vary from –9 dB at 28.4 GHz to –2 dB at 36.7 GHz. They are approximately constant in a wide frequency range from 33 to 38.6 GHz. The isolation has its absolute minimum at 37.7 GHz and is equal to –36.7 dB. Isolation greater than –20 dB has been registered in frequency range from 36.2 to 38 GHz. The nonreciprocal behaviour with insertion losses about –2 dB and isolation better than –20 dB in bandwidth *BW* equal to 1.8 GHz represents very good wideband performance for isolators at millimetre waves.

**Table 1**

| $l$ [mm] | $f_{\mathrm{min}}$ [GHz] | $IL$ [dB] | $IS$ [dB] |
|---|---|---|---|
| 12.5 | 33.1 | –3.39 | –12.7 |
| 13 | 33.1 | –2.53 | –9.52 |
| 13.5 | 34.5 | –3.36 | –11.9 |
| 14 | 33.6 | –2.26 | –11.9 |
| 14.5 | 34.9 | –2.05 | –13.7 |
| 15 | 34.8 | –2.17 | –19.9 |
| 15.5 | 36 | –2.12 | –19.4 |
| 16 | 35.9 | –2.09 | –39 |
| 16.5 | 36.8 | –2.22 | –21.4 |
| 17 | 36.6 | –2.1 | –38.7 |
| 17.6 | 37.7 | –2.35 | –36.4 |
| 18 | 37.5 | –2.24 | –22.9 |
| 18.5 | 38 | –2.09 | –47.2 |
| 19 | 39.1 | –2.35 | –28.7 |
| 19.5 | 38.5 | –2.29 | –31.5 |

In order to investigate the dependence of nonreciprocity on the coupling length we have successively modelled CFDIG structures with coupling length from 12.5 mm to 19.5 mm. Results from these numerical investigations are presented in Tabl. 1. The frequency $f_{\mathrm{min}}$ represents frequency; at which isolation $IS$ has its absolute minimum. The values of insertion losses $IL$ and isolation $IS$ shown in Tabl. 1 corresponds to frequency $f_{\mathrm{min}}$. It is evident from Tabl. 1 that in general $f_{\mathrm{min}}$ increases and $IS$ decreases with increasing the coupling length $l$. The insertion losses $IL$ keep almost constant value in the range (–2) – (–3) dB. These results imply that the longer CFDIG structures possess better nonreciprocity. Additional investigation has also shown that the longer CFDIG structures have wider bandwidth $BW$, defined as frequency band in which the absolute value of isolation is greater than 20 dB.

We have applied some efforts to find a homogeneously magnetized CFDIG structure, which has parameters close in value to those in the experimental investigation [6]. The numerical investigation has shown that the CFDIG structure with coupling length $l$ equal to 14 mm possesses the closest parameters. The frequency dependences of losses in forward and backward direction of propagation for this CFDIG structure are shown in Fig. 3. It is evident that it has worse nonreciprocal behaviour in comparison to structure with coupling length 17.6 mm (Fig. 2). The absolute value of isolation does not reach the standard value of 20 dB. Frequency band in which the absolute value of isolation is greater than 10 dB is from 33.3 to 34 GHz.

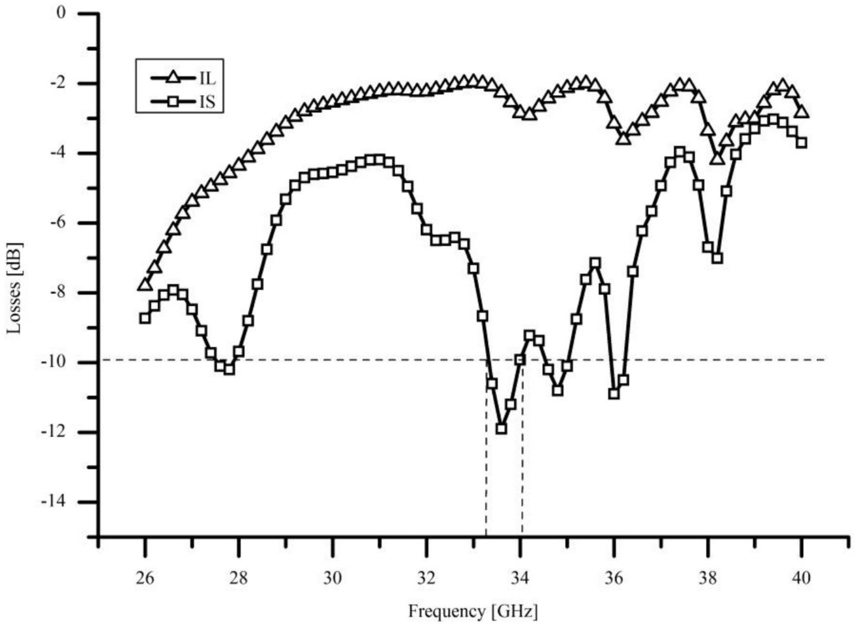


**Fig. 3.** Frequency dependences of losses for CFDIG structure with $l = 14$ mm

## 4. DISCUSSION

Geometrically identical CFDIG structure with coupling length 17.6 mm has been investigated experimentally in [6]. Insertion losses equal to –2.5 dB and an isolation equal to –16 dB have been measured at a frequency of 34 GHz. Isolation better than –10 dB has been registered in the frequency range (33.6–34.5) GHz. The corresponding insertion losses were about –3 dB. The magnetization of this structure was inhomogeneous due to the used disk-shaped permanent magnet, whose diameter is comparable with the length of the ferrite bar. The model for numerical investigation of the CFDIG structure does not permit the definition of inhomogeneously magnetized ferrite element, but only of homogeneous one. The comparison of the measured and calculated parameters has shown that the inhomogeneously magnetized CFDIG structure has worse nonreciprocity than the homogeneous one with the same coupling length. At the same time, the calculated frequency $f_{\min}$ is greater than the measured frequency of minimal isolation. As the $f_{\min}$ increases with the coupling length we can conclude that inhomogeneous magnetization makes the effective coupling length shorter. The explicit role of the existing longitudinal components of the inhomogeneous permanent magnetic field could not be understood in the frames of the model with homogeneous magnetization.

## 5. CONCLUSIONS

Coupled ferrite-dielectric image guide structure has been investigated numerically by finite element method. Frequency dependences of forward and backward losses in Ka-band have been derived for coupled ferrite-dielectric structures with different coupling lengths. Strong nonreciprocal behaviour of the coupled ferrite-dielectric structures in a wide frequency range has been registered. We can conclude that nonreciprocity depends on the coupling length and improves as it increases. The numerical results for the coupled ferrite-dielectric structure with coupling length equal to 17.6 mm have been compared with data obtained experimentally earlier. The comparison has shown that the inhomogeneously magnetized structure with coupling length 17.6 mm has lower operating frequency, which corresponds to homogeneously magnetized ferrite-dielectric structure with coupling length equal to 14 mm. The future work will aim to properly model the coupled ferrite-dielectric structure with inhomogeneous magnetization that could reveal the role of the axial components of the permanent magnetic field.

### REFERENCES

[1] Akyol, Ahmet Soydan, Lionel Edward Davis. *IEEE Trans. Microwave Theory Tech.*, 2003, **51**, 1476.

[2] Akyol, Ahmet Soydan, Lionel Edward Davis. In: *IEEE International Microwave Symposium*, 2001, 1179.

[3] Kwan, P., C. Vittoria. *IEEE Trans. Microwave Theory Tech.*, 1993, **41**, 652.

[4] Belyakov, S. V., S. S. Gigoyan, B. A. Murmuzhev. *Elektonnaya tehnika*, 1989, **419**, 42 (In Russian).

[5] Arestova, I. I., S. A. Ivanov. In: *XII ICMF (International Conference on Microwave Ferrites) Proceedings*, 1994, 188.

[6] Arestova, Iliyana, Rositza Tomova, *Radioengineering*, 2018, in processing.

[7] Marcatili, E. A. J. *Bell System Technical Journal*, 1969, **48**, 2071.

Arestova, Iliyana, Rositza Tomova, Gergina Angelova. *Annuaire de l'Universite de Sofia "St. Kliment Ohridski", Faculte de Physique*, 2010, **103**, 5.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# ФОТОМЕТРИЧНО ИЗСЛЕДВАНЕ НА НОВИ В ГАЛАКТИКАТА М31

АНТОНИЯ ВЪЛЧЕВА, НИКОЛАЙ БОЙЧЕВ

*Катедра „Астрономия“*

*Антония Вълчева, Николай Бойчев.* ФОТОМЕТРИЧНО ИЗСЛЕДВАНЕ НА НОВИ В ГАЛАКТИКАТА М31

Изследвано е фотометричното поведение на две нови, едната от които е класическа, M31N 2016-07c, а другата – повторна, M31N 2016-07e, в галактиката М31. Новите са наблюдавани през юли 2016 г., като данните покриват времеви интервал от почти три месеца. Построени са техните криви на блясъка в *BVR* филтри. Определени са типът на новата според кривата на блясъка и параметърът $t_2$ – времето за падане на блясъка с 2 зв. величини. Дискутирана е зависимостта „максимален блясък–скорост на падане на блясъка“ (maximum magnitude – rate of decline, MMRD).

*Antoniya Valcheva, Nikolay Boytchev.* PHOTOMETRIC STUDY OF NOVAE IN M31 GALAXY

otometric behavior of two novae, a classic one, M31N 2016-07c, and a recurrent one, M31N 2016-07e, in M31 galaxy is studied. The novae were registered in July 2016, and our observational data are covering a time interval of almost three months. Novae' light curves were constructed in BVR filters. The type of the novae according to their light curve and the parameter $t_2$, the time for the brightness to drop by 2 magnitudes, were determined. The relationship "maximum magnitude–rate of decline" (MMRD) is also discussed.

***Keywords:*** novae, light curve, M31, MMRD relationship
***PACS numbers:*** 97.30.Qt, 97.80.Gm, 98.52.Nr

---

*За контакти*: Антония Вълчева, Катедра „Астрономия“, Физически факултет, Софийски университет „Св. Климент Охридски“, бул. „Джеймс Баучър“ 5, 1164 София, България, Тел.: +359 2 81 61 717, E-mail: valcheva@phys.uni-sofia.bg

# 1. УВОД

Новите са тип катаклизмични променливи, които възникват в тесни двойни системи при запалване на термоядрени реакции на повърхността на бяло джудже [1 Bode, 2011]. Вторият компонент в системата е звезда от Главната последователност, субгигант или червен гигант.

Кривите на блясъка на новите дават информация за изменението на блясъка им с времето и се характеризират с покачване, максимум и спадане на блясъка. Вида на кривите на блясъка зависят от параметрите на системата, затова тяхното изследване е важен инструмент за разбиране физиката на избухванията и определяне характеристиките на тесните двойни системи, в които се пораждат новите. В зависимост от вида на кривата на блясъка новите се разделят на няколко класа според Strope et al., 2010 [18] – S (smooth), P (plateau), D (dust dips), C (cusps), O (oscillations), F (flat-topped) и J (jitters). В зависмост от скоростта на спадане на блясъка новите са разделени на *много бързи*, *бързи, нормално бързи, бавни* и *много бавни* (Payne-Gaposchkin, 1957 [14]). Този параметър е свързан с блясъка (зв. величина) на новата в максимум и от много години усилено се работи по уточняване на тази връзка. Зависимостта максимален блясък–скорост на падане на блясъка (MMRD) дава възможност директно от кривата на блясъка да се оцени на какво разстояние се намира обектът и по този начин новите могат да се използват за индикатори на разстояние във Вселената. В линеен вид зависимостта най-често се дава с формулата $M = b_n \log t_n(d) + a_n$, където $M$ е абсолютната зв. величина, $n$ е със стойност 2 или 3, а $t_n$ е времето за намаляване на блясъка с $n$ зв. величини.

Съществен напредък в калибрирането на тази зависимост се наблюдава в последните години, защото почти не остават незабелязани нови в галактиката M31. За много от тях е възможно да се построят подробни криви на блясъка, да се определят $M_{\max}$ и $t_n$ и в резултат да се намери точният вид на MMRD (Darnley et al., 2006 [5], Schafter et al., 2011[16]).

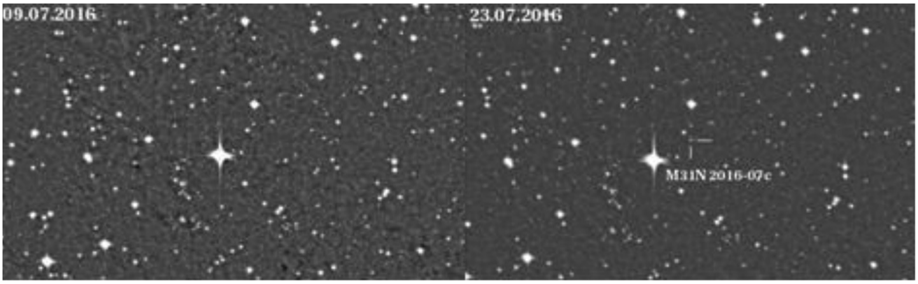
# 2. КЛАСИЧЕСКАТА НОВА M31N 2016-07C

Класическата нова M31N 2016-07c (RA(J2000)=00h43m57s.68 и DEC(J2000)=+41d34'52".8) е открита от K. Nishiyama & F. Kobachima на 21 юли 2016 г. от обсерваторията Observer Yasuo Sano, Nayoro, Япония (CBAT TOCP PNV J00435768+4134528). Фотометрични данни в *BR* филтри и уточнени координати са публикувани от Valcheva et al. [19] (ATel #9264). Паралелни спектрални наблюдения на обекта потвърждават типа и местоположението му – нова от спектрален тип FeII в галактиката M31. Спектърът се характеризира с балмерови линии (Hα и Hβ) с ширини FWHM 1300±100 km/s [3] (ATel #9296).

10 дни след откриването ѝ, Chinetti et al. [4] (ATel #9347) регистрират повторно повишаване на блясъка на новата.

### 2.1. НАБЛЮДАТЕЛНИ ДАННИ

Наблюдения на M31N 2016-07c от България са проведени в периода 22.07.2016–01.10.2016 г. в *BR* стандартни филтри с 50/70 cm Schmidt телескоп на НАО Рожен (вж. табл. 1). Фотометричните условия през наблюдателния период са различни, като пълната ширина на половината височина на звездния профил (FWHM) варира в интервала от 2.5" до 4". Всички наблюдателни кадри са обработени и е направена апертурна фотометрия на избрани обекти в полето. За целта е използван специализираният астрономически пакет за обработка и анализ на астрономически кадри IRAF. Преминаването към стандартни зв. величини е направено чрез каталога на звездни обекти в M31 на Massey et al. (2016) [13]. На фиг. 1 е показана карта за идентификация на новата в *R*-филтър.



**Фиг. 1.** Карта за идентификация на новата M31N 2016-07c: вляво – преди откриването; вдясно – след откриването. Кадрите са получени с 50/70 cm Schmidt телескоп на НАО Рожен в *R*-филтър и са с площ от 14' x 8'. Север е нагоре, изток е наляво

**Таблица 1.** Наблюдателни данни за M31N 2016-07c, получени с 50/70 cm Schmidt телескоп на НАО Рожен в периода 07.22.2016–01.10.2016 г.

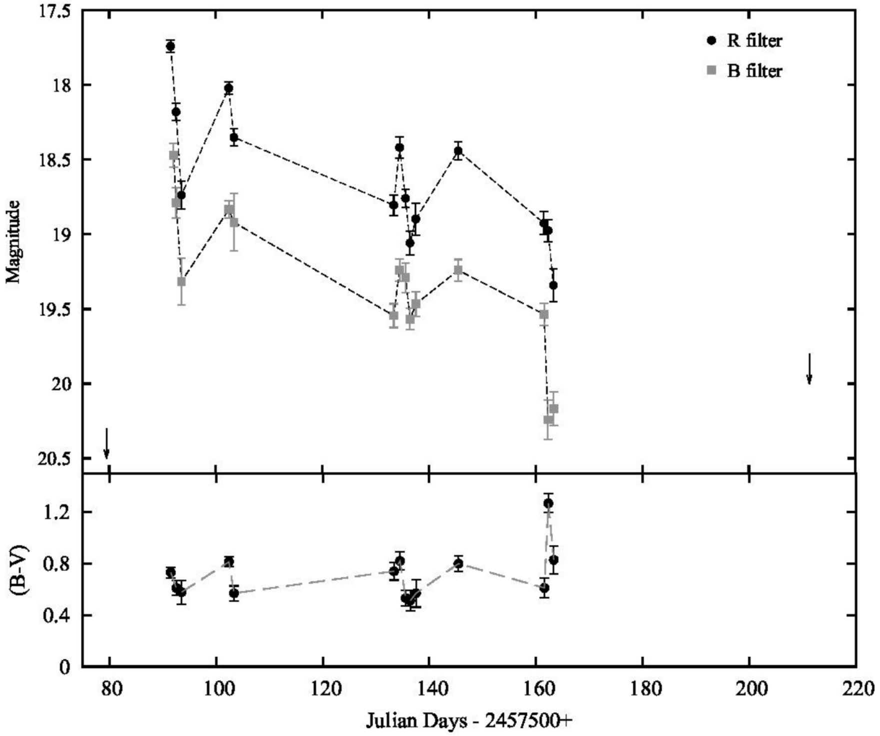
| Дата и час | *B*[mag] | *sB* [mag] | *R*[mag] | *sR*[mag] | Интеграция [s] |
|---|---|---|---|---|---|
| 2016-07-22T00:57:08 | 18.47 | 0.081 | | | 4×180 |
| 2016-07-22T00:33:55 | | | 17.74 | 0.043 | 7×180 |
| 2016-07-23T01:14:55 | 18.79 | 0.102 | | | 3×180 |
| 2016-07-23T01:03:46 | | | 18.18 | 0.059 | 5×180 |
| 2016-07-24T01:48:01 | 19.315 | 0.156 | | | 3×180 |
| 2016-07-24T01:29:51 | | | 18.737 | 0.093 | 5×180 |
| 2016-08-01T22:21:43 | 18.833 | 0.058 | | | 5×300 |
| 2016-08-01T21:53:29 | | | 18.02 | 0.039 | 5×300 |
| 2016-08-02T21:46:39 | 18.919 | 0.192 | | | 3×300 |
| 2016-08-02T21:30:26 | | | 18.350 | 0.058 | 3×300 |

| | | | | | |
|---|---|---|---|---|---|
| 2016-09-01T21:09:30 | 19.544 | 0.080 | | | 3×300 |
| 2016-09-01T20:46:30 | | | 18.804 | 0.068 | 5×300 |
| 2016-09-02T23:32:35 | 19.238 | 0.074 | | | 3×300 |
| 2016-09-02T23:12:07 | | | 18.418 | 0.071 | 3×300 |
| 2016-09-04T02:28:20 | 19.290 | 0.098 | | | 3×300 |
| 2016-09-04T01:46:03 | | | 18.758 | 0.06 | 5×300 |
| 2016-09-04T22:35:02 | 19.568 | 0.071 | | | 3×300 |
| 2016-09-04T22:12:08 | | | 19.057 | 0.078 | 5×300 |
| 2016-09-06T00:38:48 | 19.466 | 0.082 | | | 3×300 |
| 2016-09-06T00:00:16 | | | 18.897 | 0.107 | 5×300 |
| 2016-09-13T23:06:22 | 19.240 | 0.072 | | | 5×300 |
| 2016-09-13T23:23:37 | | | 18.441 | 0.059 | 5×300 |
| 2016-09-30T01:59:16 | 19,535 | 0,076 | | | 3×300 |
| 2016-09-30T01:39:20 | | | 18.925 | 0.076 | 3×300 |
| 2016-09-30T20:04:38 | 20.242 | 0.133 | | | 3×300 |
| 2016-09-30T19:42:27 | | | 18.974 | 0.074 | 3×300 |
| 2016-10-01T20:30:05 | 20.169 | 0.113 | | | 3×300 |
| 2016-10-01T20:13:08 | | | 19.341 | 0.109 | 3×300 |

## 2.2. КРИВА НА БЛЯСЪКА

Получената *BR* фотометрия за M31N 2017-07c в периода 22.07.2016 г.–01.10.2016 г. е използвана  за построяване на кривата на блясъка, показана на фиг. 2, горе. Тя се характеризира с плавно спадане на блясъка, върху което се налюдават локални минимуми и максимуми (осцилации) и в двата филтъра. Според вида на кривата на блясъка, новата спада към клас J (jitter) (Strope et al., 2010 [18]). Наблюденията преди избухването на 09.07.2016 г. и след затихването на 21.11.2016 г. на M31N 2017-07c дават възможност да се определи гранична зв. величина в *R*, над която новата не се наблюдава (стрелките на фиг. 2).

Данните от два филтъра ни дават възможност да пресметнем и проследим изменението на цвета (*B*–*R*) на новата в наблюдателния период (фиг. 2, долу). Средният цвят е ~0.6 mag, който е типичен за нова, слабо повлияна от екстинкция.



**Фиг. 2.** Горе: крива на блясъка на M31N 2016-07c в *BR* филтри. Наблюденията са получени с 50/70 cm Schmidt телескоп на НАО Рожен и са описани в табл. 1. Стрелките показват граничната зв. величина, до която новата не се наблюдава. Долу: времева еволюция на цвета (*B*–*R*)

### 2.3. РАЗШИРЕНА КРИВА НА БЛЯСЪКА

За да се направи по-детайлен анализ на фотометричното поведение на дадена нова, е важно да се привлекат всички данни от литературата. На фиг. 3 е показана разширена крива на блясъка на M31N 2017-07c в *R*-филтър, за построяването на която са използвани всички данни, публикувани от други автори. Наблюдателните данни, получени в това изследване от 50/70 cm Schmidt телескоп на НАО Рожен, са показани с плътно кръгче. Разполагаме с данни в *R*-филтър от 1.6 m Pirka телескоп на обсерваторията в Nayoro, Япония (CBAT TOCP PNV J00435768+4134528), показани на фиг. 3 с празно кръгче. Chinetti et al. [4] (ATel #9347) публикуват поредица от зв. величини, получени с 48-инчов телескоп на Palomar в *r*- (SDSS) филтър. Макар и много близки, *R*- и *r*- филтри имат различни характеристики (ефективна дължина на вълната и ширина на филтъра). За да съгласуваме тези данни с нашите, е необходимо да прехвърлим *r* зв. величини в стандартната система на Cousins. За прехвърлянето е използвано следното уравнение при предположение за плосък спектър в разглеждания спектрален диапазон:

$$m - m' = -2.5\log\left(\frac{f'\lambda_0}{f\lambda_0}\right),$$

където $m$ и $m'$ са звездните величини в двете фотометрични системи, $f_{\lambda 0}$ и $f'_{\lambda 0}$ са специфичните потоци при $m = 0$ в системата Vega.

Замествайки с характерните стойностите за $R$ (Cousins) и $r$ (SLOAN) в последното уравнение, съответно $f_{\lambda 0}=2.246\times10^{-9}$ [erg/cm$^2$/s/A] и $f'_{\lambda 0}=2.529\times10^{-9}$ [erg/cm$^2$/s/A], получаваме

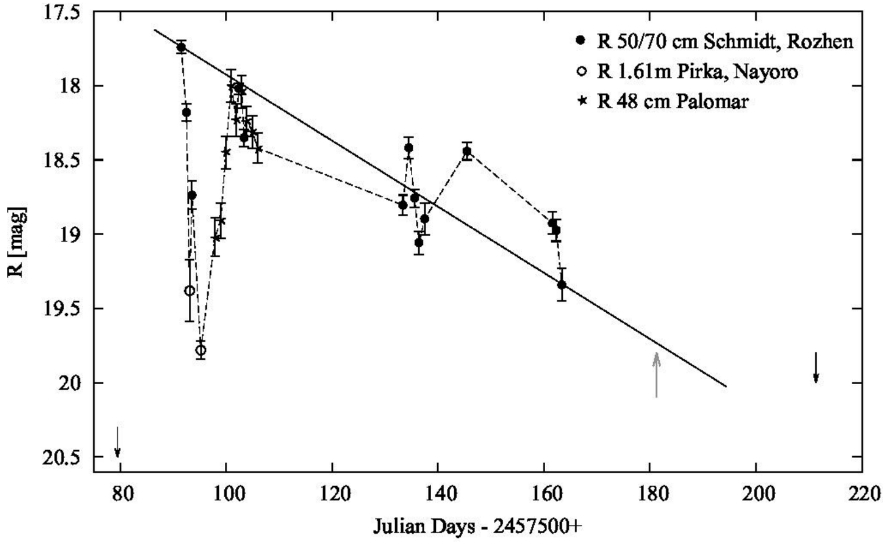
$$R - r(SLOAN) = -2.5lg\left(\frac{2.529x10^{-9}}{2.246x10^{-9}}\right),$$

$$R = r\ (SDSS) - 0.13$$

Чрез последното равенство можем да преминем от $r$ зв. величини на Chinetti et al. [4] (ATel #9347) към $R$ зв. величини в системата на Cousins, показани на фиг. 3 със звездички.

Когато кривата на блясъка описва добре намаляването на блясъка, тя дава възможност да се определят два важни параметъра на новата – максимален блясък и време за падане на блясъка с 2 или 3 зв. величини – $t_2$ ($t_3$). Познаването на тези величини за голям брой нови дава възможност да се калибрира зависимостта максимален блясък–скорост на падане на блясъка, която позволява новите да се използват като индикатори на разстояние. Тези два параметъра за M31N 2016-07c можем да получим от кривата на блясъка на фиг. 3. Тъй като нямаме данни за времето, когато новата е повишавала блясъка си, не можем да сме сигурни, че наблюдаваният максимум е истинският и не можем да пресметнем точното $t_2$. Можем обаче да определим времето, за което зв. величина ще намалее с 2 зв. величини (считано от най-ярката точка върху кривата). Това ще ни даде горна граница за $t_{2.}$ Както се вижда от фиг. 3, $t_2$ попада в интервал, извън нашите данни, и за да го определим, ще екстраполираме до $\Delta R = 2$ зв. величини. Заради вида на кривата – дълбок минимум и флуктуации, най-подходящо за определянето на $t_2$ е да направим линейна екстраполация между първата (максимумът) и последната наблюдателна точка.

От построената права получаваме $t_2 = 92.1$ дни (пунктирана стрелка на фиг. 3). Според класификацията на Payne-Gaposchkin (1957)[14] новата M31N 2016-07c се причислява към клас *бавна* и има скорост на спадане блясъка $v(t_2) = 2/t_2 = 0.02$ зв. величини/ден. Последната наблюдателна точка в кривата ни дава долна граница за $t_2$, т.е $t_2 > 69$ дни, а за неопределеност на резултата ще приемем $\sigma(t_2) = 92 - 69 = 23$ дни.



**Фиг. 3.** Крива на блясъка на M31N 2016-07c в *R*-филтър. Стрелките дават гранична зв. величина, над която новата не се наблюдава. Правата, свързваща максимума и последната наблюдателна точка, дава възможност да бъде определено времето за спадане на блясъка с две зв. величини, $t_2$ (вж. подраздел 2.3), отбелязано със сива стрелка

## 4. ПОВТОРНАТА НОВА M31N 2016-07E

### 4.1. НАБЛЮДАТЕЛНИ ДАННИ

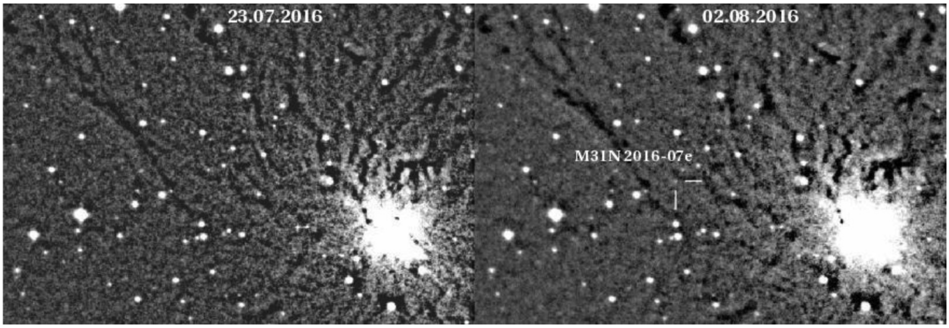
Повторната нова M31N 2016-07e (RA(J2000)=00h42m04.05s DEC(J2000)= +41°17'08.3") е открита през октомври 1990 г. от Bryan (1990) [2] – M31N 1990-10a. Второто нейно избухване е регистрирано през юли 2007 г. от Hatzidimitriou et al. (2007) [11] с 1.3m RC-телескоп в обсерваторията Skinakas, Гърция [11] (ATel #1131 – M31N 2007-07a). Последното избухване на M31N 2016-07e е докладвано на 27 юли 2016 г. от K. Nishiyama и F. Kabashima на CBAT Transient Objects Confirmation Page. Предполага се, че периодът на избухване на M31N 2016-07e е между 8 и 9 години, като обаче избухване през 1998/1999 г. не е регистрирано от никого. В M31 има само още 4 повторни нови, чийто период на избухване е по-малък от 10 години (Shafter et al. 2015 [17]).

Спектърът на новата (първи спектър до този момент), получен на 30 юли 2016 с 4m телескоп William Herschel в обсерваторията Roque de los Muchachos, Испания, е доминиран от балмерови линии (FWHM ~ 2000 km/s), което потвърждава природата на M31N 2016-07e. FeII емисионни линии, OI 7773, 8446, и CaII триплет също се наблюдават [6] (ATel #9281).

През 2016 г. M31N 2016-07e е наблюдавана доста интензивно заради връзката си с M31N 2007-07a и M31N 1990-10a. Оптични наблюдения в *BVR* филтри са проведени и публикувани от няколко международни екипа – Fabrika et al. [7] (ATel #9383), Hornoch et al. [12] (Atel #9386), Goranskij et al. [9] (ATel #9435). Повторната нова е наблюдавана и с 50/70 cm Schmidt телескоп на НАО Рожен. Получените данни в *BR* филтри са дадени в таблица 2, а карта за идентификация е показана на фиг. 4.

**Таблица 2.** Наблюдателни данни, получени с 50/70 Schmidt телескоп на НАО Рожен за M31N 2016-07e

| Дата и час | *B*[mag] | *sB* [mag] | *R*[mag] | *sR*[mag] | Интеграция [s] |
|---|---|---|---|---|---|
| 2016-08-01T22:21:43 | 19.212 | 0.095 | | | 5×300 |
| 2016-08-01T21:53:29 | | | 18.434 | 0.100 | 5×300 |
| 2016-08-02T21:46:39 | 19.095 | 0.210 | | | 3×300 |
| 2016-08-02T21:30:26 | | | 18.396 | 0.113 | 5×300 |



**Фиг. 4.** Карта за идентификация на новата M31N 2016-07e: вляво – преди откриването; вдясно – след откриването. Картите са получени с 50/70 cm Schmidt телескоп на НАО Рожен в *R*-филтър и са с площ от 14' × 8'. Север е нагоре, изток – наляво

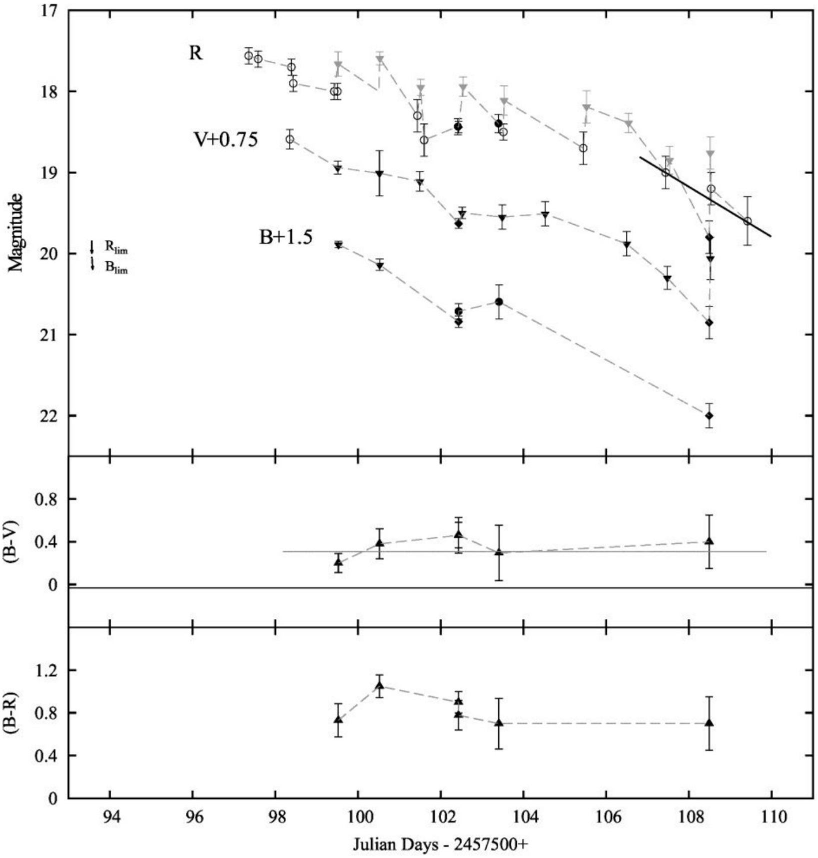
## 4.2. КРИВА НА БЛЯСЪКА

Подробна крива на блясъка на M31N 2016-07e може да бъде построена, като се добавят и всички литературни данни. Fabrika et al. [7] (ATel #9383) публикуват *BVR* фотометрия от 6 m BTA телескоп в САО, Русия (фиг. 5, плътни ромбове). Hornoch et al. [12] (ATel #9386) публикуват *VR* данни от 65 cm телескоп в Ondrejov, Чехия (фиг. 5, празни кръгчета), а Goranskij et al. [9] (ATel #9435) публикуват *BVR* данни от 50 cm Maksutov meniscus телескоп в Кримската астрономическа станция на Московския университет (фиг. 5, обърнати триъгълници). За по-добра визуализация на фиг. 5 към стойностите на *BV* зв. величини е добавена константа. Вижда се, че трите криви следват един ход на падане на блясъка, като различните флуктуации се отразяват едновременно в трите филтъра, въпреки че данните в $BV$ са по-малък брой.

На фиг. 5 е показана също и времевата еволюция на цветовете ($B–V$) и ($B–R$). Според Wright & Barlow (1975) [20] и Hachisu & Kato (2006) [10] няколко дни след оптичния максимум на новите истинският им цвят става $(B–V)_0 = –0.03$ mag (показан с права линия на средния панел). Ако определим един среден цвят от данните, можем да направим оценка за екстинкцията по лъча на зрение към новата. От данните получаваме $<B–V> = 0.37$ mag, следователно $E(B–V) = <B–V> – (B–V)_0 = 0.4$ mag и $A_V = 1.2$ mag ($A_R$~1 mag). Средният цвят ($B–R$) на новата е ~0.8 mag, което при сравнение с цвета на M31N 2016-07c (фиг. 3) също сочи към по-голяма екстинкция по лъча на зрение към новата.

Ако обърнем внимание само на кривата в $R$-филтър (фиг. 5, най-горе), можем веднага да забележим, че тя се характеризира с доста флуктуации на фона на основното спадане на блясъка. Ако тези флуктуации са действителни, кривата на блясъка на M31N 2016-07e следва да се класифицира като J (jitter) според класификацията на Strope et al. 2010 [18]. Прави впечатление обаче, че данните, показани с обърнати триъгълници, са системно по-ярки от другите такива, получени в същия или в много близък момент, като разликата винаги е ~0.4 mag. Ако тези данни не бъдат взети под внимание, кривата на блясъка може да се причисли по-скоро към клас S (smooth).

В случая на M31N 2016-07e също нямаме данни за точния момент на максимален блясък, но можем да пресметнем времето, за което зв. величина ще намалее с 2 зв. величини (считано от най-ярката точка върху кривата), което отново ще ни даде горна граница за $t_2$. От кривата се вижда, че времето за спадане на блясъка с 2 зв. величини попада във времевия интервал, за който имаме наблюдателни данни. В този случай традиционният подход е да се направи линейна интерполация между точките, намиращи се около $\Delta R = 2$ mag. Интерполиращата правата е показна на фиг. 5 в края на кривата. За $t_2$ получаваме 109.3–97.3=12±1 дни. Според класификацията на Payne-Gaposchkin (1957) [14] при това избухване M31N 2016-07e попада в клас *бързи.* За скоростта на падане на блясъка получаваме $v(t_2) = 2/t_2 = 0.17$ зв. величини/ден.



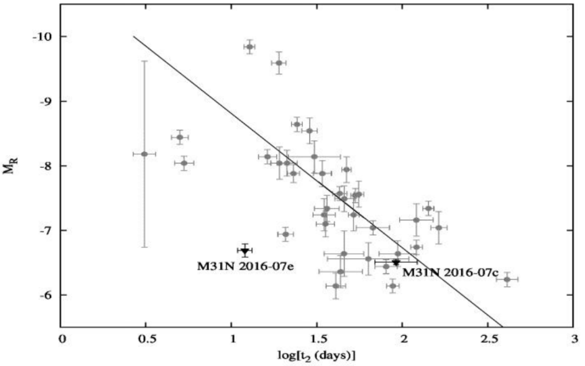
**Фиг. 5**. Горе: крива на блясъка на M31N 2016-07e в *BVR* филтри. С плътни кръгчета са показани данните от 50/70 cm Schmidt телескоп на НАО Рожен. Данните от Fabrika et al. [7] (ATel #9383) са показани с плътни ромбове; данните от Hornoch et al. [12] (ATel #9386) – с празни кръгчета, а данните от Goranskij et al. [9] (ATel #9435) с обърнати триъгълници. За по-добра визуализация, данните в *B* и *V* са отместени с констан-та. Стрелките показват гранична зв. величина в *B* и *R*. Правата линия е линейна интерпола-ция между точките (вж. подразд. 4.2). Средата: времева еволюция на цвета (*B–V*). Непрекъснатата права линия съответства на (*B–V*) = –0.03 mag. Сивата непрекъсната линия представя средната стойност на точките <*B–V*> = 0.37mag. Долу: времева еволюция на цвета (*B–R*)

# 5. ЗАВИСИМОСТ МАКСИМАЛЕН БЛЯСЪК–СКОРОСТ НА ПАДАНЕ НА БЛЯСЪКА (MMRD)

Schafter et al. (2011) [16] изследват зависимостта максимален блясък–скорост на падане на блясъка (MMRD) от спектралния клас на новите, като използват над 30 различни нови за период от 20 години. Анализът показва, че се наблюдава линейна зависимост, като най-силно се отклоняват от нея новите от спектрален клас He/N и He/Nn (вж. фиг. 19 в [16] Schafter et al., 2011). Въпреки значителната дисперсия, данните могат да бъдат апроксимирани с права линия – най-ярките нови най-общо са най-бързи, т.е. имат най-малко време за падане на блясъка с 2 зв. величини. За *R*-филтър авторите получават зависимост във вида $M_R = (10.89 \pm 0.12) + (2.08 \pm 0.077) * \log(t_2)$. При преминаване към абсолютна зв. величина е използван модул на разстоянието 24.38 mag ([8] Freedman et al., 2001) и галактична екстинкция $E(B-V) = 0.062$ mag ([15] Schlegel et al., 1998).

Фиг. 6 е построена, като са използвани данните (кръгчета) и линейният фит (правата) от Schafter et al. (2011) [16]. Резултатите, получени в това изследване, са показани с обърнати триъгълници. Използвани са същият модул на разстояние и е направена корекция само за галактичната екстинкция. За максимален блясък сме приели най-ярките точки от кривата на блясъка на M31N 2016-07c и M31N 2016-07e.

Вижда се, че M31N 2016-07c попада близко до регресионната права, получена от Schafter et al. (2011) [16], което показва, че действително новата е наблюдавана в/или много близо до максимален блясък. M31N 2016-07e попада далеч от струпването на точки и от правата. Звездата е по-слаба от очакваното и нейното положение сочи за съществено поглъщане, ~1 mag в *R*-филтър, което е в добро съгласие с оценката, направена по-рано от средния цвят $(B-V)$, или за неточно определяне на максималния блясък на новата.



**Фиг. 6.** Зависимост максимален блясък–скорост на падане на блясъка (MMRD) в *R*- филтър. Данните от Schafter et al. (2011) [16] са показани с плътни кръгчета, а регресионната правата има вида $M_R = (-10.89 \pm 0.12) + (2.08 \pm 0.077) * \log(t_2)$. С обърнат триъгълник са показани резултатите, получени в това изследване (вж. раздел 5)

## 6. ЗАКЛЮЧЕНИЕ

В настоящото изследване са построени и анализирани кривите на блясъка на класическата нова M31N 2016-07c и повторната нова M31N 2016-07e. Към оригиналния наблюдателен материал са добавени и данни от литературата, което от своя страна направи възможно определянето на вида на кривата на блясъка на новите според класификационната схема на Strope et al. (2010) [18] и времето и скоростта на падане на блясъка с две звездни величини $t_2$.

Кривата на блясъка на новата M31N 2006-07c се характеризира с дълбок минимум 2 дни след регистрирания максимум и повторно покачване на блясъка, която според Strope et al. (2010) [18] спада към клас D (dust dips). От кривата на блясъка в $R$-филтър са определени $t_2 = 92.1 \pm 23$ дни и $v(t_2) = 0.02$ зв. величини/ден.

Кривата на блясъка на повторната нова M31N 2006-07e е от клас S (smooth) и от нея получихме $t_2 = 12 \pm 1$ дни и $v(t_2) = 0.17$ зв. величини/ден. От средния цвят ($B-V$) на новата направихме оценка за поглъщането по лъча на зрение към нея ~1 mag в $R$-филтър. Това се потвърждава и от положението ѝ върху диаграмата максимален блясък–скорост на падане на блясъка (MMRD) или подсказва за неточно определяне на зв. величина в максимум. Получените резултати за M31N 2016-07c се съгласуват добре с предложената от Schafter et al. (2011) [16] линейна зависимост, построена с данни за 30 нови в M31 за период от 20 години.

**Благодарности.** Резултатите, представени в това изследване, са част от магистърската дипломна работа на Николай Бойчев „Фотометрично изследване на нови в M31“, защитена през 2017 г. в катедра „Астрономия“ на Физическия факултет. Изследването е финансирано от проект на МОН с номер ДН18/10 от 11.12.2017 г. и частично подкрепено от стипендиантската програма „За жените в науката“ за 2015 г.

IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

### REFERENCES

[1] Bode, M.F. *AN*, 2010, **331**, 160B.
[2] Bryan, J. *IAUC*, 1990, **5121**, 2B.
[3] Chinetti, K., M. J. Darnley, S. C. Williams. *ATel*, 2016a, **9296.**
[4] Chinetti, K., M. J. Darnley, N. Blagorodnova, J. D. Neill, S. C. Williams. *ATel*, 2016, **9347.**
[5] Darnley, M. J., M. F. Bode, E. Kerins, A. M. Newsam, J. An et al. *MNRAS,* 2006, **369**, 257.
[6] Ederoclite, A., M. Henze, D. Aguado et al. *ATel*, 2016, **9281.**
[7] Fabrika, S., O. Sholukhova, A. Valeev et al. *ATel*, 2016, **9383.**
[8] Freedman, W. L. et al. *ApJ*, 2001, **553**, 47.

[9] Goranskij, V., E. Barsukova, M. Henze. *ATel*, 2016, **9435.**
[10] Hachisu, I. & Kato, M. *ApJS*, 2006, **167**, 59.
[11] Hatzidimitriou, D., V. Burwitz, S. Duscha, G. Kanbach, W. Pietsch, P. Reig, G. Papamastorakis. *ATel*, 2007, **1131.**
[12] Hornoch, K., M. Henze, H. Kucakova, K. Honkova, M. Wolf. *ATel*, 2016, **9386.**
[13] Massey, Ph., K. F. Neugent, B. M. Smart. *AJ*, 2016, **152**, 62.
[14] Payne-Gaposchkin, C. The Galactic Novae. Amsterdam: North-Holland, 1957.
[15] Schlegel, D. J., D. P. Finkbeiner, M. Davis. *ApJ*, 1998, **500**, 525.
[16] Shafter, A. W., M. J. Darnley, K. Hornoch, et al. *ApJ*, 2011, **734**, 12.
[17] Shafter, A. W., M. Henze, T. A. Rector et al. *ApJS*, 2015, **216,** 34S.
[18] Strope, R.J., B. E. Schaefer, A. A. Henden. *AJ*, 2010, **140**, 34.
[19] Valcheva, A., A. Kurtenkov, E. Ovcharov, A. Kostov , M. Minev, P. Nedialkov. *ATel*, 2016, **9264.**
[20] Wright, A. E. & Barlow, M. J. *MNRAS*, 1975, **70**, 41.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# ENTROPIC THEORY OF GRAVITATION

VICTOR ATANASOV

*Department of Condensed Matter Physics*

*Виктор Атанасов.* ЕНТРОПИЙНА ТЕОРИЯ НА ГРАВИТАЦИЯТА

В настоящата статия конструираме експлицитно Маховска теория на гравитацията на основата на това, че информацията във Вселената не може да бъде унищожена (принцип на Ландауер). Ако няма загуба на информация във Вселената, то трябва да се запази сумата от ентропиите на геометричното и материалнито поле. Предлагаме локален инвариантен израз за ентропията на геометричното поле и формулираме вариационен принцип за ентропийния функционал, който произвежда ентропийни полеви уравнения. Този информационно-теоретичен подход предполага, че геометричното поле не съществува в празна от материя Вселена, материалната ентропия зависи от геометрията, материята може да обменя информация (ентропия) с геометричното поле и квантов кондензат може да канализира енергия в геометричното поле, когато е в конкретно кохерентно състояние. Ентропийните полеви уравнения притежават неинтуитивно директно куплиране между материалните полета и геометричното поле, което действа като резервоар за ентропия. Дискутират се космологични последствия, като появата на космологичната константа, както и експериментални последици, включващи взаимодействие между гравитацията и квантов кондензат. Енергетичният аспект на теорията възстановява репертоара на класическата обща теория на относителността до различна куплираща константа между полетата.

*Victor Atanasov.* ENTROPIC THEORY OF GRAVITATION

We construct a manifestly Machian theory of gravitation on the foundation that information in the universe cannot be destroyed (Landauer's principle). If no bit of information in the Universe is lost, than the sum of the entropies of the geometric and the matter fields should be conserved.

---

*For contact:* Victor Atanasov, Department of Condensed Matter Physics, Faculty of Physics, Sofia University “St. Kliment Ohridski”, 5 boul. J. Bourchier, 1164 Sofia, Phone: +359 2 8161 847, E-mail: vatanaso@phys.uni-sofia.bg

We propose a local invariant expression for the entropy of the geometric field and formulate a variational principle on the entropic functional which produces entropic field equations. This information-theoretic approach implies that the geometric field does not exist in an empty of matter Universe, the material entropy is geometry dependent, matter can exchange information (entropy) with the geometric field and a quantum condensate can channel energy into the geometric field at a particular coherent state. The entropic field equations feature a non-intuitive direct coupling between the material fields and the geometric field, which acts as an entropy reservoir. Cosmological consequences such as the emergence of the cosmological constant as well as experimental consequences involving gravity-quantum condensate interaction are discussed. The energetic aspect of the theory restores the repertoire of the classical General Relativity up to a different coupling constant between the fields.

***Keywords:*** general theory of relativity and gravitation, quantum information, quantum statistical mechanics

***PACS numbers:*** 04.90.+e, 03.67.-a, 05.30.-d

# 1. MOTIVATION

Information is physical, that is the erasure of one bit of information increases the entropy of the Universe by $k_B \log 2$ where $k_B$ is Boltzmann's constant, that is the Landauer's principle [1]. By bit here we will understand a unit of information, a variable (or binary choice) that can take only the values 0 or 1 and by information content of an object – the size of the set of instructions (number of bits) necessary to construct the object or its state. The Landauer's principle was confirmed experimentally [3] and led to the resolution of the Maxwell's demon paradox [2], in our view an example of not well explored connection between physics and information.

Maxwell's demon collects information about a system, lowers its entropy and increases its energy. The demon's operation has a net result of total conversion of heat into work, a process forbidden by the second law of thermodynamics. Bennett showed, that measurements performed by the demon in order to prepare the system for energy extraction can be done without doing work but that leaves the only one other possibility consistent with the second law, namely the erasure step, required to return the demon's memory (erase his knowledge) to its original state necessary to perform the cycle again, is dissipative. Thus, Landauer's principle resolves a century old thermodynamic gedanken paradox: the work gained by the engine $W_{\text{extracted}} = k_B T \log 2$ is converted into heat in the process of erasure of information from the demon's physical mind $W_{\text{erasure}} = -k_B T \log 2$, therefore no net work is produced in the cycle.

The resolution of the Maxwell's demon paradox with the help of the Landauer's principle points to the equivalence of conservation of energy and the conservation of information. However, the conservation of information

$$\sum_{j}^{N} I_j = \text{const}, \tag{1}$$

where $I_j$ is the information associated with a physical object or field, is not an explored venue in physics and its consequences are largely unknown. In this paper we will suppose that the information content of the Universe is constant and employ a variational principle to find the entropic equations of motion as information lost in one system is transferred as information gained in other system

$$\sum_{j}^{N} \delta I_j = -\sum_{j}^{N} \delta S_j = 0. \tag{2}$$

These systems might be a quantum condensate and a geometric, that is gravitational field. We will introduce a concept of information content (entropy) of the geometric field. The concept of information content of the matter fields is given by the von Neumann entropy. Throughout this paper by information we will understand

$$I = S_{\max} - S, \tag{3}$$

that is the difference between the maximal possible entropy of the system (total disorder; no knowledge of the microstate the system is in) and the temporary entropy.

The introduction of an appropriate expression for the entropy of the geometric field goes through the recognition that an entropy can be associated with a black hole. This entropy is equal to a quarter of the horizon area in Planck units of length [4]:

$$S \infty \textit{ Area of the Horizon.} \tag{4}$$

It has also been noted that the dependence of gravitational entropy on area is not restricted to black holes only but can be asociated with the entanglement entropy in quantum (conformal) field theories via AdS/CFT correspondence [5].

Assuming a number of restrictions, the entropy in a region of space is limited by the area of its boundary, that is the holographic principle [6]. In other words, the entropy bound [4] restricts that the physics in a region of space is characterised by the amount of information that fits on its boundary surface, at one bit per unit of Planck area $l_p^2 \approx 2.6\times10^{-66}\text{cm}^2$. Susskind's use of the term „hologram“ is justified by the hypothetic existence of preferred surface in space-time, kind of cosmological horizon, onto which all of the data in the universe can be imprinted at the same rate of a bit per unit of Planck area.

Alternatively, the main ingredient necessary to derive gravity is information [7]. In Verlinde's view that is the amount of information associated with matter and its location, measured in terms of entropy, defined as the holographic principle dictates. Changes in the entropy when matter is displaced, that is the change in the interference pattern of the hologram on the cosmological screen, leads to an entropic force, which takes the form of gravity. As the author himself admits, to a degree the result is tautological, that is traveling the path backwards from the final destination of black hole thermodynamics and the holographic principle, one is assured to end up at the beginning – the laws of gravity. However, the notion that gravity might actually be an entropic force is compelling.

Unfortunately, the holographic principle is not an idea that might lead to a local theory. Experimentally verified theories in physics are local (excluding entanglement). In local theories the number of degrees of freedom is proportional to volume. The holographic principle postulates that the entropic content is proportional to the area of surface enclosing the embedding volume [8].

We believe, that this weakness is mainly a result of the lack of an appropriate local covariant expression for the entropy of space-time. In the present article we propose such an expression and explore some of its consequences. We retain Verlinde's proposition that entropy might actually be an entropic force but re-write its formulation. We also re-write the standard approach to deriving gravity: instead of formulating a variational principle based on energy conservation, we formulate a variational principle based on information conservation.

## 2. THEORY

Let us assume that the entropy of the geometric field is given by the expression

$$S_G = \frac{1}{L^2 R}, \tag{5}$$

which is manifestly covariant, provided $R$ is the Ricci scalar curvature and $L$ is a unit of length, which at the present stage is a parameter in the theory to be built. This parameter might be equal to the Plank length or the Einstein length $L = \sqrt{1/\Lambda}$, where $\Lambda \approx 1.19\times10^{-52}\mathrm{m}^2$ is the cosmological constant, proportional to the energy density of the vacuum. The inverse of the Ricci scalar curvature is a local geometric object with the dimension of surface area

$$\frac{1}{R} \propto Area. \tag{6}$$

Indeed, this form of the local entropy suffers from a divergence in the case of $R$=0, that is flat space is associated with maximal infinite entropy, but this situation can be remedied by offsetting the singular case

$$S_G = \frac{1}{L^2\left(R+R_0\right)}. \tag{7}$$

This fixes the entropy of flat space to some maximal constant value. We are not going to explore such theory with a cut-off and focus on the main consequences of (5).

The expression for the geometric entropy can serve as a density to a more generalised estimate of the geometric entropy

$$S_G\sqrt{-g}d^4x, \tag{8}$$

which will be the main ingredient of our theory

$$\int S_G\sqrt{-g}d^4x+\int S_{QM}\sqrt{-g}d^4x=\text{const}, \tag{9}$$

where $S_{QM}$ is the entropy density of the material fields. This is the mathematical expression of the idea that information in the universe cannot be destroyed but transferred from the quantum system to the entropy reservoir, that is the geometric field, and back.

The variation of the geometric entropy amounts to

$$\delta\sqrt{-g}S_G=\frac{f'(R)}{L^2}\sqrt{-g}\delta\left(g_{\mu\nu}\right)R^{\mu\nu}+\frac{f(R)}{L^2}\delta\sqrt{-g}+\frac{1}{L^2}\left(f'(R)\sqrt{-g}g_{\mu\nu}\right)\delta R^{\mu\nu},$$

where $f'(R) = 1/R^2$ is the derivative with respect to $R$ of $f(R) = 1/R$. This expression can be varied with respect to the components of the metric tensor and the coefficients of the affine connection upon recognition of the relation $\delta R_{\mu\nu}=-\nabla_\alpha\delta\Gamma^\alpha_{\mu\nu}+\nabla_\nu\delta\Gamma^\alpha_{\mu\alpha}$ and integration by parts [9]

$$\begin{aligned}\frac{\delta\sqrt{-g}S_G}{\delta\Gamma^\alpha_{\mu\nu}}=\frac{1}{L^2}\Big[&\nabla_\alpha\left(f'(R)\sqrt{-g}g^{\mu\nu}\right)-\frac{1}{2}\nabla_\nu\left(f'(R)\sqrt{-g}g^{\mu\nu}\right)\delta^\nu_\alpha\\&-\frac{1}{2}\nabla_\mu\left(f'(R)\sqrt{-g}g^{\nu\mu}\right)\delta^\mu_\alpha\Big].\end{aligned} \tag{10}$$

Here we have symmetrised the above relation with respect to the indices $\mu$, $v$. We have used the Palatini variational method as a mathematical tool to treat the theory at hand and by introducing new independent fields (the coefficients of affine connection) to lower the order of the entropic field equations.

Its variation with respect to the metric is given by

$$\frac{\delta S_{\mathrm{G}}}{\delta g_{\mu\nu}}=\frac{1}{L^2}\left[f'(R)R^{\mu\nu}-\frac{1}{2}g^{\mu\nu}f(R)\right]=-\frac{1}{L^2R^2}\left[R^{\mu\nu}+\frac{1}{2}g^{\mu\nu}R\right]=0.$$

Note, setting it to zero means that the information we have for the space-time is maximal, that is the entropy is minimal and therefore its change (variation) as the embedding space changes its lengthscale is also vanishing.

Indeed the "vacuum" entropic equation of motion

$$\lim_{R\to\infty} S_G = 0 \qquad \& \qquad \lim_{R\to\infty} \frac{\delta S_{\mathrm{G}}}{\delta g_{\mu\nu}} = 0 \tag{11}$$

is vanishing when the geometry of space-time is singular, that is in this entropic theory of gravity $R$ – 0 corresponds to infinite entropy of space-time and infinite variation.

$$\lim_{R\to 0} S_G = \infty \qquad \& \qquad \lim_{R\to 0} \frac{\delta S_{\mathrm{G}}}{\delta g_{\mu\nu}} = \infty . \tag{12}$$

In effect, flat space-time appears to be an entropic reservoir with infinite capacity and the singular points are regions where the entropy and its variation are vanishing in accord with the classical picture. Disorder is vanishing and the information about space-time is maximal at singularities. However, $R$ – 0 and $R = \infty$ do not define affine connection. In order for affine connection to emerge in the flat space-time case matter content is necessary. The same is true for the singularity case.

The variation with respect to the affine connection coefficients yields [10, 9]

$$\frac{\delta\sqrt{-g}S_G}{\delta\Gamma^{\alpha}_{\mu\nu}} = 0 \qquad \Rightarrow \qquad \nabla_{\alpha}\left(f'(R)\sqrt{-g}g^{\mu\nu}\right) = 0, \tag{13}$$

by multiplying with $g^{\nu\alpha}$ and contracting the indices $\nu$, $\alpha$ with the help of the Kronecker symbol. This relation can be identified as the equivalent of $\nabla_{\alpha}\tilde{g}^{\mu\nu} = 0$ and entails

$$\Gamma^{\alpha}_{\mu\nu} = \left\{ \begin{matrix} \alpha \\ \mu\ \nu \end{matrix} \right\}_{\text{with respect to } \tilde{g}} , \tag{14}$$

that is the Γ's are given by the Christoffel brackets provided there exists a new metric $\tilde{g}^{\mu\nu} = f'(R)\sqrt{-g}g^{\mu\nu}$. Transforming conformally the metric

$\tilde{g}_{\mu\nu} = g_{\mu\nu}/(f'(R)\sqrt{-g})$ back to the metric $g_{\mu\nu}$ leads to [11]

$$\Gamma^{\alpha}_{\mu\nu} = \left\{ \begin{matrix} \alpha \\ \mu\ \nu \end{matrix} \right\} - \left[ \delta^{\alpha}_{(\mu}\partial_{\nu)} - \frac{1}{2} g_{\mu\nu} g^{\alpha\sigma}\partial_{\sigma} \right] \log\left(f'(R)\sqrt{-g}\right). \tag{15}$$

Here we have symmetrised with respect to the indices $\mu$, $\nu$ and the first term is the Christoffel bracket with respect to the metric $g_{\mu\nu}$.

Since the vacuum equations

$$\frac{\delta\sqrt{-g}S_G}{\delta\Gamma^{\alpha}_{\mu\nu}} = 0 : \qquad S_G\big|_{R\to 0} = \infty \quad \& \quad \nexists\, \Gamma^{\alpha}_{\mu\nu} , \tag{16}$$

imply that for $R \to \infty$ the expression $\partial_{\alpha}\log f'(R) \to \infty$, as well as $R \to 0$ leads to $\partial_{\alpha}\log f'(R) \to \infty$ we have no solution for $\Gamma^{\alpha}_{\mu\nu}$.

An important point that needs to be addressed is that (15) does not appear to define the connection since $f'(R)$ as a function of the Ricci scalar $R$ involves the derivatives of the connection. Therefore, in any theory we should be able to express $R$ in terms of the quantum matter variation of entropy which should contain only components and derivatives of the metric $g_{\mu\nu}$. Note, the vacuum case implies that thermodynamically preferred is the state of non-existence of the geometric field. A space-time can emerge only in the presence of matter

$$\frac{\delta\sqrt{-g}S_G}{\delta\Gamma^{\alpha}_{\mu\nu}}=-\frac{\delta\sqrt{-g}S_{QM}}{\delta\Gamma^{\alpha}_{\mu\nu}}\ , \tag{17}$$

which is both an equation for the coefficients of the affine connection and an indication that the theory is manifestly in accord with the ideas of Ernst Mach and bishop G. Berkeley, namely the local geometry is determined by the matter distribution in the Universe. According to Bondi-Samuel classification of statements of the Mach principle, Mach 7 fits the ideas here very closely [12]: *If one takes away all matter, there is no more space.* Note, by lack of space our theory assumes $\nexists\,\Gamma^{\alpha}_{\mu\nu}$. Therefore, the divergence of the expression for the geometric entropy for $R \rightarrow 0$ is not a serious drawback of the theory. We will show that *the presence of matter will introduce a cut-off which is intimately associated with the cosmological constant.*

Now let us explore the case when quantum matter, which carries its amount of information (entropy) fills space and can exchange information (entropy) with the entropy reservoir, namely the geometric field.

Looking for the stationary points of the information functional (9) by varying with respect to the metric components $g_{\mu\nu}$ yields

$$\frac{1}{L^2R^2}\left[R^{\mu\nu}+\frac{1}{2}g^{\mu\nu}R\right]=-\frac{1}{2}g^{\mu\nu}S_{QM}+\frac{\delta S_{QM}}{\delta g_{\mu\nu}}\ . \tag{18}$$

Contracting the indices by multiplying with $g_{\mu\nu}$ produces a scalar theory

$$R=\frac{3}{L^2\left[-2S_{QM}+\dfrac{\delta S_{QM}}{\delta g}\right]}, \tag{19}$$

which gives an expression for

$$f'(R)=-\frac{1}{R^2}=-L^4\left[-\frac{2}{3}S_{QM}+\frac{1}{3}\frac{\delta S_{QM}}{\delta g}\right]^2 \tag{20}$$

in terms of the matter content. This expression for $f'(R)$ is a function of the metric and its derivatives at most, therefore (15) defines affine connection and completes the geometric field.

However, the affine connection is governed by (17). The geometric field part is given by (10), now we need to find out an expression for the material part. Let us consider the simplest case of a non-interacting quantum condensate which possesses only kinetic energy

$$H = -\frac{\hbar^2}{2m}\Delta_{\mathrm{LB}} = -\frac{\hbar^2}{2m}\left(\hat{g}^{jk}\partial_j\partial_k - \hat{g}^{jk}\hat{\Gamma}^l_{jk}\partial_l\right) \tag{21}$$

where $\Delta_{\mathrm{LB}}$ is the Laplace-Beltrami operator defined on a three-dimensional hyperplane of space-time. $\hat{g}_{jk}$ is the three-dimensional metric and $\hat{\Gamma}^l_{jk}$ is the Christoffel symbol defined by $\hat{g}$. Provided there is curvature in the embedding space the kinetic energy of the quantum matter becomes length-scale dependent.

The variation of the energy operator with respect to the metric components is given by

$$\delta H = -\frac{\hbar^2}{2m}\delta\hat{g}^{jk}\left(\partial_j\partial_k - \hat{\Gamma}^l_{jk}\partial_l\right),$$

which yields (22)

$$\frac{\delta H}{\delta g^{\mu\nu}} = -\frac{\hbar^2}{2m}\left(\partial_\mu\partial_\nu - \hat{\Gamma}^l_{\mu\nu}\partial_l\right).$$

Note the variation of the three dimensional metric with respect to the four dimensional one is easily defined on 3+1 space-time. Here the non-vanishing contribution comes from indices running in the space part only. By contracting the indices we can obtain a scalar function which meaning might be clearer

$$\frac{\delta H}{\delta g} = g^{\mu\nu}\frac{\delta H}{\delta g^{\mu\nu}} = -\frac{\hbar^2}{2m}\left(\partial^\mu\partial_\mu - \hat{\Gamma}^\nu_{\mu\mu}\partial_\nu\right) \tag{23}$$

Indeed, for almost flat space $g_{\mu\nu} = \eta^\nu_\mu + h_{\mu\nu}$ and keeping the zero-th order terms

$$\frac{\delta H}{\delta g} \approx -\frac{\hbar^2}{2m}\partial^\mu\partial_\mu = \frac{p^2}{2m} = H_{\mathrm{flat}}. \tag{24}$$

This equals the energy operator for the quantum gas in flat space-time. The variation of the energy operator with respect to the components of the affine connection is given by

$$\delta H = -\frac{\hbar^2}{2m}\left(-\hat{g}^{jk}\delta\hat{\Gamma}^l_{jk}\partial_l\right)$$

which yields

$$\frac{\delta H}{\delta\Gamma^\alpha_{\mu\nu}} = -\frac{\hbar^2}{2m}\left(-\hat{g}^{\mu\nu}\partial_\alpha\right) \tag{25}$$

an expression which for almost flat space reduces to

$$\frac{\delta H}{\delta \Gamma^{\alpha}_{\mu\nu}} \approx \frac{i\hbar}{2m}\left(-i\hbar\partial_{\alpha}\right) = \frac{i\hbar}{2m} p_{\alpha}. \tag{26}$$

This is proportional to a component of the momentum defined for the flat space case.

The material part of the theory might be re-defined with the use of another expression for the kinetic energy term:

$$H = -\frac{h^2}{2m}\Delta_{\mathrm{LB}} = -\frac{h^2}{2m}\frac{1}{\sqrt{|\hat{g}|}}\partial_j\left(\sqrt{|\hat{g}|}\hat{g}^{jk}\partial_k\right) \tag{27}$$

where $\Delta_{\mathrm{LB}}$ is the Laplace-Beltrami operator defined on a three-dimensional hyperplane of space-time. $|\hat{g}| = \det(\hat{g})$ is the determinant of the three-dimensional metric. With this definition for the energy of the condensate

$$\frac{\delta H}{\delta \Gamma^{\alpha}_{\mu\nu}} = 0 \tag{28}$$

the variation with respect to the affine connection is vanishing and the major difficulty in solving the theory can be transferred to the metric part. Thus the expression (17) can be reduced to (13) since the entropy of the material part, that is the entropy of the quantum system is a function of its energy:

$$S_{QM}(H) = -\mathrm{tr}\left(\rho\log\rho\right) = \frac{1}{Z}\mathrm{tr}\left(\beta H e^{-\beta H}\right) + \log Z, \tag{29}$$

where $Z = \mathrm{tr}\, e^{-\beta \mathrm{H}}$ is the statistical sum emerging from the condition $\mathrm{tr}\,\rho = 1$. Here the statistical operator is equal to (see Appendix)

$$\rho = \frac{e^{-\beta H}}{Z}, \tag{30}$$

where $\beta = 1/k_{\mathrm{B}} T$. Here $k_{\mathrm{B}}$ is Boltzmann's constant and $T$ the absolute temperature.

Now we are in a position to write the variation of the entropy of the quantum system in terms of the variation of its total energy operator:

$$\delta S_{QM} = \frac{S_{QM} - \log Z}{Z}\mathrm{tr}\left(\beta\delta H e^{-\beta H}\right) - \frac{1}{Z}\mathrm{tr}\left(\beta H \beta\delta H e^{-\beta H}\right). \tag{31}$$

The value of this expression is revealed if one considers (27), therefore (28) and the variation of the entropy with respect to the components of the affine connection is vanishing. The final entropic field equation is then

$$\frac{\delta\sqrt{-g}S_G}{\delta\Gamma^{\alpha}_{\mu\nu}} = -\frac{\delta\sqrt{-g}S_{QM}}{\delta\Gamma^{\alpha}_{\mu\nu}} = 0, \tag{32}$$

which solution is given by (15) and (19).

Regarding the first form of the Hamiltonian (21), the variation of the entropy with respect to the components of the metric tensor becomes

$$\frac{\delta S_{QM}}{\delta g_{\mu\nu}} = \frac{S_{QM} - \log Z}{Z} \sum_a \beta \langle \psi_a | \frac{\delta H}{\delta g_{\mu\nu}} | \psi_a \rangle e^{-\beta E_a} - \\ -\frac{1}{Z} \sum_b \beta E_b \beta \langle \psi_b | \frac{\delta H}{\delta g_{\mu\nu}} | \psi_b \rangle e^{-\beta E_b} \quad . \tag{33}$$

where $E_a = \langle \psi_a | H | \psi_a \rangle$. Therefore, the contracted quantity for almost flat space $H \approx H_{\text{flat}}$, that is using (24),

$$\frac{\delta S_{QM}}{\delta g} \approx \left(S_{QM} - \log Z\right)^2 - \beta^2 \langle E^2 \rangle, \tag{34}$$

where $\langle E^2 \rangle$ denotes mean squared energy. The geometric field part (19) is given by

$$R \approx \frac{3L^{-2}}{-2S_{QM} + \left(S_{QM} - \log Z\right)^2 - \beta^2 \langle E^2 \rangle}, \tag{35}$$

which also solves (15).

Next, the variation with respect to the affine connection gives

$$\frac{\delta S_{QM}}{\delta \Gamma^{\alpha}_{\mu\nu}} = \frac{S_{QM} - \log Z}{Z} \sum_a \beta \langle \psi_a | \frac{\delta H}{\delta \Gamma^{\alpha}_{\mu\nu}} | \psi_a \rangle e^{-\beta E_a} - \\ -\frac{1}{Z} \sum_b \beta E_b \beta \langle \psi_b | \frac{\delta H}{\delta \Gamma^{\alpha}_{\mu\nu}} | \psi_b \rangle e^{-\beta E_b} \quad , \tag{36}$$

which due to (26) becomes

$$\frac{\delta S_{QM}}{\delta \Gamma^{\alpha}_{\mu\nu}} \approx \frac{i\hbar}{2m} \beta (S_{QM} - \log Z) \langle p_\alpha \rangle - \frac{i\hbar}{2m} \beta^2 \langle E p_\alpha \rangle, \tag{37}$$

where $\langle p_\alpha \rangle$ denotes the mean value of the $\alpha$ component of the momentum and $\langle E p_\alpha \rangle$ is the mean value of combined operator $Hp_\alpha$ Note, for a quantum condensate this quantity is vanishing due to the vanishing of the canonical momentum

$$\frac{\delta S_{QM}}{\delta \Gamma^{\alpha}_{\mu\nu}} \approx 0.$$

and for a quantum condensate (32) still applies. In this case, the equation for the connection (15) holds as long as $f'(R)$ can be expressed as a function of the metric components and its derivatives. This requirement is fulfilled by the expression for $R$, that is (35).

Let us try to estimate the large volume $V \to \infty$ limit to the statistical sums $S_{QM}$ and $\delta S_{QM} / \delta g$ using the standard statistical mechanics rule [13]

$$E_a \to \frac{p^2}{2m} \quad \& \quad \sum_a \to \frac{V}{h^3} \int_0^\infty dp\, 4\pi p^2 .$$

Effectively,

$$\begin{aligned} &\frac{\delta S_{QM}}{\delta g} \to -\frac{3}{2}, S_{QM} \to \frac{3}{2} + \log \frac{V}{\lambda_{\text{th}}^3}, \\ &-2S_{QM} + \frac{\delta S_{QM}}{\delta g} \to -\frac{9}{2} - 2\log \frac{V}{\lambda_{\text{th}}^3}, \end{aligned} \tag{38}$$

where $\lambda_{\text{th}} = \sqrt{\dfrac{2\pi h^2}{m k_{\text{B}} T}}$ is the thermal wavelength, which for the purpose of building an intuition takes the following characteristic values: i.) hydrogen atom at the temperature of the Universe 2.73 K $\lambda_{th} \approx 1$ nm; ii.) liquid helium at lambda point 2.17K $\lambda_{th} \approx 0{,}6$ nm; iii.) superconducting Cooper pairs at liquid helium temperature 4.2K $\lambda_{th} \approx 26$ nm.

As a result the entropic field equations have the following behaviour at large volume

$$R \approx -\frac{1}{L^2 \left( \dfrac{2}{3} \log \dfrac{V}{\lambda_{\text{th}}^3} \right)} \tag{39}$$

which entails cosmological consequences as the large volume limit suggests that if the principle matter in the universe is hydrogen gas and the volume of the present epoch Universe is $V \approx 10^{80}\text{m}^3$, [14], then $\frac{2}{3} \log \frac{V}{\lambda_{\text{th}}^3} \propto \times 10^2$. Therefore, the length-scale constant in the expression for the entropy density of the gravitational (geometric) field is $L = \sqrt{\dfrac{1}{10^2 \Lambda}}$, where $\Lambda$ is the cosmological term, estimated at $10^{-52}\text{m}^{-2}$, that is $L$ is of the order of one tenth of the Einstein length, then

$$R \approx -\Lambda \tag{40}$$

Therefore, *the quantum matter content can account for the small negative curvature of space-time, which we term dark-energy.*

Now, suppose the quantum condensate is a Bose-Einstein gas at ground state, that is $S_{QM} = 0$ and

$$\langle E^2 \rangle = \frac{(\log Z)^2}{\beta^2} = (2\log Z)^2 \left(\frac{1}{2} k_\mathrm{B} T\right)^2, \tag{41}$$

the entropic field equation (35) predicts

$$R \to \infty, \tag{42}$$

that is the geometry of space time diverges while the entropy of the geometric field $S_G \to 0$. For a non-interacting Bose-Einstein gas the conditions for this phenomenon can be further specified

$$2\log Z = 3 \quad \lambda_\mathrm{th} = V^{1/3}/\sqrt{e}. \tag{43}$$

The general solution for $S_{QM} \neq 0$ leading to singularity is given by

$$\log Z = S_{QM} + \sqrt{2S_{QM} + \beta^2 \langle E^2 \rangle}. \tag{44}$$

Suppose, we have a Cooper pair gas in the interior of a superconductor at liquid helium temperature. The thermal wavelength in this case is of the order of 26 nm. Since $V^{1/3}$ is a characteristic length for the gas, we may assume it is few times the coherence length for the Cooper pairs $\xi_0 \approx 80$ nm in the case of Pb ($V^{1/3}/\sqrt{e}$ 48 nm). At temperature of around 1mK, the Cooper pair thermal wavelength is $\lambda_\mathrm{th} \approx 1.7$ μm which is appropriate to test the predicted effect.

The predicted divergence in the underlying geometry as the quantum gas falls into coherent state with particular characteristics seems to violate energy conservation and appears thermodynamically improbable as $S_G + S_{QM} \to 0$, which is not a preferred process. However, *if such a process is possible locally, it represents a verifiable prediction of the theory.* The curvature of space-time need not diverge into singularity, but at this state any energetic input in the quantum system can channel into the geometric field. Similar process with an underlying energy conservation law emerges in condensed matter context[19].

## 3. ENERGETIC ASPECT

With regard to the energetic aspect of the entropic theory, it can either be i.) associated with the temperature of the compound system in a way similar to black hole thermodynamics

$$dE = TdS \infty \kappa \; d \; Horizon \; area \tag{45}$$

where $\kappa$ is the surface gravity on the horizon, provided we have an expression for the temperature T, or ii.) define the energy of the geometric field (the energy of the matter fields is given by their Lagrangian) and then obtain an expression for the temperature if neccessary.

The energy of the geometric field, by virtue of the Einstein theory of gravity, has been a controversial issue since the inception of the theory principal reason being that the canonical energy-momentum pseudotensor (Landau-Lifshitz pseudotensor) is made up of first derivatives of the metric and these vanish where the frame is locally inertial at any chosen point, that is it doesn't contain much useful information [15]. The Einstein pseudotensor [16] fails to be symmetric. The Cooperstock's hypothesis [17], namely the geometric energy only exists where the energy-momentum tensor is non-vanishing fails to account for the existence of gravitational waves [18].

Here we will use an unique definition of the local energy of the gravitational field which stems from the framework of [19]. A quantum mechanical system constrained to abide a curved hyper-plane of space-time, produces a geometric potential from the kinetic term. An energy conservation relation in a material medium involving the geometric field can be demonstrated. The energy of the geometric field is given by

$$E_G = \frac{\hbar^2}{24m} R^{3\mathrm{D}} \tag{46}$$

where $R^{3\mathrm{D}}$ is the induced three dimensional Ricci scalar curvature, and $m$ is the mass of the bosons subjected to the field (in the framework of [19] these are the Cooper pairs). It is related to the four-dimensional Ricci curvature of space-time with the relation $R = \frac{4}{3} R^{3\mathrm{D}}$ [20].

We are aware that in classical General Relativity the total energy should be an integral over space at a moment in time, not over space-time, and the Ricci scalar does not give a coordinate invariant quantity if we integrate it over space at a moment in time, that is

$$E_G = \frac{1}{4\pi} \int R_0^0 \sqrt{-g}\, d^3x = E_M = \int \left( T_0^0 - T_1^1 - T_2^2 - T_3^3 \right) \sqrt{-g}\, d^3x. \tag{47}$$

Tolman used the pseudotensor to achieve this result and a more elegant approach that did not rely upon the pseudotensor was later applied in [15] which produced the same result.

Here, we adopt a modification that renders an expression with covariant elements for energy-momentum distribution, including the contribution from gravity [21]. This modification covers arbitrary systems with time-dependence, while incorporating the demand that its energy component reduce to the above in the case of a bounded stationary distribution. As a minimum, a change to a space-time integral (i.e. replace $d^3x$ with $d^4x$) to incorporate a proper volume element. As well, we replace the $R_0^0$ component with the complete Ricci tensor in the integrand to have covariant elements. The resulting structure [21]

$$*E_i^k = \frac{1}{4\pi}\int R_i^k \sqrt{-g}\,d^4x \tag{48}$$

has a trace

$$*E_k^k = \frac{1}{4\pi}\int R\sqrt{-g}\,d^4x. \tag{49}$$

that yields the invariant field action integral for general relativity.

We are now in a position to state a definition (up to numerical factor) of the local energy of the gravitational field

$$\mathsf{E}_G = \frac{\hbar^2}{m^*}R. \tag{50}$$

Here $m^*$ is an unknown parameter in the theory with the dimension of mass (Planck mass is an option $m^* = m_p = \sqrt{\hbar c/G}$). The above expression can serve as an energy density to an energy functional

$$E_G = \frac{\hbar^2}{m^*}\int R\sqrt{-g}\,d^4x \tag{51}$$

which can be extended and include matter field via their Lagrangian density $\mathrm{L}_M$:

$$\int\left(\frac{\mathsf{h}^2}{m^*}R + \mathrm{L}_M\right)\sqrt{-g}\,d^4x = \mathrm{const}. \tag{52}$$

The similarity with the Einstein-Hilbert action is obvious and therefore its variation would yield the same equations of motion. In the case of matter free Universe, that is the "vacuum" case we have:

$$\delta E_G = 0 \Rightarrow R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = 0. \tag{53}$$

The entire repertoire of classical General Relativity is preserved. The entropic aspect implies $\lim_{R\to 0} S_G = \infty$ which is thermodynamically favourable up to the non-existence of affine connection (15) in an empty Universe, in accord with the Mach's principle of relativity. However, a minimal amount of matter would lead to the restoration of the affine connection aspect of the geometric field.

One last issue deserves attention, namely the quantum aspect. Since information content can be represented as a set of discrete units called bits, it is only natural to identify the bit as the quantum of information. The amount of dimensionless entropy corresponding to one bit is log 2, therefore a quantum of curvature can be defined with the help of (5) and (40):

$$R = \frac{1}{L^2\log 2} \approx \frac{\Lambda}{\log 2}, \tag{54}$$

where $\Lambda$ is the cosmological constant. The quantum of geometrical energy upon the use of Planck's mass becomes

$$\mathsf{E}_G = \mathsf{h}^{3/2}\sqrt{\frac{G}{c}}\,\frac{\Lambda}{\log 2} \approx 10^{-93}\ \text{eV}, \tag{55}$$

which is vanishing but non-zero. If this is the quantum of energy of gravitational waves, that is the graviton's mass $m_g$, it is well within the bound $m_g < 10^{-22}$ eV/c$^2$ [18].

## 4. CONCLUSIONS

In conclusion, we would like to go through the main ideas explored in the paper. We derive entropic equations for gravity from a Palatini variational principle based on conservation of information (maximalisation of entropy). Palatini variation is used in order to reduce the order of the governing equations and we reckon is not essential to the main results, which are: i.) a local invariant expression for the entropy of the geometric field can be defined; ii.) the geometric field does not exist in an empty of matter Universe, that is the geometric field can be constructed in accord with the Mach's principle; iii.) material entropy is geometry (lengthscale) dependent; iv.) matter can exchange information (entropy) with the geometric field; v.) a quantum condensate can channel energy into the geometric field at a particular coherent state; vi.) an expression for the local energy density of the geometric field can be defined; vii.) Einstein vacuum equations emerge as the energy conservation aspect of the theory and are thermodynamically favoured provided some matter occupies space and viii.) the cosmological constant (the energy density of dark matter) emerges within the large volume limit in the theory pointing to the *non-existence of dark energy in an empty universe.*

The relation of the proposed theory to earlier work is a work in progress, and the consistency with existing experimental results is to be further analysed at the level of recovering the newtonian gravitational interaction.

**Acknowledgements.** The author wishes to acknowledge inspiring discussions with Hristo Dimov, Dimitar Marvakov and Rossen Dandoloff.

## APPENDIX

Here the statistical operator is equal to

$$\rho = \frac{e^{-\beta H}}{Z}, \tag{56}$$

where $\beta = 1/k_{\mathrm{B}}T$. Here $k_{\mathrm{B}}$ is Boltzmann's constant and $T$ the absolute temperature. $Z = \mathrm{tr}\ {}^{e-\beta H}$ is the statistical sum emerging from the condition $\mathrm{tr}\,\rho = 1$.

This solution emerges as a result of imposing a stationary state condition $[H, \rho] = 0$, that is the quantum system is in a pure state: $[H + \delta H, \rho + \delta\rho] = 0$, which leads to

$$[H + \delta\rho, \delta H, \rho] \tag{57}$$

Inserting (30) in the last equation we get

$$\frac{1}{Z}\left[H, -\beta\delta H e^{-\beta H}\right] - \frac{\delta Z}{Z^2}\left[H, e^{-\beta H}\right] = -\frac{1}{Z}\left[\delta H, e^{-\beta H}\right],$$

which can further be reduced using the relation $[A, BC] = [A,B]C + B[A,C]$ to

$$-\beta[H, \delta H]e^{-\beta H} - \beta\delta H[H, e^{-\beta H}] = -[\delta H, e^{-\beta H}]. \tag{58}$$

This last relation is fulfilled identically provided $[H, \delta H] = 0$, that is if the Hamiltonian commutes with its variation (30) is a solution to (57) and the system remains in a pure state: $[H, \rho] = 0$ Explicitly, $\delta\rho = \delta e^{-\beta H}/Z$.

## REFERENCES

[1] Landauer, R. *IBM J. Res. Develop*., 1961 **5**, 183.

[2] Bennett, C. H. *Int. J. Theor. Phys*., 1982, **21**, 905.

[3] Berut, A., et. al. *Nature,* 2012, **483**, 187; Jun, Y., Gavrilov, M. and Bechhoefer, J. *Phys. Rev. Lett.*, 2014, **113**, 190601.

[4] Bekenstein, J.D. *Phys. Rev. D*, 1973, **7**, 2333; Bardeen, J.M., Carter, B. and Hawking, S. *Commun. Math. Phys.*, 1973, **31**, 161.

[5] Ryu, S. and Takayanagi, T., *Phys. Rev. Lett*., 2006, **96**, 181602.

[6] 't Hooft, G. *Dimensional reduction in quantum gravity*, gr-qc/9310026, 1993; Susskind, L., *J. Math. Phys.*, 1995, **36**, 6377. hep-th/9409089.

[7] Verlinde, E.J. *High Energ. Phys*., 2011, 2011: 29. doi:10.1007/JHEP04(2011)029

[8] Bousso, R. *Rev. Mod. Phys*., 2002, **74**, 825.

[9] Schrödinger, E. Space-time Structure, Cambridge University Press, p.107-108, 1950.

[10] Vollick, D. N. *Phys. Rev. D*, 2003, **68**, 063510.

[11] Hawking, S.W. and Ellis, G.F.R. The large scale structure of spacetime, Cambridge University Press, p. 42, 1973.

[12] Bondi, H. and Samuel, J. *Phys. Lett. A*, 1997, **228**, 121. arXiv:gr-qc/9607009 ; Eddington, A.S., Space time and gravitation, Cambridge University Press, p. 164, 1921.

[13] Huang, K. Statistical Mechanics, 2nd ed. John Wiley & Sons., 1987.

[14] Bars, I., Terning, J. Extra Dimensions in Space and Time. Springer, p. 27, 2009.

[15] Landau, L. D. and Lifshitz, E. M. The Classical Theory of Fields, Pergamon Press, chapter 11, sec. 96, 1951.

[16] Einstein, A. *Sitzungsber. preuss. Acad. Wiss.*, 1918, **1**, 448.

[17] Cooperstock, F.I. *Found. Phys*., 1992, **22**, 1011.

[18] Abbott, B. P., et al. *Phys. Rev. Lett.*, 2016, 116, 061102.

[19] Atanasov, V. *Advances in Condensed Matter Physics*, 2018, 1618252 https://doi.org/10.1155/2018/1618252; arXiv:1610.06020

[20] Ficken, F.A. *Ann. Math.*, 1939, **40**, 892.

[21] Cooperstock, F.I. and Dupre, M.J. *Annals of Physics*, 2013, **339**, 531.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# SENSITIVITY OF WRF MODEL RESULTS TO TOPOGRAPHY AND LAND COVER: STUDY FOR THE SOFIA REGION

EVGENI VLADIMIROV[1,2], RENETA DIMITROVA[1], VENTSISLAV DANCHOVSKI[1]

[1] *Department of Meteorology and Geophysics*
[2] *BULATSA, Sofia, Bulgaria*

*Евгени Владимиров, Ренета Димитрова, Венцислав Данчовски*. ЧУВСТВИТЕЛНОСТ НА РЕЗУЛТАТИТЕ НА МОДЕЛА WRF КЪМ ТОПОГРАФИЯТА И ПОДЛОЖНАТА ПОВЪРХНОСТ: ИЗСЛЕДВАНЕ ЗА РАЙОНА НА СОФИЯ

Най-сложна и изключително разнообразна картина на атмосферните процеси в атмосферния граничен слой се наблюдава в комплексен терен и градски райони. Описанието на такива процеси все още остава предизвикателство поради множеството фактори, които действат в мезо- и микромащаби. Две нови бази данни бяха адаптирани и използвани, за да се подобрят моделните резултати. Представена е методологията, която описва подготовката на статичните полета данни. Също така са дискутирани сравнение между резултатите от моделирането, получени с груба и с по-добра разделителна способност за топографията, както и различно представяне на подложната повърхност.

*Evgeni Vladimirov, Reneta Dimitrova, Ventsislav Danchovski*. SENSITIVITY OF THE WRF MODEL RESULTS TO TOPOGRAPHY AND LAND COVER: STUDY FOR THE SOFIA REGION

The most complicated and extremely diverse pattern of atmospheric boundary-layer processes relate to a complex terrain and urban areas. Description of such processes still remain to be a challenge due to a multiplicity of forcing factors, acting primarily at the mesoscale and microscale. Two new datasets were adapted and implemented to improve the model performance. A methodology describing preparation of both static data field is presented. A comparison between modelling results obtained with rough and fine topography and different land-use datasets is also discussed.

***Keywords:*** new static datasets implementation, model sensitivity to topography and land-use, model evaluation

***PACS numbers***: 92.60.-e, 92.60.Sz, 92.60.hf, 92.60.Fm, 47.27.nb, 92.60.Ry, 92.60.Jq, *92.60.jf, 92.60.Vb, 92.60.hk

---

*For contact:* Evgeni Vladimirov, Department of Meteorology and Geophysics, Faculty of Physics, 5 James Bourchier Blvd., 1164 Sofia, Bulgaria, GSM: +359 898 507 567, E-mail: evgeniv@uni-sofia.bg

# 1. INTRODUCTION

The most complicated and extremely diverse pattern of atmospheric boundary-layer processes relate to a complex terrain in response to a multiplicity of forcing factors, acting primarily at the mesoscale and microscale [1–12]. Terrain and land-use inhomogeneity cause thermal circulation domination under quiescent conditions, or significant modification of the large-scale synoptic flow [13–14].

In addition to natural factors urban areas affect considerably local microclimate. It is noteworthy that 54% of the world's population (almost 4 billion) now lives in urban, rather than rural areas. This is expected to swell to 66%, or more than 6 billion, by 2050 [15], as cities become centers of human activity that provide more opportunities for realization. Intense modifications of land surfaces occur through urban development [16], for example, the use of high heat capacity material for construction and roadways affects local microclimates. The differences in energy balance, temperature, humidity, and storm runoff between urban areas and rural surfaces are substantial.

The application of numerical modelling to investigate fundamental scientific tasks and solve operational tasks became more significant with the rapid development of new technology and computer science during the last few decades. This made the problem of improving numerical forecasts an increasingly salient issue, the resolution of which is highly significant to serving society's needs. Noteworthy efforts in improving operational model predictions were made on high-resolution atmospheric modelling in complex terrain. Correctly treating the land surface properties is becoming increasingly important for meteorological models to be able to capture local mesoscale circulation induced by land surface forcing [17]. The main findings from a number of studies are the importance of better description of topography and land-use inhomogeneity leading to better description of urban atmospheric processes such as clouds and precipitation, heat transfer and convection.

Sofia city is located in very complex terrain in close proximity to Vitosha mountain and better description of the topography is necessary. The standard surface datasets distributed with WRF model is global digital elevation (GMTED2010) and US Geological Survey land-use (USGS) datasets with resolution of 30-arc-second, approximately 1 km at 45° latitude. More recent MODerate resolution Imaging Spectroradiometer (MODIS) land-use dataset, modified in the International Geosphere-Biosphere Programme Data and Information System (IGBP-DIS) DISCover are also available for modelling community with resolution of 30-arc-second and 15-arc-second (approximately 500 m at 45° latitude).

Two new datasets have been implemented and adapted in this study: SRTM 1-arc-second (approximately 30 m at considering latitude) digital elevation data (SRTM, NASA; https://lta.cr.usgs.gov/SRTM1Arc) and the more recent CORINE 2012 with 3-arc-second resolution (approximately 90 m at the same latitude) Land

Cover data (CLC2012, EEA; https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012) for Europe. A methodology describing preparation of both static data field is presented. A comparison between modelling results obtained with rough and fine topography data, from one side, and different land-use category datasets, from the other side, are in the main focus of this study. The model performance is estimated by comparison against observations at singular data points.

## 2. MODEL SET UP

The well-known Weather Research and Forecasting (WRF) model, version 3.8.1 is used for the numerical experiments. WRF is a state-of-the-art atmospheric modelling system, designed for broad range of applications. These include idealized simulations for investigation of specific physical processes in the atmosphere, data assimilation, and operational forecast.

High resolution numerical modelling in both directions – horizontal (500 meters step) and vertical (50 irregular stretched vertical levels with greater density in Planetary Boundary Layer – PBL) is required for this study. Two major changes are introduced into the model in order to evaluate some improvements in its performance. A new dataset of topography with 1-arc-second resolution, and new physical surface properties via a more accurate land-use dataset with 3-arc-second resolution are implemented to WRF. A reclassifying procedure is carried out to adapt the new information from CORINE Land Cover 2012 to the existing into the model surface properties from USGS dataset.

Four nested domains are used based on a Lambert Projection (Figure 1), which essentially covered Balkan Peninsula (Domain 1, D1), Bulgaria (Domain 2, D2), Western part of Bulgaria (Domain 3, D3) and Sofia Valley (Domain 4, D4). D1 has 36×44 grid points in the horizontal with 32-km grid-point spacing; D2 – 73×65 8-km cells respectively; D3 – 69×97 2-km cells, and D4 cover Sofia Valley with 157×129 cells with resolution of 500 m.

The initial and boundary conditions (for the parent domain D1) are derived from the 0.25 degree NCEP Final Operational Model Global Tropospheric Analyses (http://rda.ucar.edu/datasets/ds083.2/) on every 6 hours. This product comes from the Global Data Assimilation System, which continuously collects observational data from the Global Telecommunications System. Data assimilation (fdda model option) is used for domain D1 at all vertical levels and for domains D2, D3 only above closest to the ground 10 model levels The fdda option did not apply for the most inner domain of specific interest.

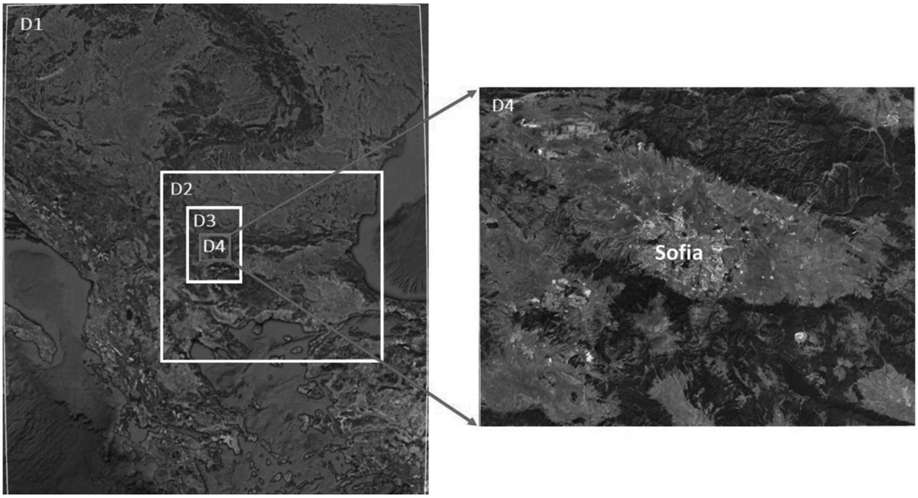


**Fig. 1.** Domains used for numerical calculations with extended view of Sofia valley (500 m grid)

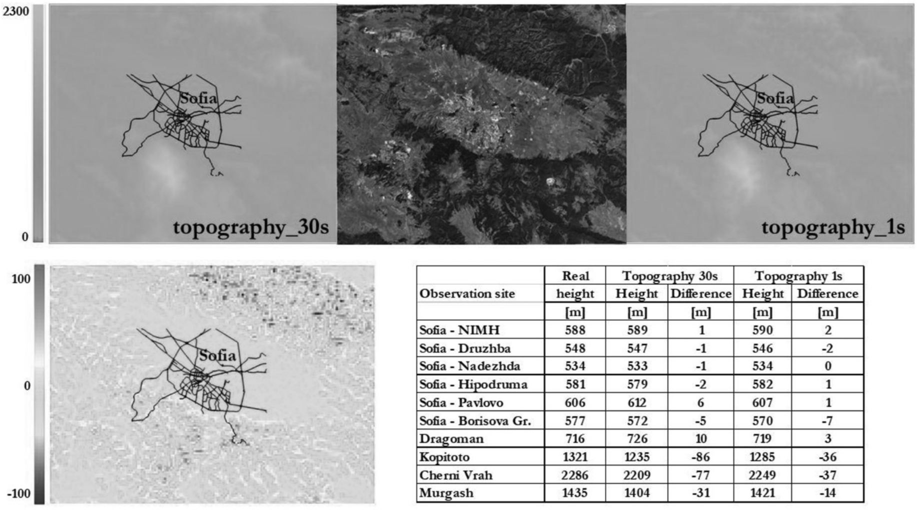
The WRF physics package includes: a sophisticated microphysics scheme [18], the Rapid Radiative Transfer Model (RRTM) longwave radiation parameterization [19], Dudhia shortwave radiation parameterization [20], Noah land surface model [17]. Simplified Arakawa–Schubert cumulus parameterization [21] is used only for domains D1 and D2. For high resolution domains D3 and D4 cumulus formation is resolved explicitly by the microphysics. The most sensitive to local surface inhomogeneity are surface fluxes and skin temperature which depend strongly on surface schemes and related PBL schemes used for simulations. Due to this reason most of the available in WRF model PBL schemes and corresponding surface schemes are used and compared. These include: Asymmetric Convective Model – ACM2 [22], Bougeault and Lacarrere–BouLac [23], Mellor–Yamada–Janjic – MYJ [24], Mellor–Yamada–Nakanishi and Niino Level 2.5 – MYNN2.5 [25], Mellor–Yamada–Nakanishi and Niino Level 3 – MYNN3 [25], Quasi-Normal Scale Elimination – QNSE [26], Bretherton and Park – UW [27] and Yonsei University – YSU [28].

## 3. METHODOLOGY FOR IMPLEMENTAION OF HIGH RESOLUTION EVEVATION AND LAND-USE DATA

WRF uses WPS (WRF pre-processing system) to prepare the boundary and initial condition fields for real-data numerical simulations. WPS is a set of three programs and each of the programs performs one stage of the work: geogrid make the initial fields and interpolates static geographical e.g. terrain and the land cover data, to the specific domain grids; ungrib extracts meteorological fields from global NCEP GRIB-formatted files; and metgrid horizontally interpolates the meteorological fields extracted by ungrib to the model grids defined by geogrid. WPS produces a specific readable binary format files for each field. Therefore to use other than the standard datasets in WRF model it is necessary to convert the downloaded data into WPS readable binary format. There are a number of available software solutions to transform the originally downloaded datasets, the most prominent of which are ARCGIS. We use R- language in this work however [29].

### 3.1. PROCEDURE OF HIGH RESOLUTION TOPOGRAPHY DATA PREPATION

The dataset is downloading from the digital elevation NASA database SRTM1Arc (https://lta.cr.usgs.gov/SRTM1Arc). A visualization tool, such as EarthExplorer (http://earthexplorer.usgs.gov/), is needed to visualize and select only the area corresponding to domain use for the numerical simulations. The selection of the appropriate area allows downloading only a part of the full global dataset. The SRTM data for the region between 22th–24th meridian and 42th–43th parallel are downloaded merged and saved in an ASCII formatted file. A slightly modified version of Fortran90 code, provided by Lorenzo Giovannini (Atmospheric Physics Group Department of Civil and Environmental Engineering, University of Trento) [30] and C program, distributed with the WPS source code, are used to convert the ASCII file into a WPS readable binary format. Very important part of the procedure is creation of index file describing all necessary information for the dataset and its geo-oriented projection. The GEOGRID.TBL file must be changed, adding the path to new dataset and describing the type of interpolation to be used. The pre-processing system is able to implement the new topography field via geogrid module.



| Observation site | Real height [m] | Topography 30s Height [m] | Topography 30s Difference [m] | Topography 1s Height [m] | Topography 1s Difference [m] |
|---|---|---|---|---|---|
| Sofia - NIMH | 588 | 589 | 1 | 590 | 2 |
| Sofia - Druzhba | 548 | 547 | -1 | 546 | -2 |
| Sofia - Nadezhda | 534 | 533 | -1 | 534 | 0 |
| Sofia - Hipodruma | 581 | 579 | -2 | 582 | 1 |
| Sofia - Pavlovo | 606 | 612 | 6 | 607 | 1 |
| Sofia - Borisova Gr. | 577 | 572 | -5 | 570 | -7 |
| Dragoman | 716 | 726 | 10 | 719 | 3 |
| Kopitoto | 1321 | 1235 | -86 | 1285 | -36 |
| Cherni Vrah | 2286 | 2209 | -77 | 2249 | -37 |
| Murgash | 1435 | 1404 | -31 | 1421 | -14 |

**Fig. 2.** Topography with 30 and 1arc-seconds and map of the domain from Google Earth (the upper panel); map of the difference between both datasets and table with differences for specific sites of model validation (down panel)

Figure 2 shows comparison between coarse and high resolution topography data. The difference is in range ± 100 m with the major deviation at high elevation points. High resolution topography reduces the difference in comparison with the real elevation for more of the observational sites, most significant for Kopitoto (50 m reduction) and Cherni Vrach (40 m reduction). The change at the Sofia Valley bottom is not significant.

### 3.2. PROCEDURE OF HIGH RESOLUTION LAND-USE DATA ADAPTATION

U.S. Geological Survey (USGS) Global Land Cover Characteristics data are based on 1-km Advanced Very High Resolution Radiometer (AVHRR) data spanning April 1992 through March 1993. The MODerate-resolution Imaging Spectroradiometer (MODIS) land cover data are coming from the instrument operating on both the Terra and Aqua spacecraft with last update from 2006. Coordination of Information on the Environment (CORINE) programme was initiated in the European Union in 1985. CORINE based on Earth observation satellite data available on a regular basis and updated on every 5–10 years with current update from 2012.

The CLC2012 dataset (http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012) consists of 44 land categories and differ from the standard USGS land-use dataset with 24 categories. Some CORINE categories is combined following the paper of Pineda et al. [31] and adapted to the USGS dataset with the same surface parameters of corresponding land category (heat capacity, moisture properties, albedo, roughness etc.). The downloaded data should be clipped for the region of interest and remapped into 24 USGS categories.

R-language is used for the procedure similar to transformation of the topography dataset – convert the ASCII file into a WPS readable binary file, construct the index file, and finally add the path to new dataset and describe the type of interpolation into GEOGRID.TBL.

Comparison between all four available land-use datasets (three distributed with WRF public version and one new (CORINE adapted to USGS) is shown in Figure 3. Urban areas grow very fast during the last decades (including Sofia) and USGS data do not represent well the current stage of the urban cover. More recent Modis_30s data with the same resolution as USGS (30 arc-sec) displays better the city extension (Figures 3a, 3b). From the other side, high resolution data provide more details on horizontal inhomogeneity and Modis_15s (Figure 3c) better represent details such as small villages and suburban areas, which are not visible in Modis_30s data. The most recent CORINE_3s data has finest resolution and better describe urban area and city park areas, missing in Modis_15s data (Figure 3d).

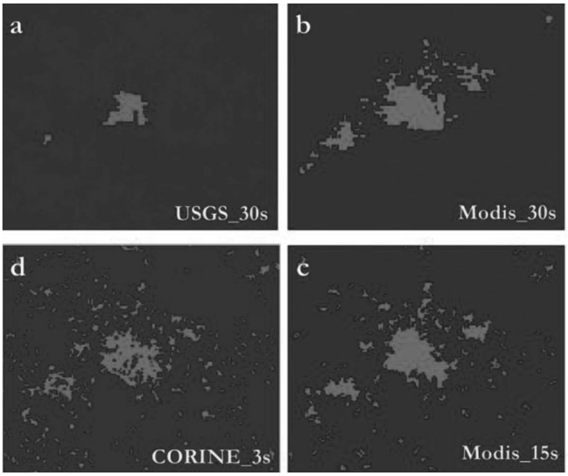


**Fig. 3.** Visualization of urban area that relate to different datasets: USGS_30s (a), Modis_30s (b), Modis_15s (c) and CORINE_3s (d), adapted to USGS categories

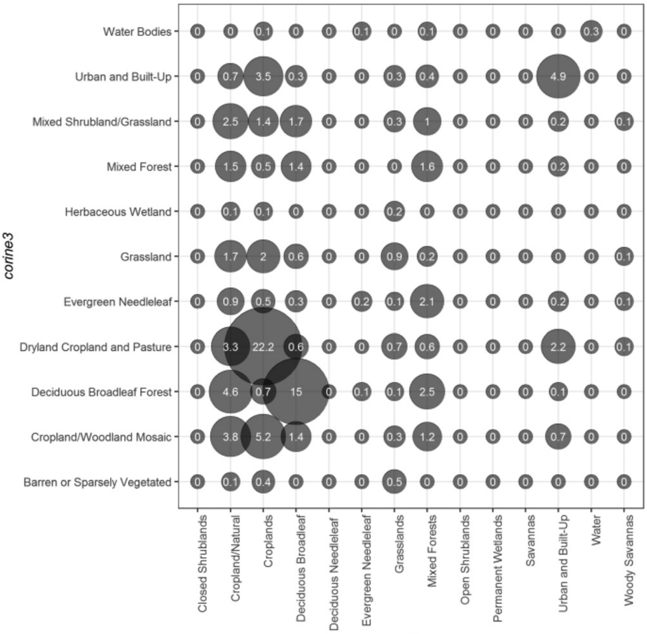


**Fig. 4.** Correspondence between different land-use categories in Modis_15s and CORINE_3s

The correspondence between different land-use categories in high resolution datasets Modis_15s and CORINE_3s is shown in Figure 4. Only presented land-use categories in the area of interest are included in this comparison. There are some differences between categories in both datasets and complete correspondence is hardly to expect. For example, CORINE_3s works with categories "Dryland cropland and Pasture", “Cropland/Woodland Mosaic” and “Mixed Shrubland/ Grassland”, while Modis_15s uses corresponding categories “Cropland/Natural”, “Croplands”, “Grassland” and “Closed Shrublands”. Due to this reason, part of each category from CORINE is spread around different similar or mixed type categories in Modis_15s. The main agreement we have for the categories “Croplands” – 22.2% and “Deciduous Broadleaf Forest” – 15.0%. Approximately a half (4.9%) from the total coverage (10.1%) in CORINE_3s by "Urban and Built-up" category corresponds to the same category in Modis_15s dataset. The total coverage by "Urban and Built-up" category in Modis_15s is 8.5% – less than in CORINE_3s dataset. We have to expect difference in parameters of surface characteristics between both land-use datasets, which will affect the surface fluxes and main meteorological fields near the surface.

## 4. RESULTS AND DISCUTION

The main goal of this study is to estimate possible improvement on model results for Sofia valley and surrounding mountains. All available observations – six sites located in Sofia (Borisova Gradina, NIMH, Druzhba, Hipodruma, Nadezhda, Pavlovo), one in Sofia Valley (Dragoman) and three mountain stations (Cherni Vrah, Murgash, Kopitoto) are used to estimate the model performance. Four of the sites are part from the national meteorological network operated by National Institute of Meteorology and hydrology (NIMH, Cherni Vrah, Murgash, Dragoman) with data taken manually on every 3 hours (synoptic stations) and the rest are automatic stations operated by Bulgarian Ministry of Environment and water (Druzhba, Hipodruma, Nadezhda, Pavlovo, Kopitoto) and Sofia University “St. Kl. Ohridski” (Borisova Gradina). Both static fields – topography and land-used affect the near surface fields of temperature and humidity.

The selected case study covers 3 days (13–16 August, 2016) with considering 12 hours of spin-up. This period is characterized with anticyclonic fair weather, dry (low humidity) and quiescent (wind speed $< 5$ m s$^{-1}$ at 850-hPa) conditions. Observations are available for temperature and relative humidity at 2 m in the area of interest. Due to described above reasons, only temperature is used for model validation as representative for any significant differences expected. The effect of moisture and wind is insignificant and substantial changes using different topography and land-use datasets is not registered.

### 4.1. HIGH RESOLUTION TOPOGRAPHY DATA – EFFECT ON MODEL PREDICTION

Two model runs are carried out using 30 s or 1 s resolution topography datasets and Modis_15s land-use dataset. The effect of more accurate topography representation leads to variance mainly in heat fluxes, skin and near surface (at

2 m) temperature. Both sites with significant difference in elevation from both topography datasets, Kopitoto – 50 m and Cherni Vrah – 40 m (see table in Figure 2) are shown as an example in Figure 5. The effect of using high resolution topography data is most significant at the mountain region showing reduction in temperature approximately with 1–1.5 degrees (Figure 5). WRF underestimates day and overestimate night temperatures at these sites in general. The difference in model results is more significant for temperature maximum prediction. Questionable improvement in model performance for temperature minimum is found. The effect of high resolution topography on model performance at the Sofia Valley is insignificant due to very small differences between both datasets.

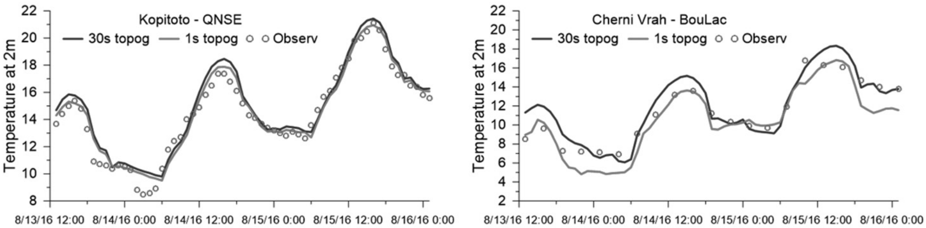


**Fig. 5.** Example of comparison of near surface temperature (in ºC) calculated with rough (30 s) and high (1 s) resolution topography against observations for two selected sites and PBL schemes

**Table 1.** Comparison between Mean Bias and Mean Error for all considering PBL schemes and both topography datasets with different resolution

| PBL scheme | Topography 30s | | Topography 1s | |
|---|---|---|---|---|
| | Mean bias | Mean error | Mean bias | Mean error |
| ACM2 | 0.80 | 1.36 | 0.44 | 1.19 |
| BouLac | 1.06 | 1.24 | 0.66 | 1.04 |
| MYJ | 1.04 | 1.24 | 0.65 | 1.01 |
| MYNN2.5 | 0.57 | 1.32 | 0.21 | 1.16 |
| MYNN3 | -0.15 | 1.18 | -0.50 | 1.32 |
| QNSE | 0.86 | 1.15 | -0.50 | 0.94 |
| UW | 0.89 | 1.12 | 0.52 | 0.95 |
| YSU | 0.67 | 1.16 | 0.34 | 1.11 |

A simple statistics is calculated using data for three available mountain observational sites – Kopitoto, Cherni Vrah and Murgash. Table 1 shows model evaluation for all considering PBL schemes in this study, comparing the results from runs using different topography datasets (with 30s and 1s resolution).

Slight improvement is registered for all PBL schemes except MYNN3, which provides the worst results in this study and behaves differently (significantly underestimate day temperature; see Figure 8) from the other schemes. QNSE and UW schemes show the best performance, following by MYJ and BouLac. More of the PBL schemes demonstrate small positive mean bias (less than 0.7°C), two of them (MYNN3 and QNSE) negative (–0.5°C). Mean error is approximately 1°C and show reduction of modelling errors for all schemes (except MYNN3) using new implemented in WRF topography.

### 4.2. HIGH RESOLUTION LAND-USE DATA – EFFECT ON MODEL PREDICTION

Two model runs are carried out using high resolution (1s) topography and different land-use datasets (CORINE and USGS). Some categories from CORINE was combined and adapted to USGS dataset, as was already mentioned in section 3.2. Comparison between adapted CORINE and USGS maps show significant difference (Figure 6). The “Urban and Build-Up” category covers the extended city area in CORINE and replace “Dryland Cropland and Pasture” and “Cropland/ Grassland/Pasture categories surrounding Sofia city in USGS data. Two of the observational sites NIMH and Druzhba are outside of the city area on USGS map. There is some difference in other categories also, for example city parks are treated as mixed forest and are not represented on USGS map.

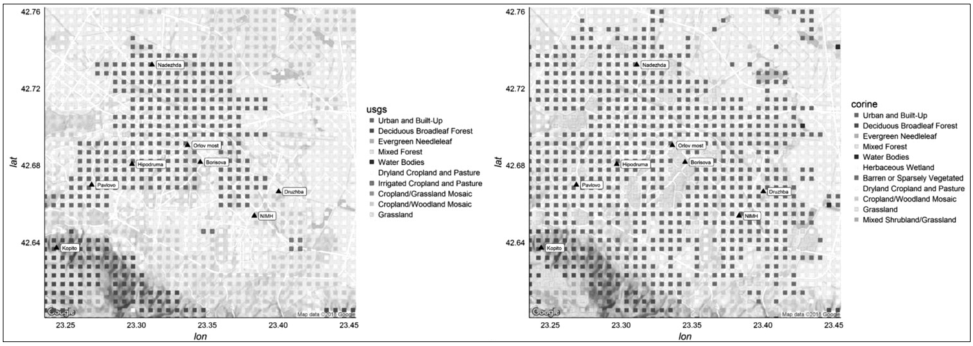


**Fig. 6.** Comparison of different land categories mapped using USGS (30 s) and adapted CORINE high resolution (3 s) land-use data

Model performance is estimated using both land-use datasets with 8 different PBL schemes. A scatterplot and coefficient of determination ($r^2$) are shown in Figure 7. There is notable improvement in model results produced by all different PBL parametrizations using high resolution adapted CORINE data. More of the schemes have coefficient of determination more than 0.9. QNSE scheme shows the best result ($r^2 = 0.93$).

Some of the observational sites show better agreement with measurements than others. One example comparing different sites, located in Sofia (modelling result using QNSE scheme) is shown in Figure 8. Very good agreement with observations is recorded at Borisova Gradina, Nadezhda and Pavlovo sites ($r^2 = 0.97$). WRF underestimates the maximum (day) temperature at two sites Hipodruma and Druzhba, and overestimates the minimum (nocturnal) emperature at NIMH site (not all sites

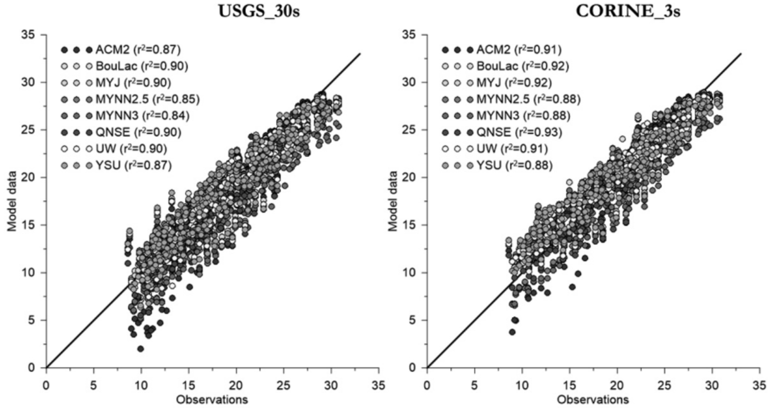


**Fig. 7.** WRF evaluation for eight PBL schemes using USGS and adapted CORINE datasets

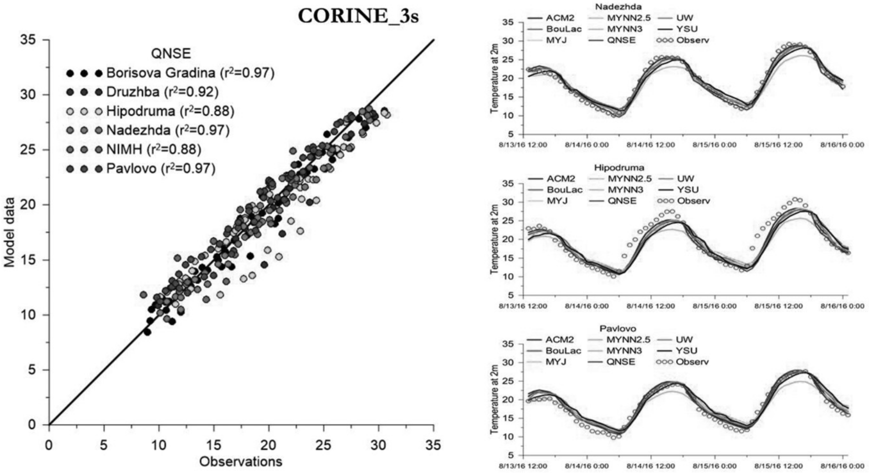


**Fig. 8.** Evaluation of different observational sites for modelling results using adapted CORINE land-use (left panel) and comparison of temperature at 2m for all PBL schemes at selected sites (right panel)

are shown in Figure 8). A shift in time with 1 hour regarding calculated temperature minimum is found at Druzhba, Hipodruma, Pavlovo and Nadezhda sites. Very fast drop in temperature after the sunset at NIMH site is not captured by the model and leads to more significant difference in the time when temperature minimum occurs. Shift with 3 hours later appears in modelling results with all considering PBL schemes in comparison with observations. According to observations the minimum recorded at NIMH occurs at midnight and 3 am for the considering period, but the minimum calculated by the model appears at 5–6 am – early morning, which is in accordance with the remainder of the observational sites at Sofia city. Agreement with observations when temperature minimum occurs is registered at Borisova Gradina only. All sites, except Borisova Gradina, are located in urban area according to model’s land cover map (see Figure 6). In spite of high resolution input data, an interpolation apply to create average land cover category for each model grid cell with 500 m resolution. The land category with the greatest weight is taken ("Urban and Built-up" for example) and small parts with different land categories such as “Dryland Cropland and Pasture” or “Cropland/Grassland/Pasture, typical for small parks are not considered in the model. Note that more of the sites are located in open space areas between the buildings, for NIMH site – large open field. The inconsistency with the land cover at the observational site affects the surface properties and probably is the reason for the shift in temperature minimum of modelling results.

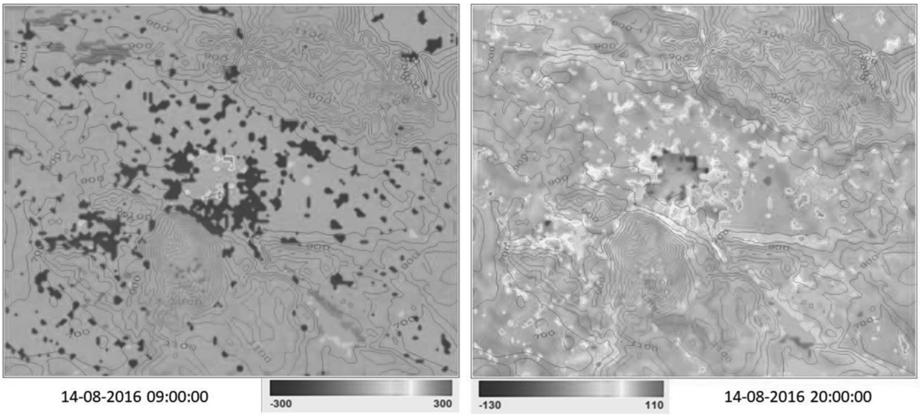


**Fig. 9.** Difference (CORINE-USGS) between ground head fluxes calculated using different land-use datasets; model data with MYJ PBL scheme are shown at 09 and 20 on August 14th, 2016

WRF model provide output for main fluxes at the ground – surface sensible, latent heat and ground heat fluxes that balanced the total net radiation (incoming and outgoing shortwave and longwave radiation). The difference in horizontal fields between model results using adapted CORINE high resolution data and USGS data is shown for the ground heat flux for one PBL scheme – MYJ at selected times (Figure 9).

Significant variances are displayed in the field of ground heat flux for the urban and park areas that appears in Sofia city using adapted CORINE data. Different land categories have diverse surface thermal properties such as thermal inertia, surface heat capacity; surface emissivity and albedo (see Table 2). Thermal inertia is a physical parameter representing the ability of a material to conduct and store heat, and in the context of planetary science, it is a measure of the sub-surface's ability to store heat during the day and reradiate it during the night. Materials with high thermal inertia value show less temperature amplitude during a full heating and cooling cycle (day heating and night cooling process) than those with lower thermal inertia. For example, the large volume of asphalt, brick, concrete and other materials give urban areas a low thermal inertia than rural areas (Table 2). Surface heat capacity is the measure of the increase in thermal energy content or heat per degree of temperature rise. It denotes the capacity of a material to store heat. Surfaces with more heat capacity like mixed forest (Table 2) will need more time and more energy from the sun in order to increase their temperature during the day. The effectiveness in emitting energy as thermal radiation is responsible mainly to faster heating and cooling of the surface in urban areas comparing with non-urban. Combination of all these parameters relate to the ground heat flux. A positive difference in ground heat flux (CORINE – USGS) appears in areas corresponding to park regions in Sofia city from the adapted CORINE categories (see Figure 6). All added "Urban and Build-Up" areas with high resolution data (see Figure 3) lead to negative difference in ground head flux during the first hours after the sunrise. In opposite the sign of the difference of ground heat flux change after the sunset and the magnitude is more than two times less.

**Table 2.** USGS LU categories (the same are used in adapted CORINE data) and their physical parameters for 'summer' season are taken from LANDUSE.TBL in WRF model. Parameters from left to right are: albedo (%), soil moisture availability (×100%), surface emissivity (%), surface roughness length (× $10^{-2}$ m), thermal inertia ($4.184\times10^{2}Jm^{-2}K^{-1}s^{-1/2}$), surface heat capacity ($\times10^{6}Jm^{-3}K^{-1}$)

| Land Categories | ALBD | SLMO | SFEM | SFZ0 | THERIN | SFHC |
|---|---|---|---|---|---|---|
| Urban/ Built-Up Land | 15 | 0.10 | 0.88 | 80 | 3 | 1.89 |
| Dryland/Cropland/ Pasture | 17 | 0.30 | 0.99 | 15 | 4 | 2.50 |
| Cropland/Woodland Mosaic | 16 | 0.35 | 0.99 | 20 | 4 | 2.50 |
| Mixed Forest | 13 | 0.30 | 0.97 | 50 | 4 | 4.18 |

## 5. CONCLUTIONS

Two new datasets have been implemented and adapted in WRF model and used in this study: SRTM 1-arc-second topography and the most recent CORINE Land Cover 2012 with 3-arc-second resolution. A reclassifying procedure based on previous works have been applied to 44 CORINE land categories and remapped to USGS 24 categories. Better representation of topography and land cover lead to change mainly in heat fluxes, skin and near surface temperature (at 2m) and these characteristics were considered in the study.

The application of high resolution topography is the most important for mountain regions, where the changes in topography elevation are significant. An increase in peaks height indicates reduction in calculated temperature approximately with 1–1.5 degrees. Implementation of high resolution topography data in WRF show improvement of model performance for mountain stations for all PBL schemes except MYNN3.

CORINE is the most recent dataset and ensures better description of actual urban areas than USGS dataset. High resolution of the new data allow to increase model's horizontal grid resolution (500 m) and update the land cover to better represent surface characteristics that may have changed in recent years. There is notable improvement in model results produced by all different PBL parametrizations with coefficient of determination between 0.88 and 0.93. Very good agreement with observations is achieved at sites: Borisova Gradina, Nadezhda and Pavlovo sites ($r^2 = 0.97$). WRF underestimates the maximum (day) temperature at two sites Hipodruma and Druzhba, and in contrary overestimates the minimum (night) temperature at NIMH site. Also a shift in calculated minimum temperature is found for all sites except Borisova Gradina. Different land categories have diverse surface thermal properties such as thermal inertia, surface heat capacity, surface emissivity and albedo. Inappropriate representation of land cover at specific sites can be the reason disagreement between the modelling result and observations.

A positive difference in ground heat flux (CORINE–USGS) appears in areas corresponding to park regions in Sofia city from the adapted CORINE categories. All added "Urban and Build-Up" areas with high resolution data lead to negative difference in ground head flux during the first hours after the sunrise. In opposite the sign of the difference of ground heat flux change after the sunset and the magnitude is more than two times less. Therefore using the appropriate dataset is necessary in efforts to improve numerical weather simulations and enhance weather forecasting capability.

**Acknowledgements.** This research is developed within the scope of the project DN4/7 (Study of the PBL structure and dynamics over complex terrain and urban area), funded by Research Fund at the Bulgarian Ministry of Education and Science.

## REFERENCES

[1] Defant, F. In: *Compendium of Meteorology,* T. M. Malone (Ed), Amer. Meteor. Soc. 1951, 655-672.

[2] Holton, J. R. *Tellus*, 1966, **19**, 199-205.

[3] Vergeiner, I., E. Dreiseitl. *Meteor. Atmos. Phys.,* 1987, **36**, 264-286.

[4] Egger, J. *Amer. Meteor. Soc.*, *Meteor. Monogr.,* 1990, **45**, 43-57.

[5] Whiteman, C. D. *AMS Meteorological Monographs*, 1990, **45**, 23, 5-42.

[6] Whiteman, C. D. Mountain meteorology: fundamentals and applications, New York, 2000.

[7] Simpson, J. E. Gravity currents, Cambridge Univ Press, 2nd Edition, 1999.

[8] Sturman, A. P., H. A. McGowan, R. A. Spronken-Smith. *Progress Physical Geography*, 1999, **23**, 611-635.

[9] Poulos, G. S., S. Zhong. *Geography Compass*, 2008, **2**, 1-24.

[10] Parish, T.R., L.D. Oolman. *J. Atmos. Sci.*, 2010, **67**, 2690–2699.

[11] Fernando, H.J.S., E.R. Pardyjak. *EOS,* 2013, **94**, 36.

[12] Fernando, H.J.S., E.R. Pardyjak, S. Di Sabatino, F.K. Chow, S.F.J. DeWekker, S.W. Hoch, J. Hacker, J.C. Pace, T. Pratt, Z. Pu, J.W Steenburgh., C.D. Whiteman, Y. Wang, D. Zajic, B. Balsley, R. Dimitrova, G.D. Emmitt, C.W. Higgins, J.C.R. Hunt, J.G. Knievel, D. Lawrence, Y. Liu, D.F. Nadeau, E. Kit, B.W. Blomquist, P. Conry, R.S. Coppersmith, E. Creegan, M. Felton, A. Grachev, N. Gunawardena, C. Hang, C.M. Hocut, G. Huynh, M.E. Jeglum, D. Jensen, V. Kulandaivelu, M. Lehner, L.S. Leo, D. Liberzon, J.D. Massey, K. McEnerney, S. Pal, T. Price, M. Sghiatti, Z. Silver, M. Thompson, H. Zhang, T. Zsedrovits. *Bull. Amer. Meteorl. Soc.*, 2015, **96**, 11, 1945-1967.

[13] Bodine, D., P. M. Klein, S. C. Arms, A. Shapiro, *Appl. Meteor. Climatol.,* 2009, **48**, 1117–1141.

[14] Dixit, P. N., D. Chen, *Theor. Appl. Climatol.,* 2011, **103**, 533–542.

[15] United Nations, Department of Economic and Social Affairs, Population Division, “World Urbanization Prospects: The 2014 Revision”, (ST/ESA/SER.A/366), 2015.

[16] Changnon, S.A. Jr. *Bull. Am. Meteorol. Soc.*, 1992, **73**, 619–627.

[17] Chen, F., J. Dudhia. *Mon. Weather. Rev.,* 2001, **129**, 569–585.

[18] Lin, Y.L., R.D. Farley, H.D. Orville. *J. Appl. Meteorol.,* 1983*,* **22**, 1065-1092.

[19] Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, S.A. Clough. *J. Geophys. Res. Atmos.,* 1997*,* **102**, 16, 663-16, 682.

[20] Dudhia, J. *J. Atmos. Sci.,* 1989*,* **46**, 3077–3107.

[21] Arakawa, A., and W. H. Schubert. *J. Atmos. Sci.,* 1974, **31**, 674-701.

[22] Pleim, J.E. *J. Appl. Meteorol. Clim*., 2007, **46(9),** 1383–1395.

[23] Bougeault, P., P. Lacarrere. *Mon. Weather Rev*., 1989, **117(8)**, 1872-1890.

[24] Janjic, Z. *Mon. Weather Rev.,* 1990, **118**, 1429-1443.

[25] Nakanishi, M., H. Nino. *Boundary-Layer Meteorol.,* 2006, **119**, 397-407.

[26] Sukoriansky, S., B. Galperin, V. Perov. *Boundary-Layer Meteorol*., 2005, **117(2)**, 231-257.

[27] Bretherton, C.S., S. Park. *Atmosphere Model. J. of Climate*, 2009, **31(2)**, 3422-3448.

[28] Hong, S., Y. Noh, J. Dudhia, *Mon. Weather Rev*., 2006, **134(9)**, 2318-2341.

[29] Arnold, D., I. Schicker, P. Seibert. High-Resolution Atmospheric Modelling in Complex Terrain for Future Climate Simulations (HiRmod), VSC Report, 2010.

[30] Pineda, N., O. Jorba, J. Jorge, J.M. Baldasano. *Int. J. Remote Sens*., 2004, **25(1)**, 129-143.

[31] Core Team. A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. 2013.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ

Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS

Volume 111

# ТЕРМОДИНАМИЧНИ УСЛОВИЯ ПРИ РАЗВИТИЕ НА ГРЪМОТЕВИЧНИ ОБЛАЦИ НАД СЕВЕРОЗАПАДНА И ЮГОЗАПАДНА БЪЛГАРИЯ

БОРЯНА МАРКОВА[1], БОРИС ЯНАКИЕВ[2], РУМЯНА МИЦЕВА[3]

*[1]Национален иститут по метеорология и хидрология, София*
*[2]38-ми Батальон за ядрена, химическа и биологична защита и екология, Мусачево*
*[3]Физически факултет*

*Боряна Маркова, Борис Янакиев, Румяна Мицева*. ТЕРМОДИНАМИЧНИ УСЛОВИЯ ПРИ РАЗВИТИЕ НА ГРЪМОТЕВИЧНИ ОБЛАЦИ НАД СЕВЕРОЗАПАДНА И ЮГОЗАПАДНА БЪЛГАРИЯ

Анализирани са статистическите параметри на разпределението на три приземни метеоелемента (температура, максимална температура и относителна влажност), на температурата на ниво 850 hPa и на два индекса на неустойчивост (*K* и *Hi*) поотделно при развитие на гръмотевични (***th***) и негръмотевични (***or***) купесто-дъждовни облаци над Северозападна (NW) и над Югозападна (SW) България. С помощта на дискриминантен анализ са определени праговите им стойности за разграничаване на ***th*** от ***or*** облаци над двата района. Анализът показва, че няма разлика в температурите при развитие на купесто-дъждовни облаци над NW и SW България.

*Boryana Markova, Boris Yanakiev, Rumjana Mitzeva.* THERMODAINAMYC CONDITIONS AT THE DEVELOPMENT OF THUNDERSTORMS OVER NORTHWESTERN AND SOUTHWESTERN BULGARIA

Statistical parameters of the distribution of three surface meteorological parameters (temperature, maximum temperature, and relative humidity), and temperature at 850 hPa, and two instability indices (K and Hi) are analyzed separately for ordinary (without lightning ***or***) and thunderstorms (***th)*** cumulonimbus, developed over the northwest (NW) and southwest (SW) Bulgaria. Using general

---

*За контакти:* Боряна Маркова, НИМХ, Цариградско шосе 66, София 1784,
GSM: +359 889517681, E-mail: boryana.markova@meteo.bg

discriminant analysis the threshold values of parameters, which are able to discriminate between ***th*** and ***or*** clouds are determined over the two regions. The results indicate that the cumulonimbus develop at similar temperature over NW and SW Bulgaria.

***Keywords:*** instability indices, thunderstorms, discriminant analyses
***PACS numbers:*** 92.60.Pw, 92.60.Qx

# 1. ВЪВЕДЕНИЕ

През последните години в Европа се наблюдава развитието на все по-мощни и опустошителни бури, свързани с интензивна гръмотевична дейност и валежи [1–4]. Известно е, че тяхното формиране и по-нататъшно развитие съществено зависят от условията в околната среда. При установяване на подходящи предиктори за развитието на гръмотевични облаци могат да се подобрят както прогнозата им, така и предупрежденията за опасни и особено опасни явления, свързани с тях.

Това е основателна причина да бъдат изследвани условията в околната среда при развитието на купесто-дъждовните облаци. В литературата много автори (например [5–9] и др.) пресмятат и разглеждат различни индекси на неустойчивост като основни характеристики на термодинамичните условия в околната среда. Обикновено се определят прагови стойности на анализираните характеристики (метеоелементи, индекси на неустойчивост), които разграничават условията за образуване на гръмотевичните от негръмотевичните облаци за даден район, като е установено, че намерените прагове са различни в зависимост от района  и сезона на разглеждане.

В България от 60-те години на миналия век в продължение на 20 години въз основа на синоптичен анализ и анализ на термодинамични характеристики на атмосферата се търсят критерии за прогноза на гръмотевична дейност и градушки [10–16]. През последните години са направени нови изследвания на условията в околната среда, при които се развиват купесто-дъждовни облаци. В [17] е предложен метод за прогноза на вероятността за развитие на гръмотевични облаци над България през летния период. Резултатите от детайлен анализ на условията, при които се развиват гръмотевичните облаци над Източна България, е представен в [18].

Целта на настоящата работа е да се намерят прагови стойности на различни метеоелементи и индекси на неустойчивост, които да послужат като индикатори за развитието на гръмотевични облаци над Западна България. Предвид разликата в релефа, климатичните особености и др. между Северозападна и Югозападна България една от задачите в работата е да се установи и дали има разлика в условията в околната среда при развитието на гръмотевични и съответно негръмотевични облаци в двата района на България.

## 2. ИЗПОЛЗВАНИ ДАННИ И МЕТОДОЛОГИЯ

В работата са изследвани условията в околната среда, при които се развиват купесто-дъждовни облаци с и без мълнии поотделно над Северозападна (NW) и Югозападна (SW) България. Случаите на гръмотевичните облаци (*thunderstorm*) са означени с ***th***, а на дъждовните негръмотевични облаци (обикновени) (*ordinary*) – са означени с ***or***.

Използваните данни са за 732 дни и са разгледани 1197 случая на купесто-дъждовни облаци, развили се в следобедните часове в месеците между април и септември през периода 2006–2009 г. Случаите на негръмотевични облаци ***or*** са 187 в Северозападна и 300 в Югозападна България. Броят случаи на развили се гръмотевични облаци ***th*** е съответно 282 в NW и 428 в SW България. За да се установи при какви условия в околната среда се развиват конвективни облаци, е необходимо да се познава профилът на температурата и влажността в атмосферата. По подобен начин, както е проведен анализът в [19] за разграничаване на условията при развитие на купесто-дъждовни облаци, развили се над Североизточна и Югоизточна България, и тук са използвани апроксимирани сондажи от GFS модел [20] за 1200 UTC. От тях са взети данни за температура и влажност на въздуха на различни атмосферни нива. Приземните метеорологични данни (налягане, относителна влажност на въздуха, температура и максимална температура на въздуха) са от измервания в 6 синоптични станции на Националния институт по метеорология и хидрология (НИМХ), разположени на територията на Западна България. Три от синоптичните станции са разположени в Северозападна България – Видин, Монтана и Враца, и три – в Югозападна България – София, Кюстендил и Сандански.

Известно е, че развитието на купесто-дъждовна облачност зависи от температурата при земята и от температурите на различни нива, от неустойчивостта и влажността на атмосферата. Във връзка с това в работата са изследвани приземната температура $T$, относителната влажност на въздуха $RH$, максималната приземна температура $T_{\max}$ и температурата на ниво 850 hPa $T_{850}$. Пресметнати са два термодинамични индекса: $K$ [21] и Humidity Index $Hi$ [22], които се използват за характеризиране неустойчивостта на атмосферата, вероятност за гръмотевична дейност и значителни валежи. Начинът им на пресмятане е показан в табл. 1.

Направен е $t$-тест с ниво на значимост $\alpha = 0.05$, с помощта на който е проверено дали има статистически значима разлика между средните стойности на анализираните характеристики на околната среда в Северозападна и Югозападна България при развитието както на гръмотевични, така и на негръмотевични облаци.

За всеки един район са пресметнати статистически параметри на разпределението на съответните величини (средна стойност, медиана, долен и горен квартил), като резултатите са представени и в таблици, и са илюстрирани с квартилна диаграма.

**Таблица 1.** Формули за пресмятане на *K* и *Hi*, формули за пресмятане на *POD* и *FAR*: $T_{850}$, $T_{700}$ и $T_{500}$ са температури [°C] на ниво 850 hPa, 700 hPa и 500 hPa съответно; $T_{d850}$ , $T_{d700}$ и $T_{d500}$ са точки на оросяване [°C] на ниво съответно 850 hPa, 700 hPa и 500 hPa; *x* – брой на коректно класифицираните гръмотевични случаи, *y* – брой на некоректно класифицираните гръмотевични случаи, *w* – брой на некоректно класифицираните негръмотевични случаи

| Параметър | Знак и мерна единица | Уравнение |
|---|---|---|
| *K*-Index | *K* [deg] | $K = (T_{850}–T_{500}) + T_{d850} – (T_{700}–T_{d700})$ |
| Humidity Index | *Hi* | $Hi = (T–T_d)_{850} + (T–T_d)_{700} + (T–T_d)_{500}$ |
| Probability of detection | *POD* | $POD = \frac{x}{x+y}$ |
| False alarm ratio | *FAR* | $FAR = \frac{w}{x+w}$ |

С помощта на обикновен дискриминантен анализ [23] са определени прагови стойности на изследваните характеристики на околната среда, над/под които има голяма вероятност за развитие на гръмотевични облаци над Северозападна и Югозападна България. В работата е направена оценка на успеваемостта на получените прагови стойности за разграничаване на гръмотевични от негръмотевични облаци, като са пресметнати *POD* (probability of detection) и *FAR* (false alarm ratio) [24]. Начинът им на пресмятане е показан в табл. 1.

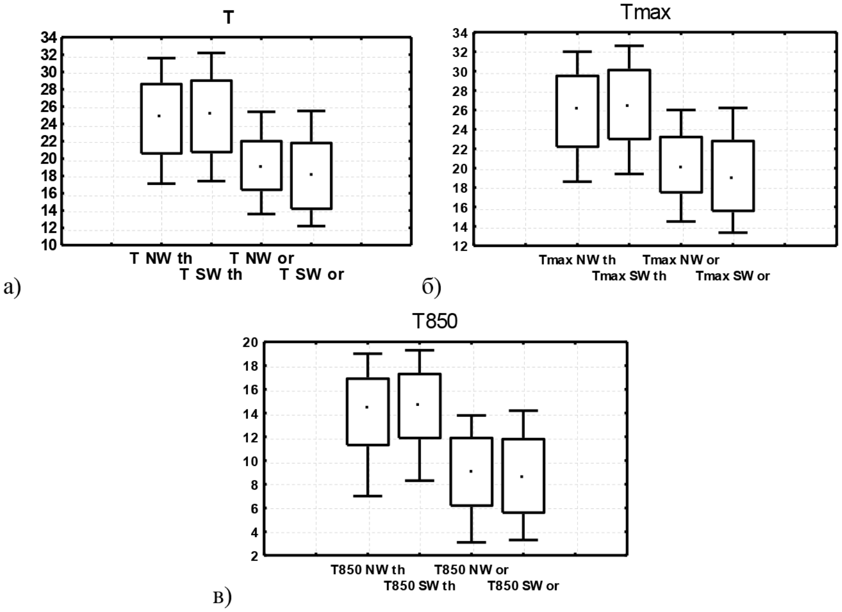
## 3. РЕЗУЛТАТИ

От информацията, представена в табл. 2, се вижда, че разликата в средната стойност на температурата при земята при развитието на гръмотевични облаци ***th*** над NW и SW България са много малки. Средната стойност на *T* в NW България е 24.5 °C, а в SW България е 24.9 °C. Такава малка разлика, под 1 °C, се наблюдава и при останалите температури, които са изследвани. Този извод е валиден и за съответните стойности на медианата, долния и горния квартил. Аналогичен резултат се наблюдава и при параметрите на разпределение за случаите на негръмотевични облаци ***or,*** развили се над двата района.

Представените резултати показват, че средните стойности на температурата при земята и във височина при развитие на гръмотевични облаци ***th*** са забележимо по-високи от съответните температури при развитие на негръмотевични валежни облаци ***or*** както в Северозападна, така и в Югозападна България. Така например, средната стойност на *T* при гръмотевичните облаци и в двата района е над 24 °C, а средната стойност за съответните температури при развитие на негръмотевични облаци е с 5–6 °C по-ниска (табл. 2).

**Таблица 2**. Параметри на разпределение на $T$, $T_{\max}$ и $T_{850}$ при развитие на негръмотевични облаци ***or*** и гръмотевични облаци ***th*** над Северзападна (NW) и Югозападна (SW) България

| Параметър | Район | Средна стойност | Медиана | Долен квартил | Горен квартил |
|---|---|---|---|---|---|
| $T$ *th* | NW | 24.5 | 24.0 | 0.6 | 28.6 |
| | SW | 24.9 | 25.0 | 20.7 | 29.0 |
| $T$ *or* | NW | 19.1 | 19.0 | 16.4 | 22.0 |
| | SW | 18.4 | 18.0 | 14.2 | 21.8 |
| $T_{\max}$ *th* | NW | 25.6 | 26.0 | 22.2 | 29.5 |
| | SW | 26.3 | 26.4 | 23.0 | 30.1 |
| $T_{\max}$ *or* | NW | 20.1 | 20.0 | 17.5 | 23.2 |
| | SW | 19.3 | 18.9 | 15.6 | 22.8 |
| $T_{850}$ *th* | NW | 13.7 | 14.3 | 11.3 | 16.9 |
| | SW | 14.4 | 14.7 | 11.9 | 1.3 |
| $T_{850}$ *or* | NW | 8.9 | 9.0 | 6.2 | 11.9 |
| | SW | 8.8 | 8.5 | 5.6 | 11.8 |



а) б) в)

**Фиг. 1**. Квартилна диаграма за: а) приземна температура $T$; б) максимална приземна температура $T_{\max}$; в) температура на височина 850 hPa $T_{850}$ при развитие на негръмотевични облаци ***or*** и гръмотевични облаци ***th*** над Северозападна (NW) и Югозападна (SW) България. Показани са 10-ти, 25-ти, 50-ти (точка), 75-ти и 90-ти процентил

Освен това както медианите, така и стойностите на долните и горните квартили на приземната температурата, на максималната температура и на температурата на ниво 850 hPa са забележимо по-високи при развитие на гръмотевични облаци, отколкото съответните стойности при развитие на негръмотевични облаци. Това добре се вижда и от фиг. 1. От направения анализ може да се каже, че гръмотевични облаци ***th*** над NW и над SW България се развиват при приблизително еднакви температури при земята и във височина. Същото е валидно и за средните стойности на температурите, при които се развиват негръмотевични облаци над двата района, но те са забележимо по-ниски от температурите, при които се развиват гръмотевичните облаци.

Резултатите от *t*-тест показват, че разликата между средните стойности на $T$, $T_{max}$ и $T_{850}$, при които се развиват купесто-дъждовни облаци над NW и SW, не е статистически значима – с ниво на значимост $\alpha = 0.05$ (вж. табл. 3), както при развитието на гръмотевични ***th***, така и при негръмотевични ***or*** облаци. От показаните в табл. 3 *p*-стойности се вижда, че с най-голяма убеденост не може да отхвърлим нулевата хипотеза за еднаквост на средните стойности над NW и SW за $T_{850}$ ***or*** и с най-малка убеденост – за $T_{850}$ ***th***.

**Таблица 3**. *p*-стойности при *t*-тест за $T$, $T_{max}$, $T_{850}$, $K$ и $Hi$ при развитие на негръмотевични облаци ***or*** и гръмотевични облаци ***th*** над Северзападна и Югозападна България

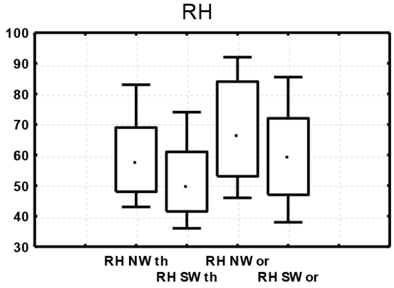
| Параметър | *p*-стойност | Параметър | *p*-стойност |
|---|---|---|---|
| $T$ ***or*** | 0.13 | $RH$ ***or*** | 0.71E-06 |
| $T$ ***th*** | 0.36 | $RH$ ***th*** | 0.63E-10 |
| $T_{max}$ ***or*** | 0.11 | $K$ ***or*** | 0.24 |
| $T_{max}$ ***th*** | 0.08 | $K$ ***th*** | 0.60 |
| $T_{850}$ ***or*** | 0.82 | $Hi$ ***or*** | 0.06 |
| $T_{850}$ ***th*** | 0.05 | $Hi$ ***th*** | 0.17 |

Относителната влажност на въздуха е друга характеристика на атмосферата, която е от съществено значение при развитието на гръмотевичните и негръмотевичните облаци.

От резултатите, представени в табл. 4 и фиг. 2, се вижда, че има разлика в средните стойности и съответните квартили не само при сравняване на относителната влажност *RH* при развитие на гръмотевични и негръмотевични облаци поотделно, а и при сравняване на съответните стойности при развитие на ***th*** и ***or***, развили се над NW, спрямо тези над SW. Оказва се, че в NW България при развитието и на гръмотевични облаци ***th*** и на негръмотевични облаци ***or*** е необходима с 7–8 % по-висока относителна влажност при земята в сравнение с условията за развитие на съответния вид купесто-дъждовни облаци в SW България. Изключително ниските *p*-стойности (вж. табл. 3) от проведения *t*-тест показват, че тази разлика е статистически значима (при ниво на значимост $\alpha = 0.05$) както при развитието на гръмотевични ***th***, така и при развитието на негръмотевични ***or*** облаци.

**Таблица 4.** Параметри на разпределение на относителната влажност *RH* при развитие на негръмотевични облаци ***or*** и гръмотевични облаци ***th*** над Северзападна (NW) и Югозападна (SW) България

| Параметър | Район | Средна стойност | Медиана | Долен квартил | Горен квартил |
|---|---|---|---|---|---|
| *RH* ***th*** | NW | 59.8 | 57.0 | 48.0 | 69.0 |
| | SW | 52.1 | 49.0 | 41.5 | 61.0 |
| *RH* ***or*** | NW | 68.3 | 66.0 | 53.0 | 84.0 |
| | SW | 60.3 | 59.0 | 47.0 | 72.0 |



**Фиг. 2**. Квартилна диаграма за относителната влажност *RH* при развитие на негръмотевични облаци ***or*** и гръмотевични облаци ***th*** над Северозападна (NW) и Югозападна (SW) България. Показани са 10-ти, 25-ти, 50-ти (точка), 75-ти и 90-ти процентил

От табл. 4 и фиг. 2 също се вижда, че средната стойност на *RH* при развитие на гръмотевичните облаци е забележимо по-ниска от средната стойност на *RH* при развитие на негръмотевичните облаци и над двата района. По-ниски са и съответните медиана, горен и долен квартил. По-ниското *RH* при развитието на гръмотевичните облаци в сравнение с *RH* при развитие на обикновени (негръмотевични) облаци е индикация, че основата на гръмотевичните облаци (ниво на кондензация) е на по-голяма височина спрямо основата на обикновените дъждовни облаци. Полученият резултат съответства на резултати, получени от други автори [25, 26], според които като показател за развитие на гръмотевични облаци може да се използва височината на основата на облаците.

След като не се наблюдава статистически значима разлика в средните стойности на температурите, отчетени при земята, ($T$ и $T_{max}$) и $T_{850}$ при развитието на гръмотевични и негръмотевични облаци над NW и над SW, възниква въпросът: съществува ли значима разлика в праговите стойности (определени за съответните температури) за разграничаване на гръмотевични от негръмотевични облаци, развили се над NW и SW България? Също така, като се наблюдава статистически значима разлика в средните стойности на относителната влажност над NW и SW България при развитието на гръмотевични и негръмотевични облаци, запазва ли се тя и в праговите стойности, получени за *RH*?

С помощта на обикновен дискриминантен анализ [23] са определени прагови стойности за $T$, $T_{max}$, $T_{850}$ и $RH$, които да разделят гръмотевичните от негръмотевичните облаци, развили се над Северозападна и Югозападна България.

**Таблица 5**. Прагови стойности на метеоелементи при земята и на $T_{850}$ за вида на облаците – гръмотевични ***th*** или негръмотевични ***or*** облаци, и успеваемостта *POD* и *FAR* на съответния праг за разграничаване на гръмотевичните от негръмотевичните облаци

| Параметър | NW | | | SW | | |
|---|---|---|---|---|---|---|
| | Праг | *POD* | *FAR* | Праг | *POD* | *FAR* |
| $T$ | > 21.8 | 0.66 | 0.21 | > 21.6 | 0.72 | 0.21 |
| $T_{max}$ | > 22.9 | 0.71 | 0.20 | > 22.9 | 0.76 | 0.19 |
| $T_{850}$ | > 11.3 | 0.74 | 0.22 | > 11.6 | 0.76 | 0.20 |
| $RH$ | < 64.1 | 0.69 | 0.29 | < 56.3 | 0.69 | 0.32 |

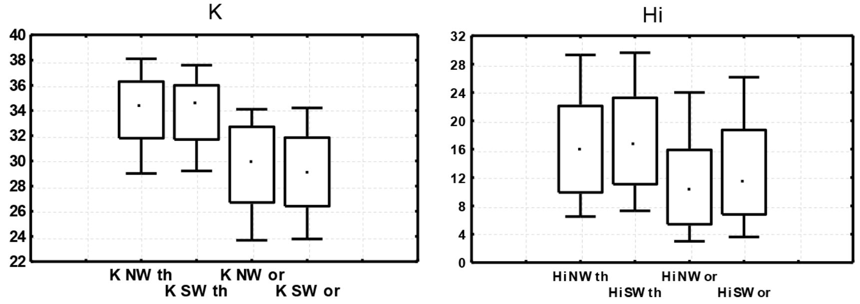
Получените прагови стойности и оценката на тяхната успеваемост посредством *POD* и *FAR* са представени в табл. 5. От таблицата се вижда, че определените прагове на температурите, над които се развиват гръмотевични облаци в NW и SW България, са много близки по стойност – най-голямата разлика е 0.3 °C. По отношение на относителната влажност *RH* обаче има съществена разлика между двата прага – за Северозападна България той е 64.1 %, а за Югозападна – 56.3%. От получените оценки *POD* и *FAR* се вижда, че като сравнително добър индикатор за гръмотевичните облаци могат да послужат $T_{max}$ и $T_{850}$, които са в състояние да класифицират коректно съответно 71% (*POD* = 0.71) и 74% (*POD* = 0.74) от ***th*** случаите, развили се над NW България, и 76% (*POD* = 0.76) над SW България. Използването на праговите стойности на тези две температури водят и до сравнително малък процент фалшива тревога, т.е. грешно класифицираните негръмотевични като гръмотевични облаци са ~20%.

При развитие на гръмотевични и негръмотевични валежни облаци над Северозападна и Югозападна България са анализирани също *K*-Index (*K*) и Humidity Index (*Hi*). Начинът на пресмятането им е даден в табл. 1.

**Таблица 6.** Параметри на разпределение на *K*-index и Humidity index *Hi* при развитие на негръмотевични облаци ***or*** и гръмотевични облаци ***th*** над Северозападна България (NW) и над Югозападна България (SW)

| Параметър | Район | Средна стойност | Медиана | Долен квартил | Горен квартил |
|---|---|---|---|---|---|
| *K* ***th*** | NW | 33.8 | 34.4 | 31.8 | 36.3 |
| | SW | 33.7 | 34.5 | 31.7 | 36.0 |
| *K* ***or*** | NW | 29.3 | 29.8 | 26.7 | 32.7 |
| | SW | 28.8 | 29.0 | 26.4 | 31.8 |
| *Hi* ***th*** | NW | 16.9 | 15.9 | 9.9 | 22.1 |
| | SW | 17.8 | 16.6 | 11.0 | 23.2 |
| *Hi* ***or*** | NW | 11.8 | 10.3 | 5.4 | 15.9 |
| | SW | 13.3 | 11.2 | 6.8 | 18.7 |

Резултатите в табл. 6 показват, че средните стойности на *K*-индекс при развитие на гръмотевични облаци ***th*** са забележимо по-високи (с около 5 °C) от съответните стойности на индекса при развитие на негръмотевични валежни облаци ***or*** както в Северозападна, така и в Югозападна България. Илюстрация на получения резултат се вижда и на фиг. 3.



а б

**Фиг. 3.** Квартилна диаграма за индекси на неустойчивост а) *K* ; б) *Hi* при развитие на негръмотевични облаци ***or*** и гръмотевични облаци ***th*** над Северозападна (NW) и Югозападна (SW) България. Показани са 10-ти, 25-ти, 50-ти (точка), 75-ти и 90-ти процентил

Получените *p*-стойности (вж. табл. 3) също показват, че няма статистически значима разлика между средните стойности за двата анализирани индекса (*K* и *Hi*) при развитие както на негръмотевични ***or,*** така и на гръмотевични ***th*** облаци над NW и SW България.

В табл. 7 са представени праговите стойности на анализираните два индекса за разграничаване на гръмотевичните ***th*** от негръмотевичните ***or*** облаци, развили се над Северозападна и Югозападна България, и успеваемостта *POD* и *FAR* на всеки един праг. Тези прагови стойности са получени с помощта на обикновен дискриминантен анализ [23].

Резултатите показват, че определените прагови стойности за *K* за развитие на гръмотевични облаци в NW и SW са доста близки: $K > 31.6$ °C за NW и $K > 31.3$ °C за SW България. Получените стойности на *POD* и *FAR* показват, че с помощта на тези прагове процентът на коректно класифицираните гръмотевични облаци е над 76 %, а процентът на фалшиво определени негръмотевични дъждовни облаци като гръмотевични (вж. стойността на *FAR*) е под 23 %. От табл. 7 се вижда, че намерените прагови стойност на *Hi* за развитие на гръмотевични облаци е по-ниска за Северозападна България ($Hi > 14.4$) спрямо намерения праг за Югозападна ($Hi > 15.6$). Използването на този праг обаче води до сравнително малък процент коректно класифицирани гръмотевични облаци – 57 % за NW и 53 % за SW България.

**Таблица 7.** Прагови стойности на *K* и Humidity index *Hi* за вида на облаците – гръмотевични ***th*** или негръмотевични ***or***, и успеваемостта *POD* и *FAR*на прага за класифициране

| Параметър | NW | | | SW | | |
|---|---|---|---|---|---|---|
| | Праг | *POD* | *FAR* | Праг | *POD* | *FAR* |
| *K* | > 31.6 | 0.76 | 0.23 | > 31.3 | 0.79 | 0.20 |
| *Hi* | > 14.4 | 0.57 | 0.26 | > 15.6 | 0.53 | 0.30 |

## 4. ЗАКЛЮЧЕНИЕ

Целта на работата е да се намерят прагови стойности на различни метеоелементи и индекси на неустойчивост, които да послужат като индикатори за развитието на гръмотевични облаци над Северозападна и Югозападна България. Причината за разглеждане на двата района поотделно е, че двете области попадат в различни климатични райони – умерено-континентален (Северозападна България) и южно-български (Югозападна България).

За целта са сравнени различни статистически параметри (средна стойност, медиана, долен и горен квартил) на температурата *T*, максималната температура $T_{max}$, относителната влажност *RH* при земята и температурата на 850 hPa $T_{850}$, както и на два от известните в литературата индекси на неустойчивост, *K* и *Hi*, при развитие на купесто-дъждовни облаци над Северозападна и Югозападна България. Чрез статистически анализ (*t*-тест с ниво на значимост $\alpha = 0.05$) е направена оценка на това, дали разликите между средните стойности на гореспоменатите параметри за NW и SW са статистически значими при развитие на гръмотевични ***th*** и негръмотевични ***or*** облаци поотделно.

Определени са прагови стойности за класифициране на гръмотевичните облаци поотделно за NW и SW. Оценката за успеваемостта на тези прагове е направена чрез изчисляване на *POD* (probability of detection) и *FAR* (false alarm ratio).

Получените резултати показват, че въпреки различните климатични особености на NW и SW България купесто-дъждовните облаци през топлото полугодие се развиват над двата района при близки приземни температури и температура на ниво 850 hPa. Сравнително близки са и условията, определени чрез *K* и Humidity index *Hi*, при развитие както на гръмотевични, така и на негръмотевични облаци над Северозападна и над Югозападна България. Макар и приземните температури при развитие на купесто-дъждовни облаци над NW и SW България да са близки, по всяка вероятност основата на тези облаци е на по-голяма височина над Югозападна България в сравнение на съответните, развили се над Северозападна България, поради установените статистически значими по-ниски стойности на относителната влажност в SW в сравнение с *RH* в NW България.

По отношение на разликата в развитието на гръмотевични и негръмотевични облаци и над двата района, за развитието на гръмотевичните облаци са необходими по-високи температури, по-ниска относителна влажност и по-неустойчива атмосфера. От определените прагови стойности на отделните термодинамични характеристики за купесто-дъждовни облаци, развили се в следобедните часове, най-добро разграничаване между гръмотевични и негръмотевични облаци, развили се над Северозападна и над Югозападна България, се получава при използване на намерените прагови стойности за $T_{max}$, $T_{850}$ и *K*-индекс. Получените прагове разграничават над 71 % гръмотевичните от негръмотевичните облаци над двата района, а фалшивата тревога е около 20 %. Резултатите показват, че определените прагови стойности за относителната влажност при земята *RH* и *Hi*-индекс самостоятелно не са подходящи за използване като индикатор за развитие на гръмотевични облаци.

Поради липсата на статистически значима разлика в средните стойности на температурите и двата индекса на неустойчивост при развитие на купесто-дъждовните облаци над Северозападна и Югозападна България е оправдано намирането на прагови стойности на тези параметри за раграничаване на гръмотевичните от негръмотевични облаци, валидни за цяла Западна България. В бъдеще е необходимо да се потърси класификационна функция, комбинация от метеоелементи и различните индекси на неустойчивост, която би могла по-добре да разграничава гръмотевични от неграмотевични облаци спрямо определените прагове на всеки един от разглежданите параметри.

## REFERENCES

[1] Bocheva, L., T. Marinova. *Journal of International Scientific Publications: Ecology and Safety*, 2016, **10**, 136-144.

[2] Bocheva, L. 15th *International Multidisciplinary Scientific GeoConference* SGEM 2015, www.sgem.org, SGEM2015 Conference Proceedings, June 18–24, 2015, **Book 4**, 889-896.

[3] Enno, S.E., A. Briede, D.Valiukas. *Theor. Appl. Climatol.*, 2012, DOI 10.1007/s00704-012-0666-2.

[4] Bielec-Bakowska, Z. *Environ. Socio.-econ. Stud.*, 2014, **2,** 3, 33-46. DOI: 10.1515/environ-2015-0041.

[5] Haklander, A. J., A. Van Delden. *Atmospheric Research*, 2003, **67–68**, 273-299.
[6] Huntrieser, H., H. Schiesser, W. Schmid, A. Waldvogl. *Weath. and Forecas.*, 1997, **12,** 108-125.
[7] Jacovides, C. P., T. Yonetani. *Weath. and Forecas.*, 1990, **5**, 559-569.
[8] Bunkers, M. J., J. R. Wetenkamp Jr., and J. J. Schild. *Weath. and Forecas.*, 2010, **25**, 799-814.
[9] Savvidou, K., A. Orphanou, D. Charalambous, P. Lingis, and S. Michaelides. *Adv. Geosci.*, 2010, **23**, 79-85.
[10] Stanchev, K. *Proceedings of the Institute of Hydrology and Meteorology*, 1965, **V**, 5-28 (in Bulgarian).
[11] Stanchev, K. *Proceedings of the Institute of Hydrology and Meteorology,* 1968, **XV** (in Bulgarian).
[12] Sirakova, M, N. Godev. *Hydrology and Meteorology*, 1970, **No 3**, 39-47 (in Bulgarian).
[13] Stanchev, K., P. Boev, R. Petrov. *Hydrology and Meteorology, XXII*, 1973, **4**, 53-62, (in Bulgarian).
[14] Boev, P., S. Stoyanov. *II International Conference on Hailstorms and Hail suppression*, Sofia, 1984, pp. 185–191 (in Russian).
[15] Boev, P., A. Marinov. *Hydrology and Meteorology*, *XXXIII*, 1984, **4**, 10-17 (in Bulgarian).
[16] Sirakov, D., M Sirakova. *Annual of Sofia University „St. Kliment Ohridski", Faculty of Physics,* 1986, **79**, 186-191 (in Bulgarian).
[17] Tsenova, B., A. Bogatchev. In: *Preprints: 16th Inter. Conf. on Clouds and Precipitation, July 30-August 03*, 2012, Leipzig, Germany.
[18] Markova, B., *PhD Thesis, Sofia University*, 2013, 35-77 (in Bulgarian).
[19] Markova, B., R. Mitzeva, T. Todorova. *Annual of Sofia University „St. Kliment Ohridski", Faculty of Physics,* 2015, **10**, 71-82 (in Bulgarian).
[20] http://ready.arl.noaa.gov/NOAA Air Resources Laboratory.
[21] George, J. In: *Weather Forecasting for Aeronautics,* Academic Press. 1960, 407-415.
[22] Litynska, Z, J. Parfiniewicz, H. Pinkowski. *WMO Bull*, 1976, **450**, 128-130.
[23] StatSoft, Inc., 2001: STATISTICA (data analysis software system), version 6.1, www.statsoft.com
[24] Donaldson, R., R. Dyer, M. Krauss. Preprints: *9th Conf. Severe Local Storms*, Norman, Oklahoma. Amer. Meteor. Soc., 1975, 321-326.
[25] Kaltenbock, R., G. Diendorfer, N. Dotzek. *Atmos. Res.*, 2009, **93**, 381-396.
[26] Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowsky. *Wea. Forecasting*, 2003, **18**, 1243-1261.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ

Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS

Volume 111

# THE EFFECT OF GLOBAL WARMING ON DYNAMICS, MICROPHYSICS AND ELECTRIFICATION OF CONVECTIVE CLOUD – NUMERICAL SIMULATIONS

BORYANA TSENOVA[1], RUMJANA MITZEVA[2], HRISTO IVANOV[2,3]

[1] *National Institute of Meteorology and Hydrology, BAS*
[2] *Faculty of Physics*
[3] *Chemical, Biological, Radiological and Nuclear Defense and Ecology, Musachevo, Bulgaria*

*Боряна Ценова, Румяна Мицева, Христо Иванов.* ВЛИЯНИЕ НА ГЛОБАЛНОТО ЗАТОПЛЯНЕ ВЪРХУ ДИНАМИКАТА, МИКРОФИЗИКАТА И НАЕЛЕКТРИЗИРАНЕТО НА КОНВЕКТИВЕН ОБЛАК – ЧИСЛЕНИ СИМУЛАЦИИ

Изследвано е влиянието на глобалното затопляне върху динамиката, микрофизиката и наелектризирането на мощен летен купесто-дъждовен облак с помощта на мезомащабния модел MesoNH. Симулиран е случай на облак при използване на реален профил на температурата в околната среда и повишена температура с 3 и 5 градуса. Резултатите показват, че затоплянето не оказва съществено влияние върху динамиката на симулираните облаци. Повишаването на температурата води до увеличаване на масата на хидрометеорите в течна фаза и повишаване на акумулирания течен валеж, както и до съществена разлика в знака и разпределението на заряда, носен от различните хидрометеори.

*Boryana Tsenova, Rumjana Mitzeva, Christo Ivanov*. THE EFFECT OF GLOBAL WARMING ON DYNAMICS, MICROPHYSICS AND ELECTRIFICATION OF CONVECTIVE CLOUDS – NUMERICAL SIMULATIONS

The effect of global warming on summer convective cloud dynamics, microphysics and electrification is studied using MesoNH model. A cloud case is simulated using: a real temperature profile and warmer respectively with 3 and 5 degrees. Results show that warming does not affect significantly

---

*For contact:* Boryana Tsenova, NIMH, 66 Tsarigradsko chausse Blvd, GSM: +359 886 687 485, E-mail: boryana.tsenova@meteo.bg

cloud dynamics. The increase in temperature leads to an increase of liquid hydrometeors mixing ratio and of accumulated liquid precipitation, as well to significant differences in charge sign carried by different hydrometeors.

***Keywords:*** global warming, convective clouds, numerical simulation, dynamics, microphysics, electrification

***PACS numbers:*** 92.60 Pw,92.60 Qx, 92.60 Ry

## 1. INTRODUCTION

The mean global surface temperature increases since the end of the 19th century [1]. The temperature increase was considerable between 1900 and 1940, as after 1970 until our days. However, this increase is not the same over the entire planet and in some places it is significantly higher than the defined global mean values. It is expected that the increase in temperature will have a significant impact on the development of convective clouds, thunderstorms and precipitation. Based on several studies such as [2–4] and others, it was stated [1] that the projected global warming will lead to more frequent and more severe extreme precipitation events. Trenberth et al. [5] studied the impact of various thermodynamical factors on precipitation, and concluded that the increased moisture content as a result of climate warming would have a significant impact on precipitation amount and intensity. Mccaul et al. [6] examined the sensitivity of supercell storms to environmental temperature and found that the updraft velocity and precipitation efficiency are higher at a colder environment, while the peak precipitation rate in a warmer environment is comparable to that in colder environment. Numerical simulations of Takemi et al. [7] with WRF model reveal the high dependence of the precipitation intensity from mesoscale convective systems on the temperature lapse rate. In the frame of their model simulations they found that with the increase of the lapse rate the mean precipitation intensity increases while the maximum precipitation intensity decreases. The author stresses on the need for diagnosis of stability in climate simulations and the need of further investigations on the effects of cloud microphysics on precipitation. Brandiyska et al. [8] using RAMS studied the effect of the expected changes of tropospheric temperature profile on the dynamical and microphysical characteristics of individual summertime convective storms and on the processes of precipitation development in these storms. Their results showed that projected warming lead to a decrease of precipitation from moderate cloud cases, while it leads to an increase of severe cloud cases precipitation. The main reason for the opposite direction of the impact of the projected tropospheric changes on different clouds lies in the ice phase evolution. The aim of the present study is to test the impact of expected changes in tropospheric temperature on convective clouds electrification having in mind the relationships between cloud electric charge structure and thundercloud microphysics and dynamics. Numerical simulations are performed using MésoNH model.

## 2. MÉSONH MODEL

The MésoNH is a non-hydrostatic mesoscale model developped in Laboratoire d'Aérologie and Météo-France [9]. The model integrates a system of equations able to simulate ideal and real atmospheric flows ranging from large eddy turbulent motion to the synoptic scale. The mixed-phase microphysical scheme in MésoNH follows the approach of Lin et al. [10] that is a three-class ice parameterization coupled to a Kessler scheme [11] used for the warm processes. The scheme follows the evolution of the mixing ratios of six water species: *rv* (vapour), *rc* and *rr* (cloud and rain drops) and *ri*, *rs* and *rg* (pristine ice, snow and graupel). The concentration of the precipitating particles is parameterized according to Caniaux et al. [12]. The pristine ice category is initiated by two heterogeneous nucleation processes: formation of ice embryos in a supersaturated environment over ice (deposition) following Meyers et al. [13], and freezing of supercooled droplets. In the model, the secondary production of ice crystals or rime-splintering mechanism is following Hallett and Mossop [14]. The homogeneous nucleation of pristine ice starts at temperatures lower than −35 °C. Ice crystals grow by water vapour deposition. The snow phase is initiated by autoconversion of primary ice crystals and it grows by deposition of water vapour, by aggregation through small crystal collection and by the riming produced by impaction of cloud droplets and of raindrops. Graupel particles are produced by the heavy riming of snow or by rain freezing when supercooled raindrops come in contact with pristine ice crystals. According to the heat balance equation and the efficiency of their collecting capacity, graupel particles can grow in dry and in wet mode (when riming is very intense and the excess of non-freezable liquid water at the surface of the graupel is shed and forms raindrops). At temperatures above 0 °C, ice particles melt into cloud and rain drops. Cloud droplet autoconversion, accretion and rain evaporation follow the Kessler scheme [11].

## 3. PARAMETERIZATION OF CHARGE SEPARATION IN THUNDERSTORM

The analytical expressions of the charging rates highly relies on the microphysical scheme:

$$\frac{dq_{xy}}{dt} = \int_0^{+\infty}\int_0^{+\infty} \left( \frac{\pi}{4} \delta q(1 - E_{xy})(D_x - D_y)^2 (V_x - V_y(n_x(D_x) n_y(D_y) dD_x dD_y \right)$$

where $D_x$ and $D_y$ are the diameters for hydrometeors *x* and *y*, respectively. $|V_x - V_y|$ is the relative fall speed, $n_x$ and $n_y$ are the number concentrations of hydrometeors *x* and *y*, respectively, and $E_{xy}$ is the collection efficiency. The collection efficiency depends on the temperature and follows Kajikawa and Heymsfield [15] for ice-snow and snow-graupel collisions and Mansell et al. [16] for ice-graupel collisions.

The expression of the charge exchanged is

$$\delta q = B\, d^a V^b \delta Q(RAR, T),$$

where $B$, $a$, and $b$ are constants depending on the size of small ice particles, on the relative velocity $V$ of the interacting ice particles, and on the sign of charge transfer and are tabulated in Saunders et al. [17]; $\delta Q$ is the charge determined from the parameterization scheme for non-inductive charging proposed in Brooks et al [18] giving the relationships between the separated charge and cloud temperature $T$ and rime accretion rate $RAR = EW.V$, where $EW$ is the effective water content.

$$\delta Q = 6.74RAR - 1.36(-T) + 10.05 \quad \text{if } RAR > RAR_{cr},$$

$$\delta Q = 3.03 - 10.59RAR + 2.95RAR^2 \quad \text{if } 0 < RAR < RAR_{cr},$$

where $RAR_{cr} = -1.47 + 0.2(-T)$.

## 4. NUMERICAL SIMULATIONS AND RESULTS

For the aim of the study three cloud cases are simulated. For the case named as «original» or dT0 the sounding presented in Fig. 1a is used. The other two cloud cases are simulated using the modified according to Fig. 1b temperature profile. The modified soundings are named as dT3 and dT5 (the number 3 or 5 corresponding to an increase of surface temperature with 3 or 5 °C respectively). The sounding modification follows Santer et al. [19] findings that changes in temperature profile depend on latitude. According to Santer et al. [19] the increase in temperature is higher in upper troposphere. It is about 1.5 times higher than surface temperature increase and has a maximum at around 300 hPa. As there is not any consensus about changes in relative humidity profile [20], for our simulations it is assumed to be identical for cases dT3 and dT5 with dT0. By keeping relative humidity constant, it is clear that the specific humidity increases with the increase of temperature.

All simulations are performed over a domain of 25×20 km, 28 vertical levels, a horizontal step of 500 m with a perturbation of 1.5 °C and sizes of 10×10×2 km in the thermodynamic field. The simulations time is 40 min with a time step of 1s. For case dT0 the zero isotherm height is about 3.7 km, while for cases dT3 and dT5 about 4.0 and 4.5 km, respectively. The increase of the environmental temperature leads to an increase in altitude of the –40 °C isotherm, which means that the homogenous freezing of cloud water droplets is at higher altitude over the ground. For cloud case dT0 the height of –40 °C isotherm is at 9 km, while for dT3 — at 9.6 km, and for dT5 — at 10.5 km. Due to the presumption accepted here that relative humidity does not change with temperature changes, the lifted condensation level height is approximately the same for the three simulated clouds dT0, dT3 and dT5 (Table 1).

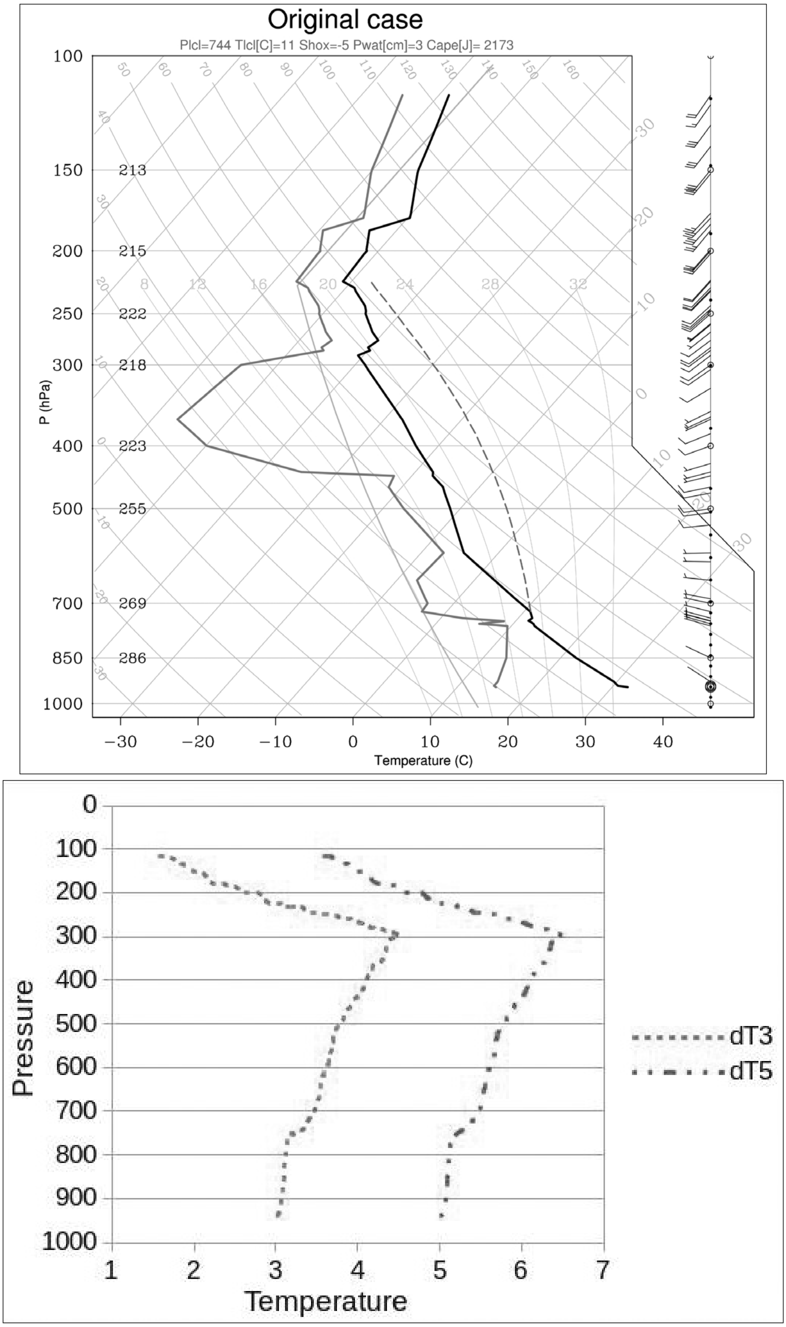


**Fig. 1.** a) Sounding used for case dT0 (original) simulation;
b) temperature changes at different pressure height for cloud cases dT3 and dT5

**Table 1.** Height and temperature of lifted condensation level,
CAPE and Lifted index for cloud cases dT0, dT3 and dT5

| | dT0 | dT3 | dT5 |
|---|---|---|---|
| Lifted condensation level height [m] | 2717 | 2723 | 2727 |
| Lifted condensation level temperature [°C] | 11 | 14 | 16 |
| CAPE [J/kg] | 2173 | 2481 | 3162 |
| Lifted index | –7 | –7.4 | –8.1 |

From Table 1 it is visible that cloud base for the three considered cases is at about 2.7 km. However the cloud base temperature for dT0 is 11 $^{o}$C, while for dT3 and dT5 it is respectively with 3 and 5 $^{o}$C warmer. The calculated CAPE and Lifted index based on soundings dT0, dT3 and dT5 show that for all considered cases the atmosphere is unstable and with temperature increase the instability increases significantly.

### 4.1. IMPACT OF TROPOSPHERE WARMING ON CONVECTIVE CLOUD DYNAMICS

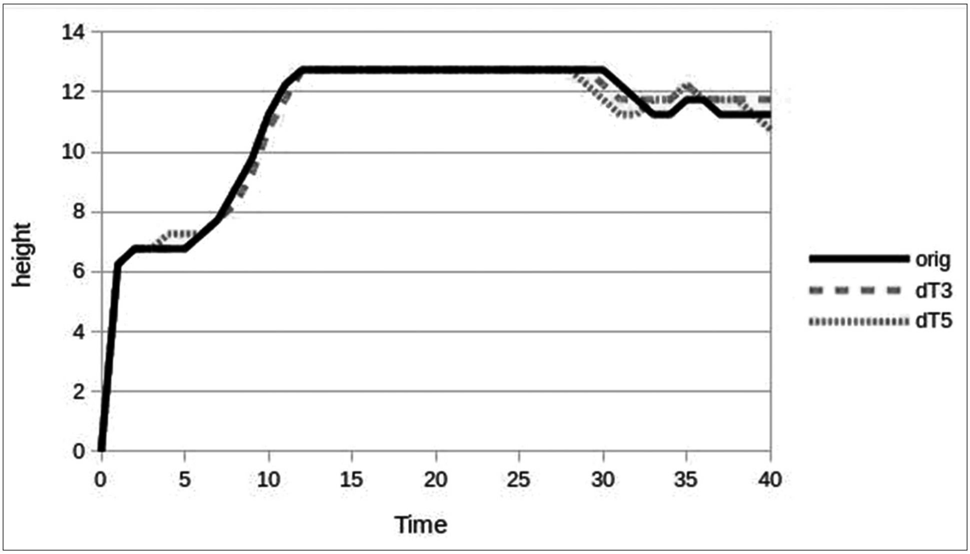


**Fig. 2.** Cloud cases dT0 (orig.), dT3 and dT5 top height AGL
as a function of time after cloud case formation

Analysis shows that for the three considered cases condensation starts 5 min after the beginning of simulations. From Fig. 3 it is visible that the three simulated clouds tops increase during the following 13 min and until the 28 min, clouds tops are at the same height (about 12–13 km AGL). Despite the approximately same altitudes of the simulated clouds tops, the lowest temperatures reached by the clouds are warmer with about 3 and 5 °C respectively for dT3 and dT5.

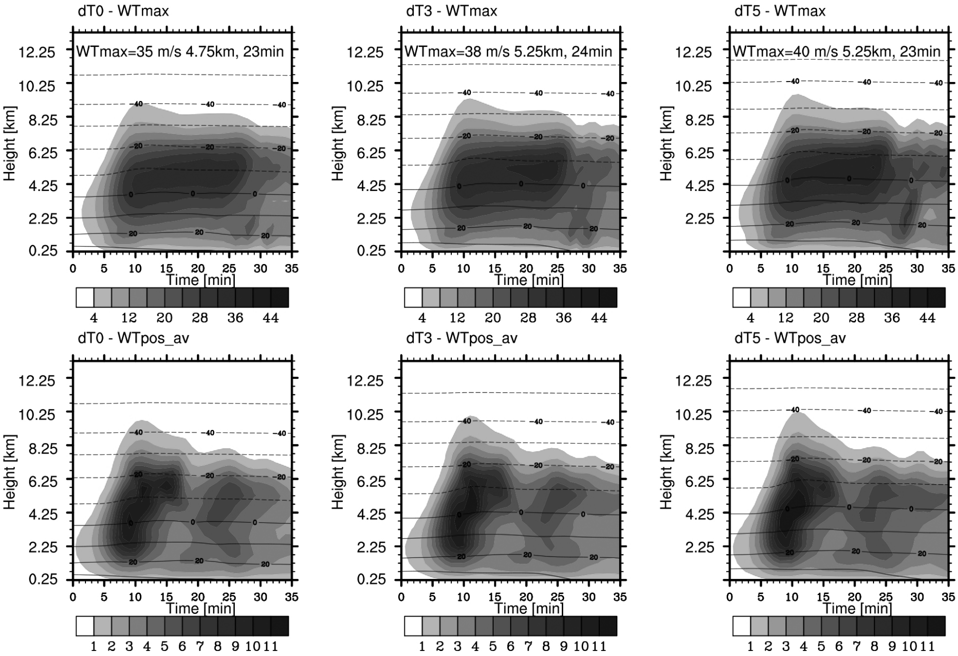


**Fig. 3.** Maximum updraft velocity, WTmax (top panel) and horizontally averaged updraft velocities, WTpos_av (bottom panel) as a function of time and height for the three cloud cases dT0, dT3 and dT5

During cloud development (Fig. 3, botom panel) it is visible that for the three considered cloud cases the horizontally averaged updraft velocities WTpos_av are strong and between 23 and 28 min a second maximum is visible (at temperatures between +1 and –17 °C for dT0, between +2 and –12 °C for dT3, and for dT5 – between +5 and –10 °C ). Results show that in warmer environements the maximum updraft velocity increases – it is 35 m/s for dT0, 38 m/s for dT3 and 40 m/s for dT5. These largest values of the maximum updraft velocity are reached between 23 and 24 min for the three considered clouds, at cloud temperature –7.5 °C for dT0, –7.4 °C for dT3 and –5.1 °C for dT5. Our results show that the increase in temperature profile has a weak effect on cloud dynamics leading to an increase of updraft velocity.

Fig.4 shows the horizontally integrated mixing ratios in kg/kg of cloud and rain droplets as a function of time and height for the three simulated cloud cases dT0, dT3 and dT5. The horizontally integrated values of cloud water droplets in dT0 are lowest. The maximum mixing ratio *RC*max = 5.5 g/kg is also the smallest in comparison with cases dT3 and dT5 (with *RC*max respectively 7.1 and 7.9 g/kg) and it is reached at lowest altitude in the cloud (5.25 km) and latest (at 18 min after cloud formation). In dT3 and dT5 the maximum of *RC* is reached at higher altitudes (5.75 km and 6.25 km respectively). However, cloud temperatures are similar: –11 °C for dT0 and dT5 and –10 °C for dT3. In the three simulated cases there is cloud water until about –40 °C, which corresponds to an altitude of about 9 km for dT0, about 9.6 km for dT3 and about 10.2 km for dT5.

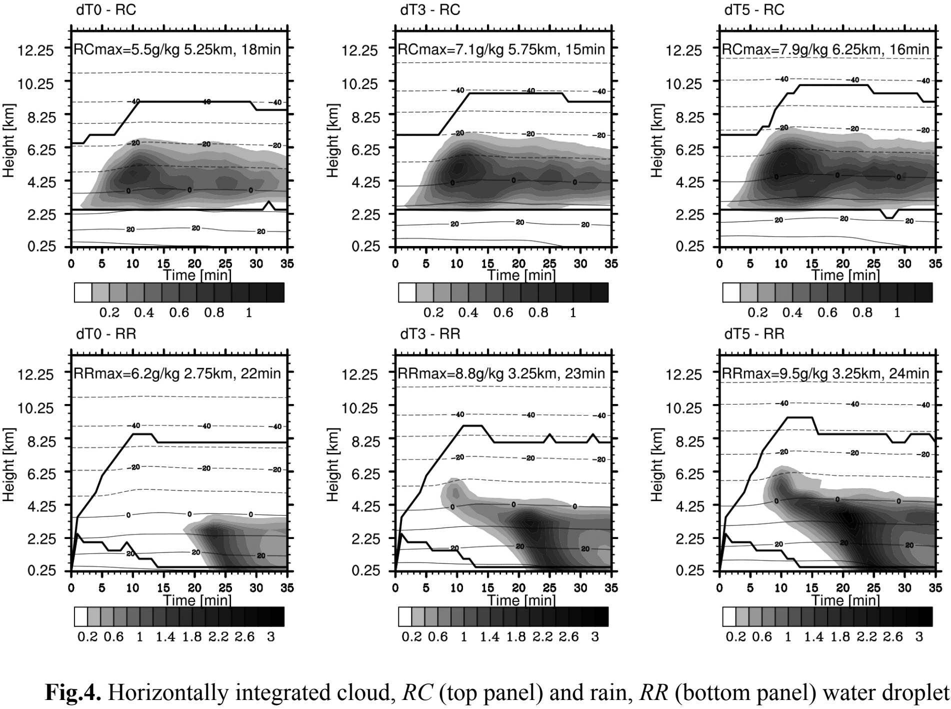


**Fig.4.** Horizontally integrated cloud, *RC* (top panel) and rain, *RR* (bottom panel) water droplets mixing ratios [kg/kg] as a function of time and height for the three considered cases dT0, dT3 and dT5 (the corresponding maximum values in g/kg are indicated above in each panel)

From Fig. 4 it is visible that the horizontally integrated rain water mixing ratio *RR* has high values at negative temperatures 10 min after cloud formation for cases dT3 and dT5, reaching 0.8 kg/kg and 1.2 kg/kg respectively, while for dT0 case it is below 0.2 kg/kg at negative cloud temperatures. For the three cases the maximum values of rain water mixing ratios are at positive cloud temperatures (at 6.5 $^{o}$C, for dT0 and dT3 and at 9 $^{o}$C for dT5). In case dT0 the value of maximum rain water mixing ratio is the lowest in comparison to dT3 and dT5 (respectively 6.2, 8.8 and 9.5 g/kg). Our results show that the increase of tropospheric temperature leads to an increase of liquid water (cloud and rain) mixing ratio in the simulated clouds. This result is not surprising, having in mind the higher cloud base temperature in clouds dT3 and dT5 in comparison to dT0 and the highest quantity of water vapour during clouds formation in warmer environment.

Fig. 5 shows the horizontally integrated mixing ratios in kg/kg of ice crystals, snow particles and graupel as a function of time and height for the three simulated cloud cases dT0, dT3 and dT5. It has to be stressed that at temperatures below –40 $^{o}$C the mixing ratio of the horizontally integrated ice crystals is considerable in dT0, while at these cloud temperatures it is lower in dT3 and negligible in dT5. The maximum value of ice crystals mixing ratio in the three clouds is similar (about 1.7 g/kg). It is reached approximately at the same time (about 14 min after cloud formation) and at similar altitudes above ground level (about 8.5 km). However, the cloud temperature in the different clouds at the level of the maximum ice crystals mixing ratio achievement differs considerably. For dT0, *RI*max is reached at –33.6 $^{o}$C; for dT3 — at –28.6 $^{o}$C, and for dT5 — at –30 $^{o}$C

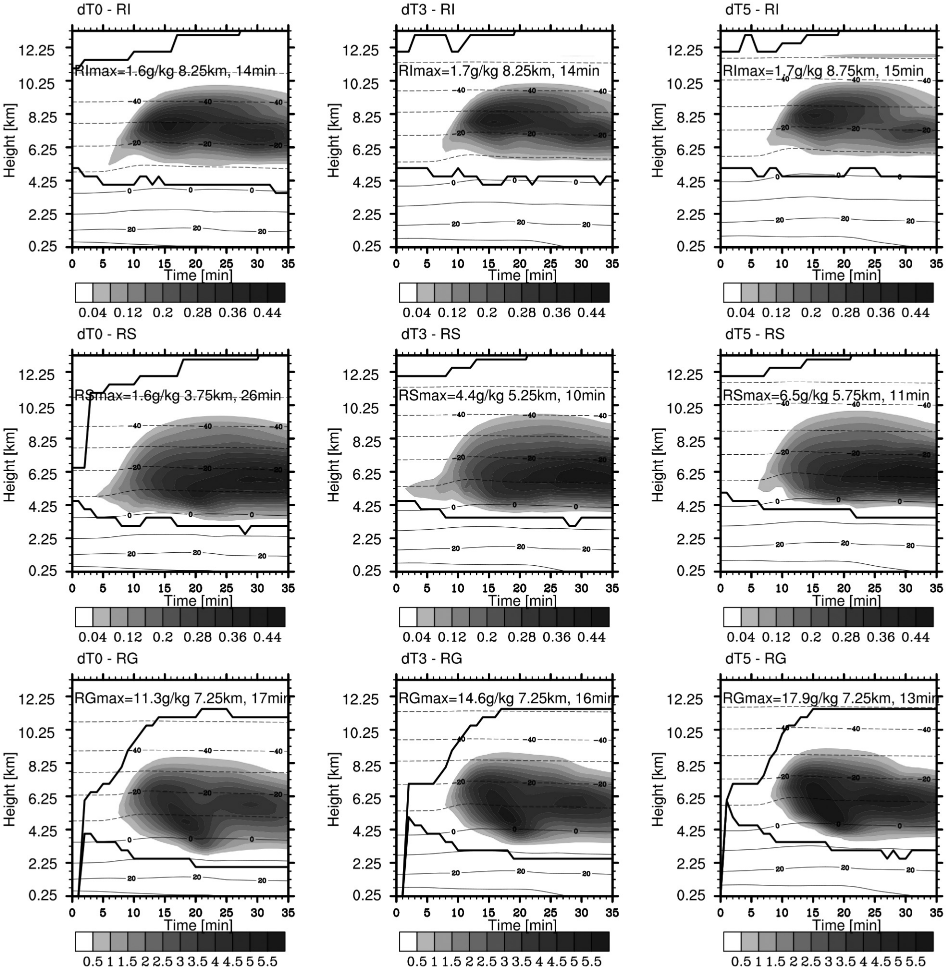


**Fig. 5.** Horizontally integrated ice crystals, *RI* (top panel), snow particles, *RS* (middle panel) and graupel, *RG* (bottom panel) mixing ratios [kg/kg] as a function of time and height for the three simulated cases dT0, dT3 and dT5 (the corresponding maximum values in g/kg are indicated above in each panel)

Similarly to ice crystals, horizontally integrated mixing ratios of snow particles are similar for the three cloud cases. However, in warmer environment the maximum of the horizontally integrated values of the mixing ratio are in a wider part of the cloud, for a longer time. The absolute maximum of snow particles mixing ratio *RS*max differs considerably in the three clouds. In the coldest environment it has the lowest value, which is achieved latest and at lower altitudes. The maximum values of graupel mixing ratios (11.3 g/kg, 14.6 g/kg and 17.9 g/kg respectively for dT0, dT3 and dT5) are achieved at the same altitude (7.25 km), which corresponds however to different cloud temperatures (–26 °C, –21 °C and –18 °C, respectively). Considering the horizontally integrated graupel mixing ratios for the three simulated clouds, one can see a second maximum, more pronounced for dT3 and dT5 cases, corresponding to the visible second maximum of horizontally averaged updraft velocities in Fig. 3 (bottom panel). Our results show that the increase in tropospheric temperature leads to an increase of snow and graupel mixing ratios.

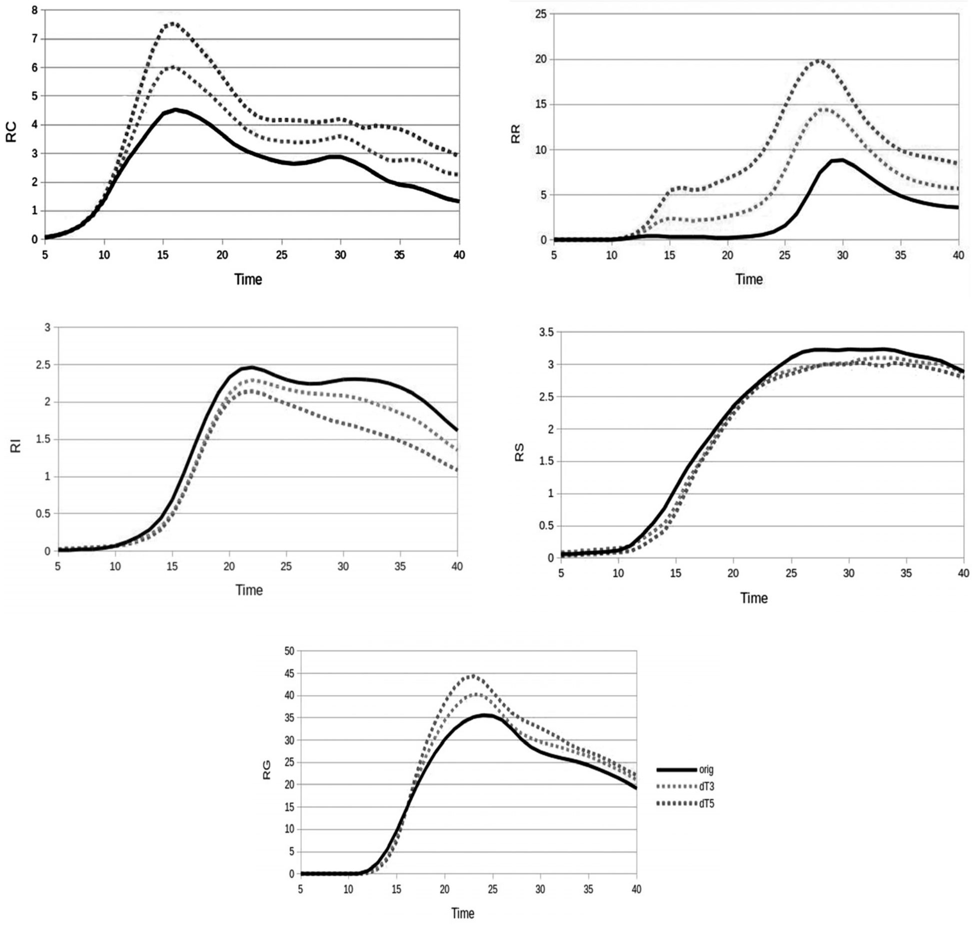


**Fig. 6.** Integrated cloud (*RC*) and rain (*RR*) water, ice crystals (*RI*), snow (*RS*) and graupel (*RG*) particles mixing ratios [kg/kg] as a function of time

The effect of tropospheric warming on integrated cloud water mixing ratio RC is well visible in Fig. 6. In the three simulated clouds, integrated cloud water is similar until the 12 min. Then, in warmer clouds dT3 and dT5 it increases considerably. Similar is the evolution of integrated rain water *RR* and graupel

*RG* mixing ratios that increase with temperature increase (Fig. 6). The effect of atmospheric warming on integrated ice crystals *RI* and snow *RS* mixing ratios is opposite, leading to their decr ease. The increase of integrated rain water *RR* and graupel *RG* mixing ratios leads to higher values of liquid precipitation from dT3 and dT5 in comparison to dT0 (Table 2).

**Table 2.** Maximum accumulated liquid precipitation from the three simulated cloud cases dT0, dT3 and dT5

| Simulated cloud cases | Maximum accumulated liquid precipitation [mm] |
|---|---|
| dT0 | 23.249 |
| dT3 | 32.596 |
| dT5 | 38.142 |

## 4.3. IMPACT OF TROPOSPHERE WARMING ON CONVECTIVE CLOUD ELECTRIFICATION

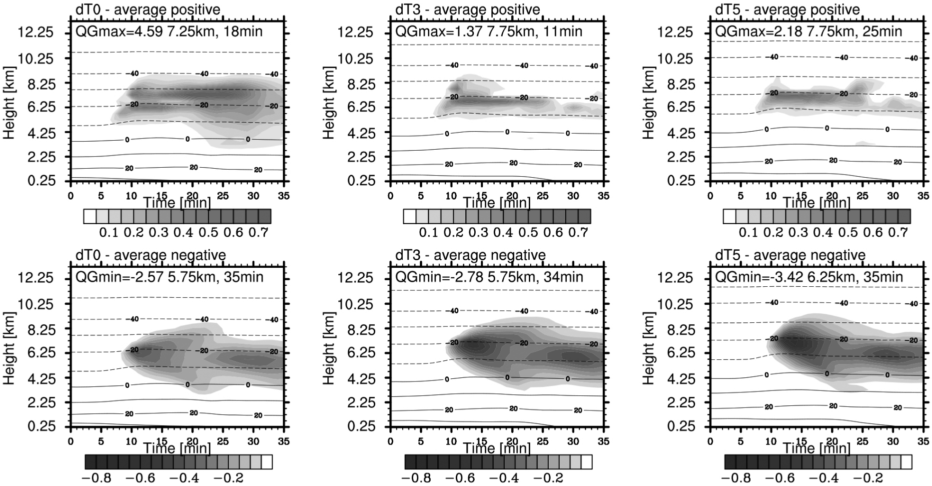


**Fig. 7.** Horizontally averaged positive (top panel) and negative (botom panel) graupel charge density [nC/m$^3$] QG as a function of time and height for the three cloud cases dT0, dT3 and dT5 (the corresponding maxima as well time and height of their achievement are indicated above in each plots)

Fig. 7 shows the evolution of the horizontally averaged positive and negative graupel charge density for the three simulated cloud cases. It is visible that the average positive graupel charge density for dT0 case at temperatures below –20 $^{o}$C after 15 min is larger than these in dT3 and dT5. Also Qgmax for this cloud case is considerably larger in comparison to the warmer clouds (4.59 nC/m$^3$ for dT0, while for dT3 and dT5 it is respectively 1.37 and 2.18 nC/m$^3$). The maxima positive graupel charge densities are achieved almost at the same height (about 7.5 km) in the three clouds, but the time of their achievements differs considerably (18 min, 11 min and 25 min respectively for dT0, dT3 and dT5). The tendency of the negative graupel charge density is opposite — there is a slight increase of maximum negative graupel charge density with the warming of the cloud. Our results show that the increase of tropospheric temperature leads to an increase of negative and to a decrease of the positive charge carried by graupel.

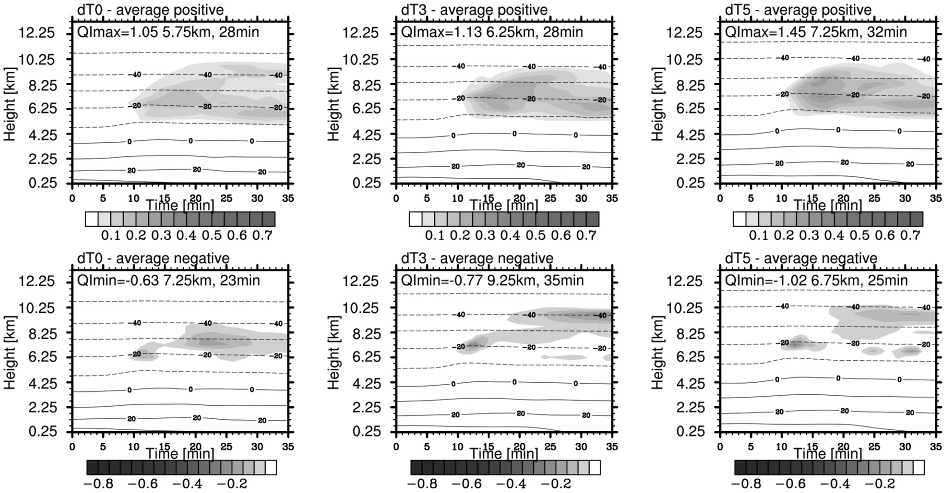


**Fig. 8.** Horizontally averaged positive (top panel) and negative (botom panel) ice crystals charge density [nC/m3] QI as a function of time and height for the three cloud cases dT0, dT3 and dT5 (the corresponding maxima as well time and height of their achievement are indicated above in each panel)

From Fig. 8. it is visible that the horizontally averaged positive ice crystals charge density in dT0 in the temperature interval <–10,–40> °C is smaller than these in clouds developed in warmer environment. However, in this case the average positive ice crystals charge density at temperatures below –40 °C is bigger. There is a slight increase with the warming of the maximum positive (respectively 1.05 nC/m$^3$, 1.13 nC/m$^3$ and 1.45 nC/m$^3$ for dT0, dT3 and dT5) as well of the maximum negative (–0.63 nC/m$^3$, –0.77 nC/m$^3$ and –1.02 nC/m$^3$ for dT0, dT3 and dT5 respectively) ice crystals charge density.

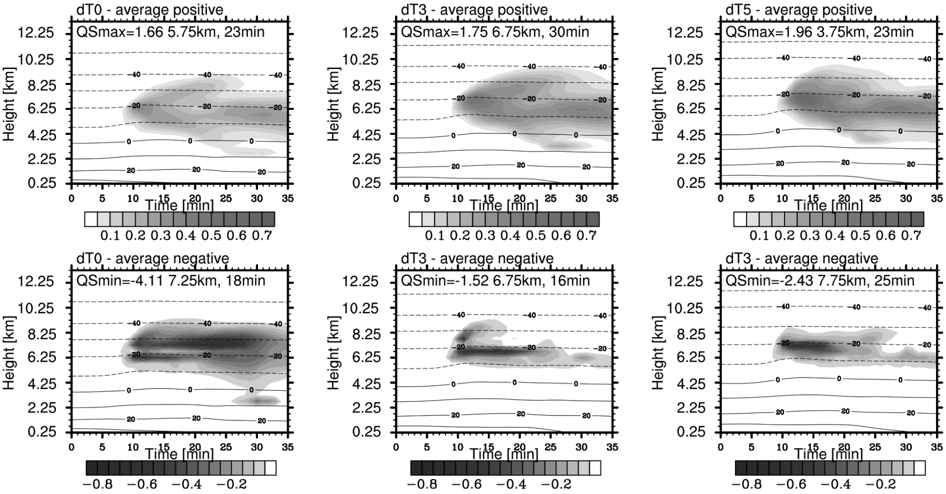


**Fig. 9.** Horizontally averaged positive (top panel) and negative (botom panel) snow particles charge density [nC/m3] QS as a function of time and height for the three cloud cases dT0, dT3 and dT5 (the corresponding maxima as well time and height of their achievement are indicated above in each panel)

Fig. 9 shows the evolution of the horizontally averaged positive and negative snow particles charge density for the three simulated cloud cases. It is visible that at cloud temperatures above –20 °C, the average positive snow charge density in dT0 is smaller in comparison to dT3 and dT5, while at temperature below –20 °C, the average negative snow charge density is bigger in dT0. There is a slight increase of the maximum positive snow charge density with warming. The maximum negative snow charge density differs considerably in the three simulated cases: –4.11 nC/m$^3$ for dT0, –1.52 nC/m$^3$ for dT3 and –2.43 nC/m$^3$ for dT5. Our results show that the tropospheric warming leads to an increase of the positive and to a decrease of the negative charge density due to snow particles.

The analysis of cloud and rain water charge densities in the simulated cloud cases (not shown here) shows that the increase of tropospheric temperature leads to an increase of negative cloud water charge density and to a decrease of the positive rain water charge density.

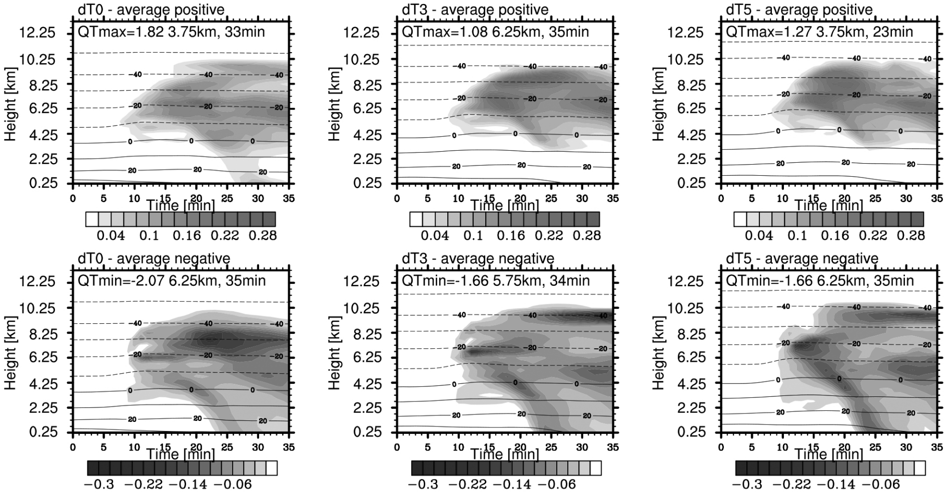


**Fig. 10.** Horizontally averaged positive (top panel) and negative (bottom panel) total charge density [nC/m$^3$] QT as a function of time and height for the three cloud cases dT0, dT3 and dT5 (the corresponding maxima as well time and height of their achievement are indicated above in each panel)

Fig. 10 shows the horizontally averaged positive and negative total charge density (which is the sum of the densities of the charges carried by all hydrometeors) as a function of time and height. From this figure one can see that warming leads to the increase of the average positive total charge densities, but to a decrease of its maximum values. The maximum and average negative total charge densities also decrease in clouds developed in warmer environment.

In Fig.11 the average total charge evolution for the three simulated cloud cases is presented. The main differences in cloud charge distribution between dT0 and the warmer dT3 and dT5 cloud cases are after 20 min. At temperatures between –25 °C and –39 °C the average total charge is negative in dT0, while in dT3 and dT5 it is positive.

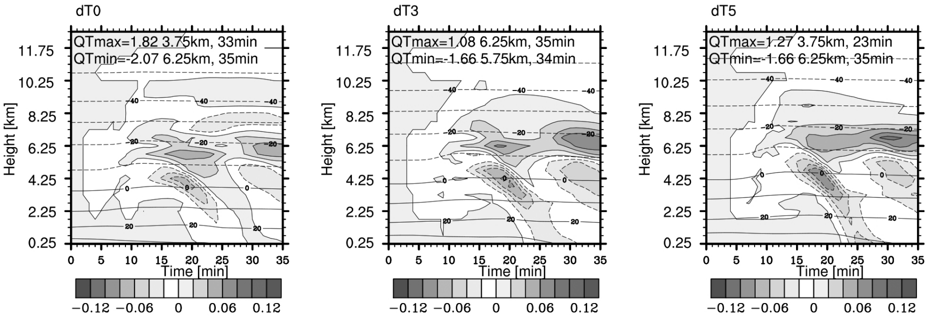


**Fig. 11**. Horizontally averaged total charge density [nC/m$^3$] *QT* as a function of time and height for the three cloud cases dT0, dT3 and dT5 (the corresponding maxima as well time and height of their achievement are indicated above in each panel)

This is a result of the higher negative charge density carried by snow particles in this cloud region in dT0 cloud in comparison to the warmer clouds. The average positive total charge density at positive temperatures in dT0 is due to the higher values of the positive rain water charge density in this cloud case. Our results show that tropospheric warming leads to an increase of the positive total charge density at negative cloud temperatures and to its decrease at positive cloud temperatures.

## 5. CONCLUSION

In the present study the impact of expected changes in temperature and humidity on termodynamic and electrical characteristics of convective clouds is studied using the non-hydrostatic model MesoNH. Results show that the increase of temperature and specific humidity affects slightly the cloud dynamics in direction of updraft velocity increase. However, for the simulated cloud cases (the original one dT0, and the modified dT3 and dT5), there is no visible effect on their power, because their tops reach tropopause level and above the atmosphere is stable. Cloud development in warmer and more humid environement (dT3 and dT5) leads to an increase of liquid (cloud and rain) water and graupel mixing ratios. The increase of cloud and rain water mixing ratios can be explained by the higher water vapor mixing ratio in warmer cloud temperatures as are in dT3 and dT5 cloud cases. The higher concentration of cloud droplets in warmer clouds (dT3 and dT5) in comparison to the original (dT0) leads to an increase of graupel mixing ratio, as more cloud droplets freeze on riming graupel surface. The higher precipitating hydrometeors mixing ratios explain the visible increase of accumulated precipitation from the warmer clouds in comparison to the original. Our results are similar to those obtained in Brandiyska et al.[8].

Related to the impact of the warming on cloud electrification, our results show that it leads to:

- an increase of negative and a decrease of positive graupel charge density;
- an increase of maximum positive and negative charge carried by ice crystals;
- an increase of positive and a decrease of negative charge carried by snow particles;
- an increase of negative cloud droplets charge density and to a decrease of positive rain droplets charge density.

These results show that the warming of the troposphere affects in different directions the charging of the hydrometeors. As a result, the mean positive total charge density increases at negative temperatures and decreases at positive cloud temperatures. The warming leads to a decrease of maximum positive and negative total charge densities.

However, it has to be stressed that the results presented here are obtained for the simulated here cloud cases and following the presumption that the relative humidity will remain constant at global warming. However, if the humidity over continents does not increase enough to keep the relative humidity constant, the impact of such termodynamical changes in the troposphere on convective clouds electrification will be less pronounced.

## REFERENCES

[1] IPCC, 2013: Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, 2013, Cambridge.

[2] Semenov, V. and L. Bengtsson. Report of the Intergovernmental Panel on Climate Change (IPCC), 2007.

[3] Kharin, V. and F. Zwiers. Report of the Intergovernmental Panel on Climate Change (IPCC), 2007

[4] Meehl, G. and others. Report of the Intergovernmental Panel on Climate Change (IPCC), 2007.

[5] Trenberth, K, A. Dai, R. Rasmussen, and D. Parsons. *B.Am.Meteorol.Soc.,* 2003, **84**, 1205-1217.

[6] McCaul, Jr., E.W., C. Cohen, C. Kirkpatric. *Mon. Weather Re.*, 2005, **133**, 3015–3037.

[7] Takemi, T. *Atmos. Res.*, 2010, **96**, 273-285.

[8] Brandiyska, A., R. Mitzeva, B. Tsenova and J. Latham. *Idojaras*, 2012, **116**, 4, 253-280.

[9] Meso-NH Scientific documentation version MASDEV4-10 (version August 20, 2014), http://mesonh.aero.obs-mip.fr /mesonh52/BooksAndGuides.

[10] Lin, Y.-L., R. Farley, H. Orville. *J. Clim. Appl. Meteorol.*, 1983, **22**, 1065-1092.

[11] Kessler, E. *Meteor. Monogr.*, 1969, Bolton, **10**, 32.

[12] Caniaux, G., J.-L. Redelsperger, J.-P. Lafore. *J. Atmos. Sci.,* 1994, **51**, 046-2074.

[13] Meyers, M.P., P. DeMott, W. Cotton. *J. Appl. Meteorol.*, 1992, **31**, 708-721.

[14] Hallett, J., S. Mossop. *Nature*, 1974, 249, 26-28.

[15] Kajikawa, M. and A. J. Heymsfield. *J. Atmos. Sci.*, 1989, **46** (20), 3108–3121.

[16] Mansell, E. R., D. R. MacGorman, C. L. Ziegler, J. M. Straka. *J. Geophys. Res.*, 2005, **110**, D12101.

[17] Saunders, C. P. R., W. D. Keith, R. P. Mitzeva. *J. Geophys. Res.*, 1991, **96**, 11007-11017.

[18] Brooks, M., C. P. R. Saunders, R. P. Mitzeva, S. L. Peck. *J. Atmos. Res*, 1997, **43**, 277-295.

[19] Santer, B.D, R. Sausen, T. Wigley, J. Boyle, A. Rao, K. Doutriaux, C. Hansen, J. Meehl, G. Roeckner, E. Ruedy, R. Schmidt and K. Taylor. *J.Geophys. Res.*, 2003, **108**, D1, 4001.

[20] Elliot, W.P. and J. Angell. *Geophys. Res. Lett.*, 1997, **24**, 41-44

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# REGIONAL MAGNITUDE RELATIONS FOR THE BALKAN PENINSULA

RENETA RAYKOVA[1], MARIA FILIPOVA[1,2], MILEN TSEKOV[1]

[1]*Department of Meteorology and Geophysics*
[2]*Department of Seismology, National Institute of Geophysics, Geodesy and Geography, BAS*

*Ренета Райкова, Мария Филипова, Милен Цеков.* РЕГИОНАЛНИ МАГНИТУДНИ ЗАВИСИМОСТИ ЗА БАЛКАНСКИЯ ПОЛУОСТРОВ

Виртуалната сеизмична мрежа на Софийския университет (ВСМСУ) е основана през 2015 г. в рамките на научен проект, финансиран от Софийския университет „Св. Климент Охридски“. ВСМСУ се състои от 16 сеизмични станции от различни национални и международни мрежи в Югоизточна Европа със свободен достъп до данните в реално време. Получихме стойности за няколко вида магнитуди ($M_l$, $M_d$, $M_b$, $M_S$), анализирайки 25 земетресения в района на Балканския полуостров чрез измерване на максималната амплитуда и нейния период за определени сеизмични вълнови пакети, както и на продължителността на земетръсните записи. Приложихме множествен линеен регресионен анализ за получаване на седем магнитудни зависимости, специфични за района на Балканския полуостров.

*Reneta Raykova, Maria Filipova, Milen Tsekov.* REGIONAL MAGNITUDE RELATIONS FOR THE BALKAN PENINSULA

The Virtual Seismic Network of Sofia University (VSNSU) was established in 2015 as a part of a scientific research project funded by Sofia University “St. Kliment Ohridski”. VSNSU consists of sixteen seismic stations belonging to several national and international networks in southeastern Europe with open access to near real-time data. We estimated values for several magnitude types ($M_l$, $M_d$, $M_b$, $M_S$) analyzing 25 earthquakes in the Balkan Peninsula region by measuring the maximum amplitude

---

*For contact:* Reneta Raykova, Department of Meteorology and Geophysics, Faculty of Physics, Sofia University „St. Kliment Ohridski”, 5 J. Bourchier Blvd, 1165 Sofia, Phone: +359 2 8161389, E-mail: rraykova@phys.uni-sofia.bg

and its period for certain seismic waves, as well as the earthquake's duration. We applied multiple linear regression method to obtain seven magnitude relations, specific for the Balkan Peninsula region.

***Keywords:*** earthquakes, seismograms, magnitude, Balkan Peninsula
***PACS numbers:*** 91.30.px

# 1. INTRODUCTION

The magnitude gives an estimate of the energy release during an earthquake. It is one of the most important earthquake parameters. Various magnitude types use different features of a seismogram record to characterize earthquake size. The magnitude estimates of a given seismic event obtained by the seismogram records from different stations may differ, because of the radiation pattern of respective seismic waves, differences in the Earth's structure between the hypocentral zone and the corresponding seismic stations, the quality of the records, etc. The magnitudes reported by different seismic agencies for specific earthquake may also vary significantly, even by more than one magnitude unit [1].

The Balkan Peninsula is one of the most seismically active regions in Europe. However, a homogenized catalog of modern earthquake activity in the region does not exist. Different national seismic agencies report magnitudes based on different network specific calibrations. Thus, the magnitude estimates for individual Balkan Peninsula's earthquakes reported by different national seismic networks, generally differ. There are two approaches to construct magnitude homogenized seismic event databases: (1) using general orthogonal regression (or other proper regression) method to calculate regression relations between the differing network magnitude scales, or (2) using original seismogram records from stations belonging to different seismic network and a reference magnitude scale to construct new magnitude relations. In this work we follow the second approach and constructed Balkan Peninsula specific magnitude relations based on seismic records from the Virtual Seismic Network of Sofia University (VSNSU) and magnitude estimates from IDC [2] as reference values in the calibration procedure. In a previous study [3] 490 individual measurements from seismic records of 11 earthquakes Balkan Peninsula earthquakes were used to calibrate 7 magnitude scales of 4 magnitude types (local magnitude $M_l$, duration magnitude $M_d$ – 2 scales, body wave magnitude $M_b$ – 2 scales, and surface wave magnitude $M_S$ – 2 scales). Here, following the same procedure as in [3], we extend the number of measurements to 758 (measurements from seismograms of 25 Balkan Peninsula earthquakes, presented in [3] and [4]) in order to revise the specific for the region magnitude relations obtained in [3].

The structure of the paper is as follows. In Section 2 we present the data and procedures to obtain magnitude values for a given magnitude type and we commented on some peculiarities and problems in magnitude estimations. In Section 3 we describe the application of multiple linear regression analysis method to calculate magnitude relations specific for the Balkan Peninsula region from the available data. We compare our calibration relations with these obtained in [3] and with magnitude relations calibrated for other regions of the globe. In Section 4 we summarize our findings and we outline some future work to improve magnitude relations for the earthquakes in the Balkan Peninsula region.

## 2. MAGNITUDE MEASUREMENTS

Virtual Seismic Network of Sofia University (VSNSU) was established in 2015 as a part of a scientific project funded by Sofia University. The 16 VSNSU seismic stations cover relatively uniformly the Balkan Peninsula region and they have open access to their seismic records in near real time. Seismic records from twenty five earthquakes were analyzed in the studied region in order to obtain several magnitude relations specific for the Balkan Peninsula. Fig. 1 shows the location of the VSNSU seismic stations and the epicenters of the analyzed earthquakes. The procedure for magnitude calculations, based on the respective seismogram measurements, was explained in detail in [3] and [4]. Every seismic record was processed and converted to the simulated record of an instrument appropriate to measure relevant amplitudes and their periods for a specific magnitude. The duration of the earthquakes was measured from the unfiltered seismic records. An example for amplitude, period and duration measurements for different magnitudes is shown in Fig. 2. Magnitude relations calibrated by several authors for certain regions of the globe were used in [3] and [4] to calculate magnitude values from the measured parameters. Normally these calibrations are site specific and they may not be valid in other regions including the Balkan Peninsula.

Estimates for seven magnitudes of four magnitude types ($M_d$, $M_l$, $M_S$, $M_b$) were obtained by the same procedure described in [3] and [4] using seismograms of 25 earthquakes, located in the Balkan Peninsula region with magnitude $\geq 4.0$ (as reported by [5]). Magnitudes were estimated for every seismic station that recorded a respective earthquake with sufficient quality. Next, we calculated the relevant event magnitudes (for each magnitude type) for each of 25 earthquakes by averaging station magnitude estimates. Thus we obtained the VSNSU magnitudes by "global" relations in order to compare them with regional magnitude estimates, defined in the next Section.

We identified some station magnitude estimates that are disproportionately higher or lower than the other station magnitude estimates for the same seismic event. This discrepancy may be explained by a combination of several factors. Firstly, this problem is related to the large uncertainty in the estimation of important earthquake parameters such as epicenter coordinates, hypocenter depth and origin time. The values of these parameters reported by different seismic agencies often vary significantly [1]. For example, for one of the selected events the variation in earthquake's depth reported by different seismic centers is comparable to the epicentral distance to the closest station. Secondly, the seismic waveforms are influenced also by seismic source peculiarities like the nature of rapture process. For example, if the rapture process is relatively slow or if it develops in several consecutive branches, the seismic wave amplitudes would be smaller and the record durations would be longer in respect to a "standard" earthquake. Thirdly, the quality of the magnitude estimates is affected by the seismic noise level, especially for seismic stations in immediate vicinity of coastal regions. Noisy records make difficult identification and measurement of important earthquake parameters such as amplitude, period and duration.

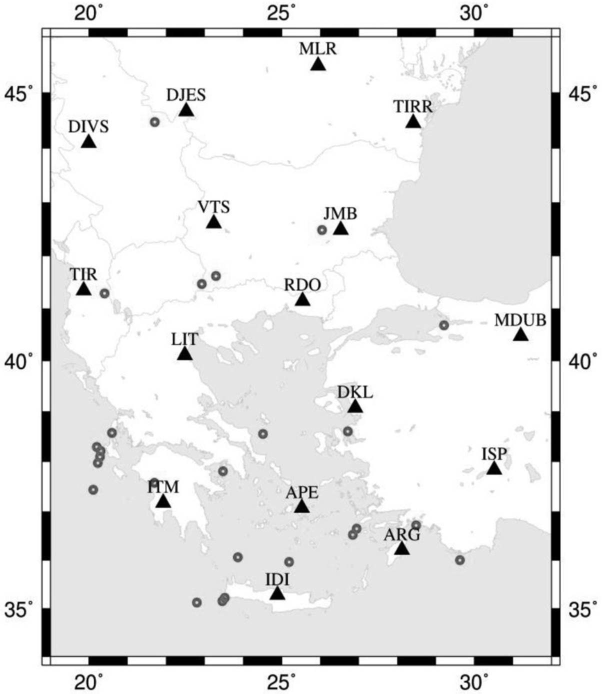


**Fig. 1.** Map of the Balkan Peninsula region with locations of the VSMSU seismic stations (black triangles) and the epicenters of the selected seismic events (red circles)

## 3. MAGNITUDE RELATIONS FOR THE BALKAN PENINSULA REGION

We used the multiple linear regression analysis method to obtain magnitude relations for each magnitude type that is specific to the Balkan Peninsula region. The multiple linear regression is a common statistical procedure used to find the relationship between a response variable and two or more explanatory variables by fitting a linear equation to observed data. It is applied to variety of statistical problems including for magnitude scale calibrations [6, 7]. We used the R free software environment [8] for statistical computations to implement the method.

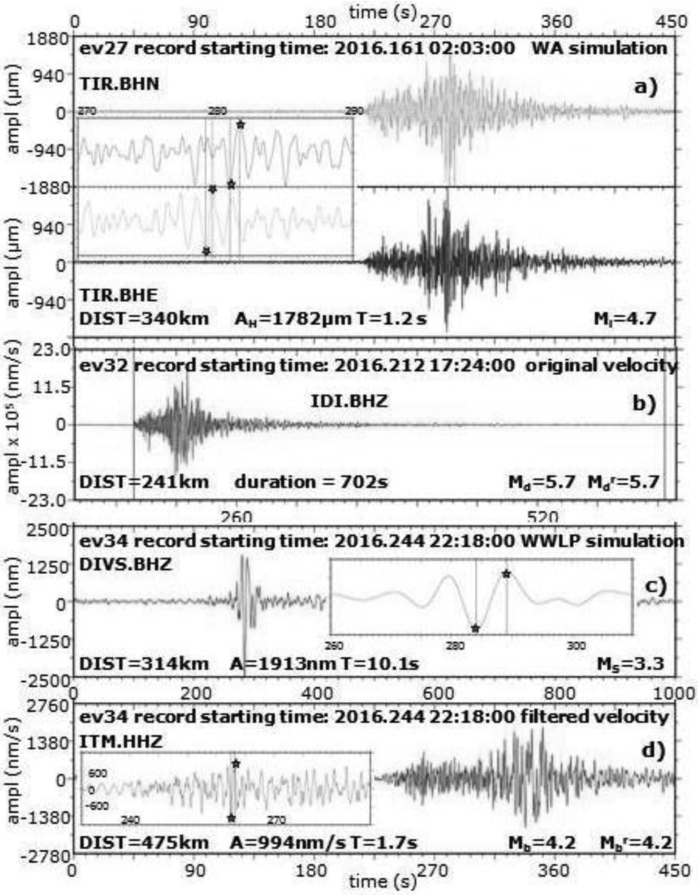


**Fig. 2.** Measurements to obtain magnitude estimates of different types: (a) for local magnitude $M_l$ at station TIR; (b) for duration magnitudes $M_d$ and $M_d^r$ at station IDI; (c) for surface wave magnitudes $M_S$ and $M_{S\_BB}$ at station DIVS; (d) for body wave magnitude $M_b$ at station ITM

We applied multiple linear regression to the set of magnitude measurements reported in [3] and [4]. Thus, we determined the specific coefficients $A_i$, $B_i$, $C_i$ and $F_i$ in the generalized relations for $M_d$, $M_l$, $M_S$, and $M_b$:

$$M_l = A_1 \times \log(A_H) + B_1 \times \log(R) + C_1 \times R + F_1$$
$$M_b = A_2 \times \log(V) + B_2 \times \log(D) + F_2$$
$$M_b^r = A_3 \times \log(V) + B_3 \times \log(R) + F_3$$

$$M_S = A_4 \times \log(A/T) + B_4 \times \log(\Delta) + F_4$$
$$M_{S\text{–}BB} = A_5 \times \log(V/2\pi) + B_5 \times \log(\Delta) + F_5$$
$$M_d = A_6 \times \log(\tau) + B_6 \times D + F_6$$
$$M_\mathrm{b}^\mathrm{r} = A_7 \times \log(\tau) + B_7 \times R + F_7.$$

Here, $A$ and $A_H$ denote the relevant wave amplitudes in [µm] or [nm], $T$ – period of the relevant amplitude in [s], $V$ – amplitude of the velocity in the relevant seismic record in [m m/s], $t$ – duration of the earthquake record in [s], $R$ – hypocentral distance in [km], $D$ – epicentral distance in [km], $\Delta$ – epicentral distance in [deg]. The sign “×” in the equations denotes multiplication. The difference between the two body wave magnitudes is that $M_\mathrm{b}$ uses epicentral distance while $M_\mathrm{b}^\mathrm{r}$ – hypocentral distance. The difference between the two surface wave magnitudes is that $M_\mathrm{S}$ uses WWWLP instrument simulation while $M_\mathrm{S\_BB}$ – the broad-band long-period filtered velocity records. The difference between the two duration magnitudes is that $M_\mathrm{d}$ uses epicentral distance while $M_\mathrm{d}^\mathrm{r}$ – hypocentral distance. See [3] and [4] for more details about magnitudes definition.

We used the corresponding transformations of the relevant seismogram parameters (see the generalized relations above) as explanatory variables in the regressions. We adopted corresponding magnitudes obtained by IDC [2] (provided by ISC [1]) as reference values. Body wave magnitudes were calibrated with respect to reported by IDC $M_\mathrm{b}$ magnitude estimates, surface wave magnitudes were calibrated with respect to reported by IDC $\mathrm{M_S}$ magnitude estimates, while local and duration magnitudes were calibrated with respect to IDC $M_\mathrm{l}$ magnitude estimates. Reference magnitude values for each selected earthquake are listed in Table 1.

**Table 1.** Reference magnitudes obtained by IDC [2]. Event labels follow [3] and [4]

| Event | $M_\mathrm{l}$ (IDC) | $M_\mathrm{S}$ (IDC) | $M_\mathrm{b}$ (IDC) | Event | $M_\mathrm{l}$ (IDC) | $M_\mathrm{S}$ (IDC) | $M_\mathrm{b}$ (IDC) |
|---|---|---|---|---|---|---|---|
| Ev01 | 4.0 | 3.4 | 4.2 | Ev18 | 4.3 | 3.7 | 4.2 |
| Ev02 | 4.2 | 3.6 | 4.5 | Ev20 | 3.7 | 3.3 | 4.0 |
| Ev03 | - | - | 3.7 | Ev21 | 3.7 | 3.3 | 3.6 |
| Ev04 | - | 3.8 | 4.2 | Ev25 | 3.5 | 3.5 | 3.9 |
| Ev05 | 4.0 | 3.2 | 3.8 | Ev27 | 3.7 | 3.5 | 4.3 |
| Ev06 | 4.4 | 4.7 | 4.8 | Ev28 | 3.4 | 3.2 | 3.8 |
| Ev07 | 4.0 | 2.9 | 3.6 | Ev29 | - | - | - |
| Ev08 | 3.6 | 3.1 | 3.9 | Ev32 | 4.2 | 4.3 | 4.8 |
| Ev09 | 3.8 | 3.3 | 4.1 | Ev33 | 3.6 | 3.0 | 4.0 |
| Ev11 | 4.3 | 3.9 | 4.4 | Ev34 | - | - | - |
| Ev14 | 3.9 | 3.4 | 3.8 | Ev36 | 3.4 | - | 3.9 |
| Ev15 | - | 4.8 | 4.8 | Ev37 | 4.1 | 3.6 | 4.5 |
| Ev17 | - | 3.8 | 4.0 | | | | |

The number of waveforms used in the regression for the corresponding magnitude relations is summarized in Table 2. The number of waveforms used in [3] is presented for comparison.

**Table 2.** Number of waveforms used for calibration of specific magnitude scales

| Magnitude scale | $M_l$ | $M_b$ | $M_b^r$ | $M_S$ | $M_{S\_BB}$ | $M_d$ | $M_d^r$ |
|---|---|---|---|---|---|---|---|
| Number of waveforms (this study) | 94 | 132 | 131 | 142 | 144 | 56 | 59 |
| Number of waveforms (in [3]) | 55 | 91 | 91 | 93 | 95 | 32 | 33 |

Applying the multiple linear regression method we obtained the following magnitude relations:

$$M_l = (0.14 \pm 0.14) \times \log(A_H) - (1.01 \pm 0.64) \times \log(R) + (0.0020 \pm 0.0009) \times R + (5.31 \pm 1.33)$$

$$M_b = (0.49 \pm 0.05) \times \log(V \log(1.08 \pm 0.12) \times \log(D) - (0.28 \pm 0.46)$$

$$M_b^r = (0.47 \pm 0.05) \times \log(V) + (1.10 \pm 0.13) \times \log(R) - (0.30 \pm 0.46)$$

$$M_S = (0.70 \pm 0.03) \times \log(A/T) + (0.55 \pm 0.11) \times \log(\Delta) + (1.74 \pm 0.11)$$

$$M_{S\_BB} = (0.71 \pm 0.03) \times \log(V/2\pi) + (0.53 \pm 0.11) \times \log(\Delta) + (1.72 \pm 0.11)$$

$$M_d = (0.37 \pm 0.16) \times \log(\tau) - (0.00032 \pm 0.00032) \times D + (3.05 \pm 0.41)$$

$$M_d^r = (0.36 \pm 0.17) \times \log(\tau) - (0.00026 \pm 0.00033) \times R + (3.04 \pm 0.43).$$

The comparison with relations obtained in [3] emphasizes several important inferences. Magnitude relations for the body wave magnitudes and the surface wave magnitudes differ less between the two studies than the local magnitude and the duration magnitudes. It is probably related to the larger numbers of measurements used in the calculation of the magnitude relations for the body wave magnitudes and the surface wave magnitudes, decreasing statistical fluctuations and leading to stability of the estimates. Uncertainties of the estimates for the body wave and surface wave magnitudes are smaller in this study where more data are used in respect to the data used in [3]. It is an indication that the magnitude relations may approach stability. Both relations for body wave magnitude, $M_b$ and $M_b^r$, have similar regression coefficients and uncertainties. Thus, it is justified in future studies based on larger data sets to use only one of the relations. Similar behavior is observed also for the surface wave magnitude scales, $M_S$ and $M_{S\_BB}$. Therefore only the $M_{S\_BB}$ may be used in future research since there is no necessity to simulate WWWLP records contrary to the $M$s estimates. Uncertainties of the regression coefficients for the duration magnitude relations also diminish in comparison with the previous work [3]. Duration magnitude practically does not depend on the distance, since the corresponding calibration coefficients in both $M_d$ and $M_d^r$ relations are close to zero. Both duration magnitude scales, $M_d$ and $M_d^r$, have similar regression coefficients and uncertainties; hence it is justified to use only one of them in future studies. The local magnitude depends on distance only by the logarithm term in the regression, while the calibration coefficient for the linear distance term is approximately zero. Regression coefficient $A_1$ in local magnitude relation is characterized by higher uncertainties than this in [3], while other coefficients have smaller uncertainties. More seismic waveforms measurements are required to reach stability for the local magnitude relation.

We recalculated magnitude estimates for all 758 sets of measurements from individual seismic waveforms by obtained magnitude calibration relations. Next, we estimated different magnitudes (for each magnitude type separately) for each analyzed earthquake as average of relevant station magnitudes. Figs. 3–6 present comparison between magnitude estimates for the analyzed 25 seismic events, calculated on the base of the magnitude relations obtained in this study, relations obtained in [3] and magnitude estimates by relations calibrated for other regions of the globe (for details see [3] and [4]). Relevant reference magnitude values, as given by IDC, are also presented in Figs. 3–6.

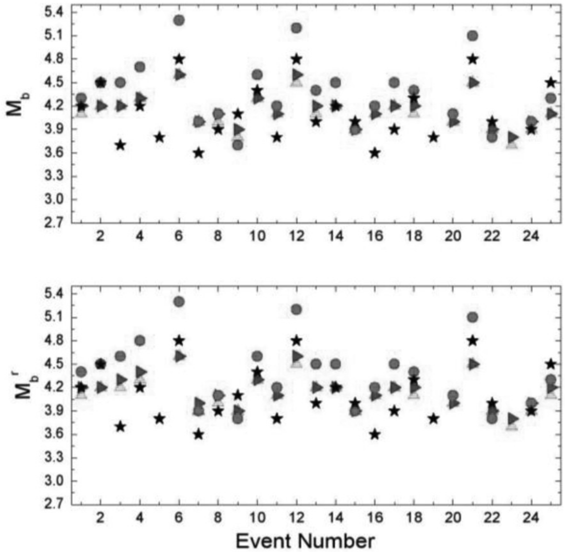


**Fig. 3**. Comparison between body wave magnitude estimates. Black stars denote reference magnitudes, green triangles – estimates by relations obtained in this study, violet triangles – magnitude estimates by relations obtained in [3], red dots represent estimates based on “global” relations, used for initial magnitude estimates in [3] and [4]

Differences between “local” (Balkan Peninsula specific) and “global” estimates for surface wave magnitudes are smaller than these for other magnitude types. It is an indication that the “global” relation for the surface wave magnitude may be considered relatively appropriate also for Balkan Peninsula’s earthquakes. For the majority of the considered earthquakes single event magnitudes based on the two different relations (in [3] and in this study) coincide completely, except for 3 $M_{S\_BB}$ estimates and 5 $M_S$ estimates which differ by 0.1 magnitude units.

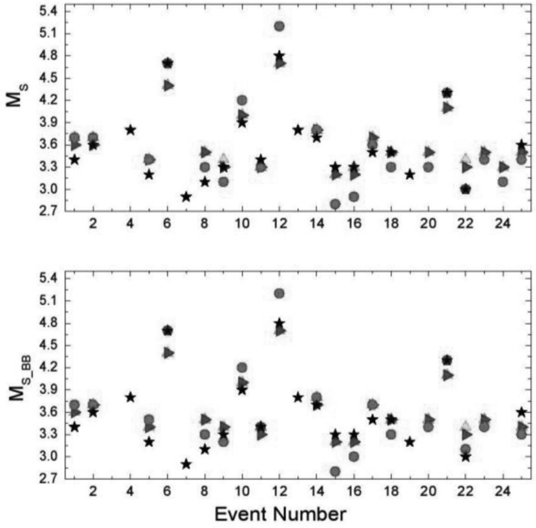


**Fig. 4**. Comparison between surface wave magnitude estimates. See Fig. 3 for used symbols

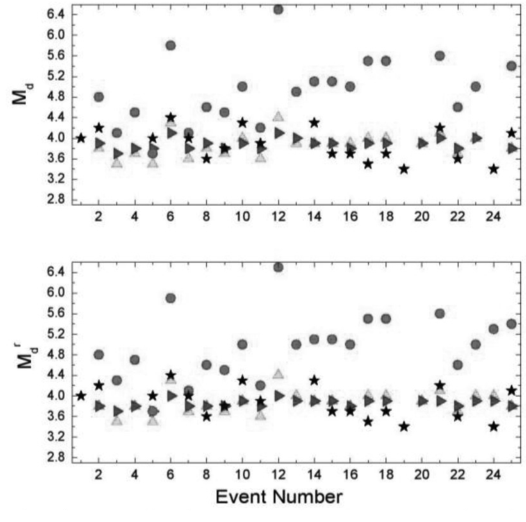


**Fig. 5**. Comparison between duration magnitude estimates. See Fig. 3 for used symbols

Body wave magnitude estimates in this study and in [3] also differ by no more than 0.1 magnitude units. Differences between “local” and “global” body wave magnitude estimates are relatively low for most of the selected earthquakes, although for a limited number of seismic events the difference is up to 0.6–0.7

magnitude units.

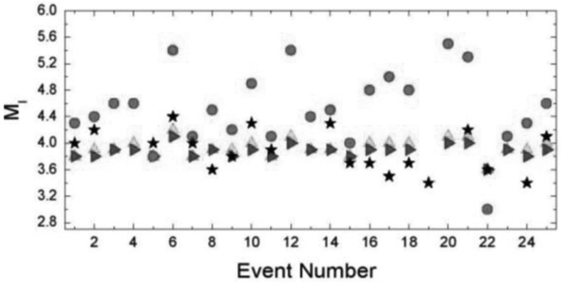


**Fig. 6**. Comparison between local magnitude estimates. See Fig. 3 for used symbols

Local magnitude estimates in this study and in [3] also differ by no more than 0.1 magnitude units. It is interesting that the difference between these "local" estimates and the "global" $M_l$ magnitude relations may be quite large (up to 1.4 magnitude units), indicating that "global" $M_l$ magnitude calibration is not appropriate for application to Balkan Peninsula earthquakes.

Relatively large are differences (up to 0.3 magnitude units) between the duration magnitude estimates in this study and in [3]. As for the local magnitude estimations, the "global" calibration relation for duration magnitude is not suitable for Balkan Peninsula earthquakes.

## 4. SUMMARY AND CONCLUSIONS

In this paper we presented new magnitude calibration relations for local magnitude, body wave magnitude (2 scales), surface wave magnitude (2 scales), and duration magnitude (2 scales) derived by VSNSU seismic record measurements. The obtained relations are representative for the earthquakes in the Balkan Peninsula region. Our results improve and precise the magnitude relations obtained in the previous work [3].

Obtained magnitude relations are still preliminary despite the improvements. Many more seismic records should be processed and used in the calibration procedure to obtain more precise and fully operational relations for the Balkan Peninsula region. We plan to process more VSNSU seismogram records including also earthquakes with $M < 4.0$, in order to decrease the uncertainties of the regression coefficients. The obtained Balkan Peninsula magnitude relations may improve the magnitude determination of earthquakes especially in regions near to political borders between the different countries, where magnitude estimates are usually based on seismic stations records from single national network with limited azimuth coverage.

**Acknowledgements.** This paper was published with the financial support of project 80-10-200/2017 funded by Scientific Research Fund of Sofia University. Authors thank Prof. Ivan-Asen Petkov Foundation granting semestrial scholarship to bachelor students in geophysics in Sofia University. M. Filipova received the scholarship for three semesters between summer 2016 and summer 2017.

## REFERENCES

[1] ISC International Seismic Center, www.isc.ac.uk, 2017.

[2] IDC International Data Center at CTBTO, www.ctbto.org, 2017.

[3] Raykova, R., M. Tsekov, V. Protopopova, M. Filipova, L. Dimitrova, L. Dimova, In: *Proceedings of the 3rd Bulgarian National Congress on Physical Sciences,* 2016, S0623, 1-10, Heron Press, Sofia.

[4] Filipova, M., R. Raykova. *J. of Physics and Technology,* 2017, **1**, 2 (in press).

[5] EMSC-CSEM, European Mediterranean Seismic Center, www.esmc-csem.org, 2017.

[6] Kaviris G., P. Papadimitrios, and K. Makropoulos. In: *Proceedings 11th International Congress, Athens*, 2007.

[7] Craiu, M., A. Craiu, C. Ionescu, M. Popa, and M. Radulian. *Romanian Reports in Physics*, 2012, **64(4)**, 109.

[8] Hothorn T., B. Everitt. A Handbook of Statistical Analyses Using R. Boca Raton, Florida, 2014.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# ДИСТАНЦИОННИ МЕТОДИ ЗА ГАМА-КАРТОГРАФИРАНЕ С ПОМОЩТА НА БЕЗПИЛОТНИ ЛЕТАТЕЛНИ СРЕДСТВА (БЛС)

ИВАН ИЛИЕВ[1,2], ПЛАМЕН ДАНКОВ[1]

*[1] Катедра „Радиофизика и електроника“*
*[2] „Тита Консулт“ ООД, София, България*

Иван Илиев, Пламен Данков. ДИСТАНЦИОННИ МЕТОДИ ЗА ГАМА-КАРТО-ГРАФИРАНЕ С ПОМОЩТА НА БЕЗПИЛОТНИ ЛЕТАТЕЛНИ СРЕДСТВА (БЛС)

Разгледана е технологията за дистанционно картографиране на терени с предполагаемо гама радиоактивно замърсяване с помощта на безпилотни летателни системи (БЛС). Дискутират се принципите на избор на гама-детектор, подходящ за вграждане в БЛС. Описани са подходите за създаване на летателен план при линеен и точков модел на облитане и техниките на картографиране на гама-радиоактивното замърсяване. Представени са данни от тестово измерване на изкуствено създадено радиоактивно замърсяване и са разгледани техниките за събиране на данни. Накрая е представен конкретен пример за обработване на получените резултати и създаване на карта на предполагаемото радиоактивно замърсяване.

*Ivan Iliev, Plamen Dankov.* DISTANCE METHODS FOR GAMMA-MAPPING WITH UNMANNED AERIAL VEHICLES (UAV)

The technology for remote mapping of terrains with suspected gamma radioactive contamination using unmanned aerial vehicles (UAV) has been explored. The principles of selection of gamma-detectors suitable for embedding in UAV have been discussed. Described are the approaches to creating a flight plan for a line and point model for data collection from the terrain and the techniques of gamma-radioactive contamination mapping. Test data from artificially generated radioactive contamination are presented and data collection techniques are described. Finally, a concrete example has been given for data processing and mapping of the supposed radioactive contamination.

---

*За контакти*: Иван И. Илиев, „Тита Консулт“ ООД, бул. Дж. Баучър 5А, 1164 София, тел.:+359 888 550 089, E-mail: i.iliev@thetaconsult.com

***Keywords:*** airborne remote sensing, radioactive pollution, gamma-background, scintillating detector, CsI, gamma-mapping, UAV

***PACS numbers:*** 89.60.-k, 98.70.Vc, 29.40.-n, 92.60.hx, 89.90.+n

## 1. ВЪВЕДЕНИЕ

За изследване на обширни територии от земната повърхност и близко прилежащата атмосфера – картографиране, земеделие, природни бедствия и катаклизми, аварии, дива природа, населени територии, автомобилен трафик, качество на водата и въздуха, локални замърсявания и пр., един от най-подходящите методи е наблюдението от въздуха (airborne remote sensing). Към тази широка категория от дейности можем да прибавим и дистанционното изследване на радиоактивни замърсявания от въздуха.

Ако знаем местоположението на дадено радиоактивно замърсяване на земята, можем да се насочим към него с подходящата измервателна апаратура и да го оценим качествено и количествено. За безопасността на персонала, който би извършил подобно измерване с ръчно преносима апаратура или натоварена в движещо се по земната повърхност превозно средство (кола, камион и др.), от съществено значение е информацията за типа на очакваното лъчение. Това може да бъде предвидено чрез познаване на радионуклидния състав на замърсяването. Друга полезна за персонала информация би била една, макар и груба, количествена оценка на активността и нейното разпределение по земната повърхност – дали цялата активност е концентрирана върху по-малка, или разпределена върху по-голяма площ. За да си набавим такава полезна предварителна информация за относително големи територии, най-удобно би било да ги облетим с летателен апарат (пилотируем или безпилотен), носещ на борда спектрометрични детектори за радиоактивност.

Гама-лъчението е дълбоко проникващо йонизиращо лъчение (радиация) от естествени или изкуствени източници и затова гама- спектрометрията е много мощен инструмент за мониторинг и оценка на радиацията в околната среда. Изследванията на гама-лъчението се извършват от въздухоплавателни средства, полеви превозни средства, пеша, в сондажи, на морското дъно и в лаборатории. Наземните и особено въздушните измервания на гама-лъчение обхващат големи участъци от земната повърхност. Много национални и регионални организации са съставили и публикували подробни гама-радиометрични карти. Такива стандартизирани карти на наземните концентрации на радиация и радиоактивни материали могат да бъдат сравнени и регионално обединени, показвайки общи регионални тенденции в радионуклидното разпространение и извършване на радиологичната оценка на околната среда. Този процес се поддържа изключително активно от International Atomic Energy Agency (IAEA) [1–4].

Гама-картографията от въздуха с пилотируеми летателни системи датира от десетки години. Най-активни са държавите и организациите, които изследват обширни и труднодостъпни територии (например в Австралия [5]). Тук има натрупан много опит, включително и в България [6–8].

Идеята за използване на безпилотни летателни системи (БЛС) за дистанционна гама-радиометрия и гама-картографиране датира от по-скоро време – от 10–11 години. В обзора [9] е дадена подробна картина на развитието на тази методика. Всъщност по-сериозно развитие идеята претърпява след ядрения инцидент във Fukushima Daiichi Nuclear Power Plant през март 2011 г., като след тази година започва да се гледа много по-сериозно и на дистанционните измервания при аварии с помощта на управляеми роботи (вж. например [10]), каквито са БЛС (или дроновете). При това БЛС могат да се прилагат както при радиоактивни инциденти (например [11]) с оглед на получаване на спешна информация без риск за живота на специалистите, работещи по случая, така и за регулярни измервания и картография на локални замърсители (например изследване на урaниевата аномалия в Чехия [12]). Гама-радиоактивните измервания с БЛС могат да се разделят условно на две категории: с безпилотни самолети и с квадрокоптери [9]. Първите се използват за бързи и относително по-груби измервания от височина 30–50 m, а вторите – за по-точни и локализирани измервания от височина типично до 10 m, включително и във вътрешността на сгради и помещения.

За да се изгради гама-карта с добра разделителна способност и в същото време методът ни да бъде икономически ефективен, е нужно да се открие балансът между необходимата чувствителност на детекторната система, от една страна, и скоростта и височината на полета, от друга. По-високата скорост и по-голямата височина позволяват по-бързо облитане на дадена територия, което спестява летателно време и по-бързо се получава резултатът от измерването. За да се постигне това, са необходими детектори с по-голяма чувствителност, което от своя страна увеличава техния обем/тегло и цена. Намирането на баланса между двете се постига чрез решаването на редица физични задачи.

С развитието на безпилотните летателни системи (БЛС) се увеличават и възможностите за картографиране на земната повърхност от въздуха. Към БЛС можем да причислим и малките спътници, които биха ни послужили за наблюдение на радиационната обстановка по орбитата им в близкия Космос (Spaceborn remote sensing) (вж. например [13,14]). БЛС позволяват изследването на територии, криещи различни опасности – както радиационни, така и от всякакъв друг характер, без физически да се застрашава наземният екипаж, управляващ летателното средство. Това тяхно предимство разкрива нови перспективи както за развитието на дозиметрията, така и за службите, реагиращи в аварийни ситуации.

В настоящата работа ние сме споделили опита си от последната година, свързан с приложение на БЛС за дистанционно гама- картографиране. Това включва: анализ на подходящите гама-детектори, анализ и дискусия на лабораторни и полеви експерименти с поставено (изкуствено създадено) радиоактивно замърсяване, съставяне на летателни планове за облитане над замърсяването, обработка на данните и принципите на изграждане на картата на обекта – гама-картографирането. Накрая са направени изводи и обсъдени насоки за бъдещата ни работа по дистанционно гама-картографиране с помощта на БЛС.

## 2. СЪВМЕСТЕН ПРОЕКТ ЗА ДИСТАНЦИОННА ГАМА-КАТРОГРАФИЯ С ПОМОЩТА НА БЛС МЕЖДУ ФИЗИЧЕСКИЯ ФАКУЛТЕТ И ФИРМА „ТИТА КОНСУЛТ“ ООД

В началото на 2017 г. екип от докторанти от Физическия факултет на Софийския университет „Св. Климент Охридски“, ръководени от доц. Пламен Данков от катедра „Радиофизика и електроника“, започна работа по нов проект, фокусиран върху изследване на възможностите за развитие на дистанционните методи за измерване и оценка на радиационната обстановка върху земната повърхност – на открито и закрито. Това е методът на въздушното гама-картографиране, който е относително нова тематика за екипа на проекта. Членове на този екип са взели участие и в първия и засега единствен експеримент по въздушно гама-картографиране, осъществен от българската фирма „Тита Консулт“ ООД, на относително голяма открита площ на българска територия с помощта на пилотируеми летателни средства [6, 15]. Това е един много успешен експеримент, при който са постигнати значими и използваеми за практиката резултати и това дава надежда, че споменатият метод може да се развива и в бъдеще.

Новата цел обаче е използването на БЛС с фиксирано или ротиращо крило, което изисква цялостно преосмисляне на измерителната постановка, използваните детектори, електронните схеми за реализация на дистанционните измервания, придобиване на опит при съставяне на ефективни летателни планове и нов подход при реализация на метода за гама-картографиране.

Като първи проект на екипа в тази област през настоящата година са решени няколко задачи с важно за бъдещото продължаване на работата значение. На първо място това е сравнителен анализ и избор на подходящ гама-детектор и електронна схема за измерване. Това включва проучване, подбор, лабораторни (наземни) експерименти и анализ на налично и ново оборудване с оглед на три важни параметъра – чувствителност на детектора, тегло на цялата измерителна схема и ефективност по отношение на консумацията на енергия от батерията. Следващата важна задача е проучване на приложимостта на метода за дистанционно гама-картографиране с БЛС за различни терени – на открито поле, планинска среда, градска и плътна градска среда и във вътрешността на сгради. Това е свързано с използването на нов подход и нова технология за създаване на летателния план на използваното БЛС. Много важна задача на проекта е и реализацията на планиран експеримент за гама-картографиране на открита малка площ с оглед на доказване на практическата работоспособност на метода. Това е относително сложен експеримент и допълнително е използвана консултантска помощ на външни за екипа експерти по БЛС. Накрая са обобщени резултатите от проучвателната и експерименталната работа по проекта и е предложен новият метод за относително не много скъпо дистанционно гама-картографиране на малки площи с помощта на БЛС. Получени са и полезни резултати относно чувствителността на метода и експерименталната неопределеност, както и препоръки за бъдещо практическо използване на метода.

С развитието на новите технологии в областта на автоматизацията и по-продължителната автономност на БЛС се откриват и нови възможности за избягване на досег на човека с високи нива на радиоактивност при откриване на такива източници. Подобна радиоактивност би могла да бъде остатъчно замърсяване вследствие на ядрени аварии, замърсяване в близост до уранови мини, в складови помещения, както и повишена естествена радиоактивност в почвата, подпочвените скали и др.

За осигуряване на безопасност на персонала и населението първото необходимо условие е провеждане на дистанционно изследване на радиационната обстановка. Един слабо познат в България метод е въздушното гама-картографиране. Докторант от настоящия екип е имал възможността да участва в първото и вероятно единственото до момента въздушно гама-картографиране, извършено на територията на България от българска фирма със съдействието на Софийския университет „Св. Климент Охридски“. За целта е използвана апаратура, закупена от чуждестранен производител без достатъчно приложен опит в областта. Наблюденията по време на картографирането разкриха много възможности на този дистанционен метод, които не са достатъчно добре развити. Идеите ни за бъдещо развитие на метода и възможността да бъде приложен посредством БЛС ни подтикна към разработването на въпросния проект.

За изследването на познатите методи и развитието на нови ни е необходимо експериментално потвърждение на резултатите. За реализацията на такъв експеримент сме подбрали няколко вида детектори и изчислителна електроника. Тази апаратура се предвижда да работи на борда на БЛС по време на полет. Това предполага високи изисквания към чувствителността и бързодействието от физична гледна точка. От техническа гледна точка за продължителен и ефективен полет е необходимо апаратурата да бъде лека и с понижена консумацията на енергия. Освен груба оценка на радиоактивността на обследваните обекти е добре да познаваме и произхода на тази радиоактивност, за да преценим дали тя представлява интерес за нас. Това става чрез спектрометрични детектори, които от своя страна имат по-специални изисквания към електрониката, генерират по-големи масиви от данни и материалът, от който са изработени, е по-скъп. Сложността на проблема, пред който заставаме, ни принуждава да го решим на малки стъпки. Първата от тях би била изборът и набавянето на необходимия за експеримента детектор. За целта ни е необходимо сравнение на познатите до момента детектори, подходящи за гама-спектрометрия, проучване на пазара и закупуването на подходящ детектор. Към момента екипът ни сътрудничи тясно с фирма „Тита Консулт" ООД (с опит във въздушното гама-картографиране), както и с водещи фирми, произвеждащи дозиметрична апаратура: Mirion Technologies, Ortec, Flir и Thermo-Fischer, разработващи също и гама-детектори.

В България все повече се развиват и безпилотните летателни средства, усъвършенстват се комуникациите с тях, позиционирането, увеличаването на продължителността на полета и полезната маса. Засега повечето се използват главно за заснемане на земната повърхност с оптични, инфрачервени или хипер-спектрални камери, но има потенциал да се използват и за дистанционно гама-картографиране. В България има достатъчно специалисти по БЛС, включително и в Софийския университет (преподаватели в магистърска програма «Аеро-космическо инженерство и комуникации» във Физическия факултет), които с готовност ще подпомогнат още в самото начало реализацията на подобен дистанционен метод за гама-картографиране с помощта на БЛС и ще ни осигурят необходимата техника. Това ще е една наистина нова възможност за практическо и изключително полезно използване на БЛС.

Работата по един такъв проект за развитие на нов метод за дистанционно гама-картографиране с помощта на БЛС е свързана с преследването на няколко основни цели.

♦ Най-важната цел е да се докаже работоспособността и ефективността на метода на радиационното гама-катрографиране от въздуха с помощта на БЛС с фиксирано или ротиращо крило. Това наистина е важна цел в момента, защото досега няма ясни и категорични доказателства за приложимостта на подобна технология на радиационно картографиране с използване на относително евтини и икономични летателни средства, с прилагане на малки по маса, но достатъчно ефективни гама-детектори и накрая – възможността за реализация на такъв летателен план и използване на такъв полезен товар, който да осигурява качествено, с достатъчно висока чувствителност и разделителна способност радиационно гама-картографиране отначало на открити терени, а в последствие и на терени на закрито и сложни градски терени.

♦ Друга цел, която има важно практическо значение и за по-нататъшната работа в това направление, е осъществяването на полезен сравнителен анализ на наличните и на новите съвременни детектори на гама-лъчение не само по главните им характеристики (ефективност, мъртво време, разделителна способност по енергии и пр.), но в допълнение – по полезната маса и енергийната ефективност. Подобен анализ не е реализиран досега, защото теглото на детектора и консумираната мощност не са от основно значение при стандартните приложения. При използване в БЛС обаче тези характеристики са особено важни и това е необходимо, за да се намери подходящият баланс.

♦ Конкретна цел е придобиването на опит в планирането на подходящи профили за облитане на дадена област (създаване на подходящ летателен план) в зависимост от нейния характер и поставените цели на изследването. Това е важна цел, защото съставянето на този план за облитане зависи от модела на евентуалното радиоактивно замърсяване на дадената област – точков, линеен или непрекъснат.

♦ Експерименталното потвърждение на възможността за постигане на достатъчна точност при въздушното гама-картографиране, сравнено с обикновеното наземно картографиране, също е специално търсен резултат.

♦ Цел е и обучение на млади хора – студенти, докторанти и млади учени, в експериментална работа и опит във въздушното гама-картографиране. За част от екипа тази дейност е пряко свързана с тематиката на дисертационния им труд. Цел е и подпомагане на дейност по усъвършенстване на лабораторни практикуми, защити на дипломни и курсови работи, дисертационни трудове и др.

♦ Ключова цел е да се набави една достатъчно добра собствена техническа база от измерителни средства (детектори, електронни и комуникационни схеми, методи и софтуерни средства) и достатъчно подготвен екип, който да развива метода на въздушна гама-картография с помощта на БЛС.

## 3. СРАВНЕНИЕ МЕЖДУ РАЗЛИЧНИ ГАМА-ДЕТЕКТОРИ С ЦЕЛ ВГРАЖДАНЕ В БЛС

Това е първият важен въпрос за решаване: сравнение на различни гама-детектори по три от техните характеристики. В сравнението участва не само материалът, от който е изработена активната част на детектора, а цялостно работещата детектираща система. Това включва самият детектор (сцинтилатор или друг тип), фотоелектронен умножител (ФЕУ) или друг преобразовател на светлинни импулси в електрически, предусилвател, усилвател, електроника, броеща и записваща импулсите, и захранващ блок. Трите параметъра, които са най-важни за прилагането на детектиращата система в безпилотните методи за изследване, са: тегло, чувствителност и ниво на консумация на електрическа енергия. Поради структурата им газоразрядните броячи са относително леки, но тяхната ефективност е ниска. Чувствителността на сцинтилационните детектори зависи от размерите, плътността и ефективния атомен номер $Z_{\text{еф}}$ на сцинтилатора (ефективността във фотопика е пропорционална на $Z^5$).

Класическите ФЕУ са с голямо усилване, но се изработват от стъкло, което ги прави тежки и чупливи. Преди да се проектира цялостна детекторна система, е направено сравнение между готови детекторни системи, използвани в преносима дозиметрична апаратура, където параметрите тегло, ефективност и консумация на ел. енергия също са решаващи. Според каталожната информация, дадена от производителите на детектори, е избран сцинтилационен детектор от CsI. За първоначалните тестови полети е избран детектор № 2 (табл. 1). Поради добрата си чувствителност, малки размери и сравнително ниска цена би било рентабилно детекторът да се използва, когато не е нужна спектрометрия. За спектрометрични измервания по-подходящ би бил детектор № 5, който е от същия материал (CsI), но с по-голям обем, което го прави по-чувствителен. Част от обема на сцинтилатора на детектор № 5 е изработена от кристал, чувствителен към неутрони (Li-6). Цената му е по-висока, но би била оправдана предвид това, че ще позволи създаване на гама-карта на няколко отделни радионуклида и откриване на неутронни източници само с едно облитане на терена.

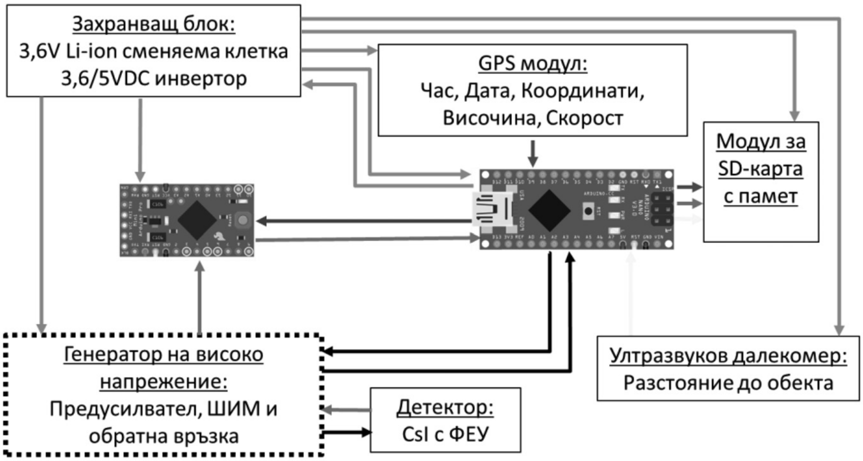
**Таблица 1**. Сравнителна таблица на изследваните детектори

| № | Вид/материал | Размери [mm] | Чувствителност [cps/μS/h] (Cs-137) | Тегло [g] |
|---|---|---|---|---|
| 1 | ГМ-брояч | 80 ґ | 8 | 8 |
| 2 | CsI | 38 ґ | 210 | 40 |
| 3 | CZT | 1000 $mm^3$ | 1000 | 60 |
| 4 | NaI | 25 ґ | 290 | 120 |
| 5 | CsI+Li | 38 ґ | 1500 | 550 |
| 6 | NaI | 63 ґ | 4600 | 1400 |
| 7 | Органичен сцинтилатор | 75 ґ | 5800 | 1400 |

Към всеки от детекторите като неразделна част е прикрепен фотоелектронен умножител (ФЕУ), който изисква добра защита от механично счупване на стъкления му корпус и стабилизирано високо захранващо напрежение (до около 1000 V DC). За първоначалните експерименти достатъчна ще бъде система само с възможност да определя общата активност на замърсяването, без спектрометрия (както е представена на фиг. 1). При нея импулсите от ФЕУ се подават към операционен усилвател, формираща верига и брояч. Броячът е реализиран с два контролера на базата на ATmega328p [16]. Единият контролер служи само за отброяване на импулсите в режим на прекъсване, а вторият ги записва на карта с памет. Към записа се добавят и данните от GPS приемник и ултразвуков далекомер, необходим за определяне на относителната височина на полета, а спрямо нея и замърсяването, отнесено към земната повърхност.

Преди да се осъществи полет, детекторната система е тествана в условия, максимално близки до реалните, с помощта на контролни източници. Използваните източници са от два различни радионуклида: Cs-137 и Co-60. Подбраните радионуклиди са едни от най-често срещаните в радиоактивни замърсявания, причинени от човешка дейност. В този експеримент са постигнати очакваните резултати от предварителните теоретични пресмятания, което доказва, че системата работи и може да се пристъпи към реалните полеви тестове. Важен резултат, получен от експеримента, е откриването на височината, от която облитането би било оптимално по отношение на скорост, безопасност и минимална детектируема активност (МДА).

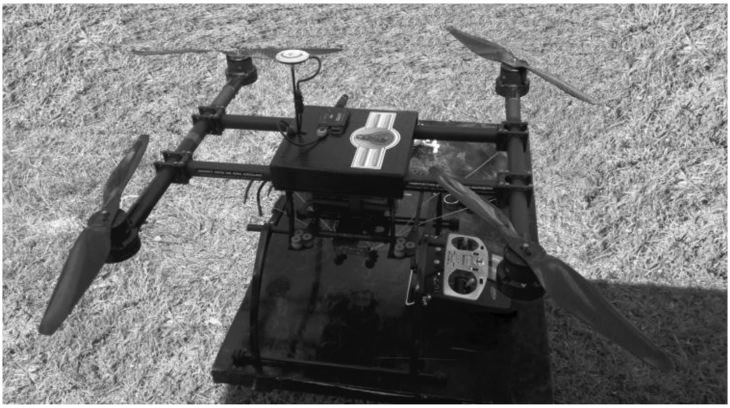
Следващата стъпка е извършване на софтуерна симулация на поглъщане на гама-лъчението в различни материали. Симулацията е извършена с два софтуерни продукта [17]: MicroShield и RadPro за потвърждаване на резултатите. Въведените параметри са за три различни дебелини на слоя въздух при влажност 50%, температура 20 °C и налягане 950 hPa (дебелината на слоя въздух отговаря на относителната височината на полета). Освен въздух са въведени и четири вида поглътители от очакваните материали, покриващи радиоактивното замърсяване. За да се въведат коректни данни, експериментално е определена плътността на почва, пясък, тревиста и храстовидна растителност. Резултатите показват, че отслабването на гама-лъчението във всеки от въведените поглътители е значително и прилагането на поглътител в полевите експерименти ще изисква източници с по-голяма активност от тези, с които разполагаме към момента. Това определи и бъдещия ни подход към полевия експеримент (вж. следващия раздел) – да се реализира открито изкуствено замърсяване с цел ефективно приложение на технологията за гама-картографиране с помощта на БЛС.



**Фиг. 1**. Блок схема на детекторната система за запис на данните

# 4. ПРОВЕЖДАНЕ НА ТЕСТОВ ПОЛЕВИ ЕКСПЕРИМЕНТ С ИЗБРАНАТА ГАМА-ДЕТЕКТОРНА СИСТЕМА

В тази част е описано провеждането на първите полеви експерименти с БЛС, носещо на борда си детекторна система за гама-лъчение. За полета е избрано БЛС, тип ротиращо крило, използвано за учебни цели в Геолого-географския факултет на Софийския университет [18] – вж. и илюстрацията на фиг. 2. То представлява електрически квадрокоптер с обща подемна сила на четирите ротора около 100 N. Наличната батерия е литиево-полимерна с напрежение 22.2 V и капацитет 10 Ah, което на практика осигурява полет с продължителност около 10 min с подбраната детекторна система на борда.



**Фиг. 2**. Квадрокоптер с обща подемна сила на четирите ротора около 100 N [18], притежание на Университетския център за въздушно наблюдение към СУ

AUTO 115200

Distance: 0.0501 km
Prev: 14.54 m AZ: 113
Home

GEO 42.673967 23.330377
SRTM 609.27m

☐ Grid View KML
GoogleSatelliteMa
Status: loaded tiles

Lat
Long
Alt (abs) 2

©2017 Google · Map data ©2017 Tele Atlas, Imagery ©2017 TerraMetrics

Waypoints

WP Radius 30 Loiter Radius Default Alt 100 Relative ☐ Verify Height Alt Warn 0 ☐ Spline

| | Command | Dela | | | | Lat | Long | Alt | Delete | Up | Down | Grad % | Angle | Dist | AZ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ▷ 9 | WAYPOINT | 0 | 0 | 0 | 0 | 42.6736410 | 23.3302245 | 611 | X | | | 0.0 | 0.0 | 4.0 | 153 |
| 10 | WAYPOINT | 0 | 0 | 0 | 0 | 42.6736311 | 23.3302314 | 611 | X | | | 0.0 | 0.0 | 1.2 | 153 |

**Фиг. 3**. Летателен план, изготвен със софтуера MissionPlaner

Като полеви терен за тестовия полет е избрана част от паркинга пред Физическия факултет с общи размери около 10 г.

| Точка № | Северна ширина [deg] | Източна дължина [deg] | Скорост на броене [cps] |
|---|---|---|---|
| 1 | 42.6736268 | 23.3302206 | 16 |
| 2 | 42.6736268 | 23.3302206 | 17 |
| 3 | 42.6736589 | 23.3301986 | 17 |
| 4 | 42.6736911 | 23.3301765 | 29 |
| 5 | 42.6737014 | 23.3301694 | 18 |
| 6 | 42.6737053 | 23.3301804 | 17 |
| 7 | 42.6737053 | 23.3301804 | 18 |
| 8 | 42.6736731 | 23.3302025 | 16 |
| 9 | 42.6736410 | 23.3302245 | 17 |
| 10 | 42.6736311 | 23.3302314 | 17 |
| 11 | 42.6736353 | 23.3302421 | 17 |
| 12 | 42.6736353 | 23.3302421 | 18 |
| 13 | 42.6736674 | 23.3302201 | 18 |
| 14 | 42.6736995 | 23.3301980 | 16 |
| 15 | 42.6737092 | 23.3301914 | 16 |
| 16 | 42.6737131 | 23.3302024 | 19 |
| 17 | 42.6737131 | 23.3302024 | 17 |
| 18 | 42.6736809 | 23.3302245 | 18 |
| 19 | 42.6736488 | 23.3302465 | 31 |
| 20 | 42.6736395 | 23.3302529 | 16 |
| 21 | 42.6736438 | 23.3302637 | 17 |
| 22 | 42.6736438 | 23.3302637 | 17 |
| 23 | 42.6736759 | 23.3302416 | 18 |
| 24 | 42.6737080 | 23.3302196 | 19 |
| 25 | 42.6737170 | 23.3302134 | 18 |

При него профилите на облитане представляват успоредни прави линии на разстояние, равно на височината на полета (2 m). Скоростта на полета е съобразена както с неговата височина, така и с чувствителността на детектора, скоростта на броене (табл. 2) и желаната МДА.

Измереният естествен фон на площадката, предвидена за експеримента, е средно 0.08 μSv/h. При такъв фон откликът на избрания детектор № 2 е около 0.08×210 = 16.8 ≈ 17 cps. Статистическата неопределеност би била $17^{½}$ = 4.12 cps. Теоретично няма как при тези условия на измерване (детектор, време на интегриране, геометрия и гама-фон) да постигнем МДА, по-малко от 4.12 cps нетна стойност, равняваща се на мощност на дозата ~0.02 μSv/h (закръглена нетна стойност). Тоест, за да открием използваните източници с нашата детекторна система, всеки един от тях трябва да създава поле с мощност на дозата минимум 0.02 μSv/h на разстояние 2 m от него (колкото е очакваната височина на полета).

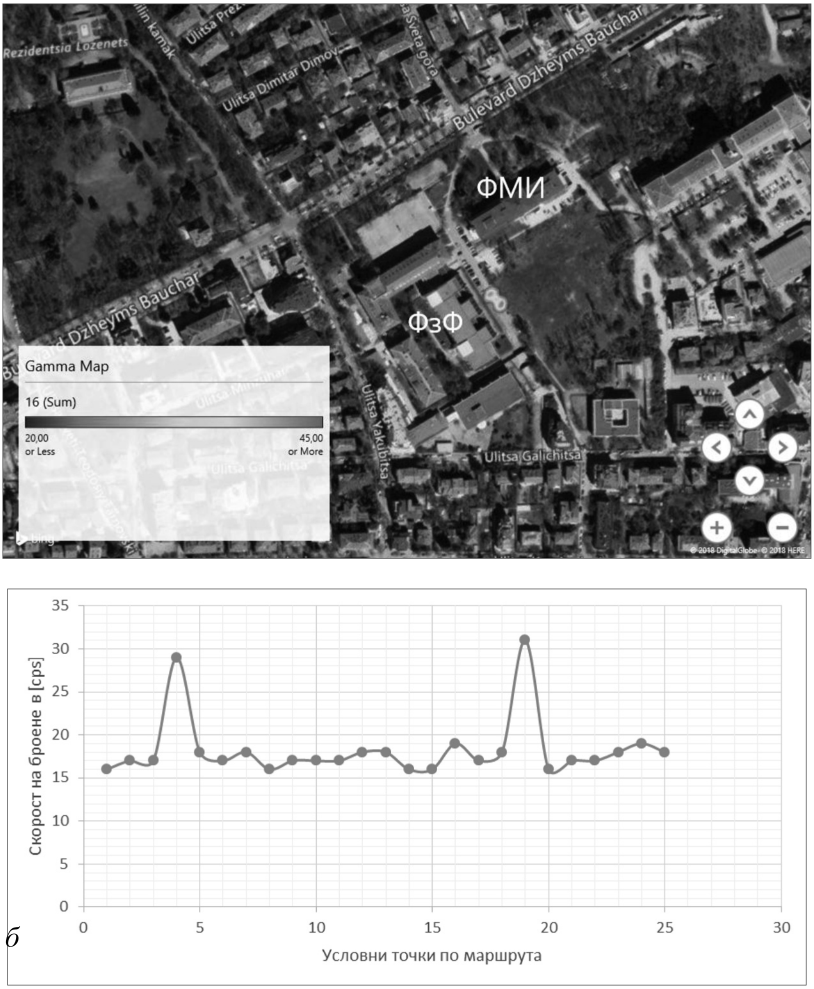
От табл. 2 (т. 4 и т. 19) и фиг. 4 се вижда, че откликът в детектора в точките над източниците е около 30 cps, или при фон 17 cps нетната стойност е около 13 cps. Това показва, че източниците, подбрани за експеримента, създават поле, което е достатъчно силно, или по-конкретно ~0.06 μSv/h.

При повторното облитане с цел по-точно определяне на активността на източниците времето за интегриране може да се увеличи чрез зависване на БЛС само над точките, в които е забелязано завишаване на фона.

За така създаденото изкуствено замърсяване и установени условия за измерване за реализацията на предварителните тестови изследвания не беше използвана реална БЛС и експериментът се проведе с помощта на дистанционна щанга, носена на ръце и поддържаща детекторната система на височината на полета на БЛС. Впоследствие бе проведен и редовен експеримент с БЛС, но резултатите ще бъдат описани на по-късен етап и по-подробно, като ще се разгледат и други технически въпроси за тази технология. Един от тези въпроси е свързан със съвсем конкретното откриване на източниците чрез наслагване на гама-картата върху детайлно изображение, създадено от фотозаснемане на терена. Това заснемане лесно би могло да се осъществи по време на полет с помощта на миниатюрна камера с висока разделителна способност, монтирана на БЛС.

Допълнително затруднение предизвика използването на програмите MissionPlaner и 3D Maps (към пакета на Microsoft Excel) поради факта, че те са адаптирани към по-едромащабни карти. В настоящия експеримент се използваха техните пределни възможности. За по-дребномащабни обследвания е необходимо квадрокоптерът да е снабден с различна навигационна система, по-точна от GPS (споменатият лидар или друго).

След приключване на тестовите измервания на симулационен полет данните са свалени от картата с памет на детекторната система и обработени на преносим компютър на място в близост до обследваната територия. Целта на експресния анализ на данните е да се открият местата със замърсяване, за да се планира втори полет, използващ линеен или точков модел на разпределение на замърсяването.



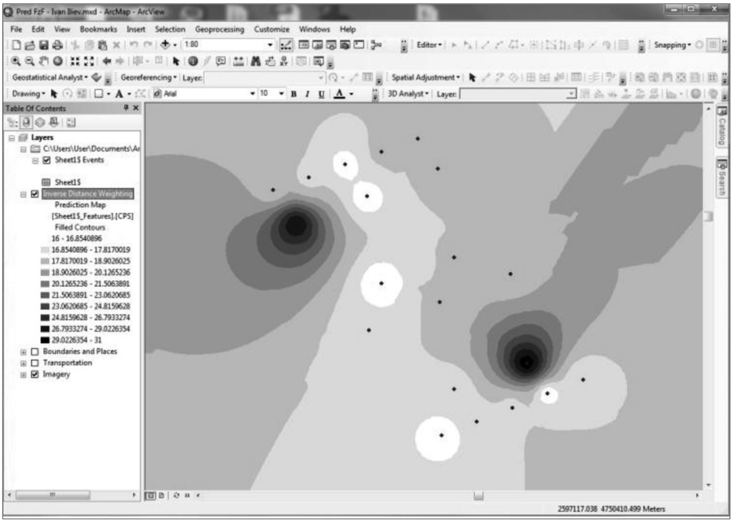
**Фиг. 4.** Графично определяне на разпределението на замърсяването (*а*) и установената скорост на броене (б)

Докато непрекъснатият модел е необходим за локализиране на замърсяването, то точковият и линейният модел могат да ни дадат по-точна представа за активността. В нашия случай за втория полет е избран точков модел, приложен в две точки над две открити замърсявания. На фиг. 4 са демонстрирани два метода, използвани за експресен анализ на замърсяването посредством Microsoft Excel.

При втория полет е необходимо зависване над центъра на откритото замърсяване с цел по-продължително измерване. При такова зависване има риск БЛС да отвее замърсяването, с което може дори самото БЛС да се замърси, затова се предпочита по-продължително зависване пред по-малката височина.

Накрая се извършва окончателна обработка на данните от измерването. След лабораторните експерименти детекторната система е програмирана да записва данните от измерването във файлов формат, подходящ за директно прехвърляне в програмата ArcGis, предоставена за целите на експеримента от фирма „Тита Консулт“. На фиг. 5 е представена картата, генерирана с ArcGis и съдържаща измерените скорости на броене. Картата позволява множество манипулации и интерпретации на резултатите с цел по-лесен достъп до информацията.

След първоначалния полет се наблюдават две петна със завишено ниво на радиоактивност (фиг. 4*б*). При втория полет координатите на тези петна се задават като точки на зависване на БЛС за определено време. Предвид първоначалната груба оценка, на база на непрекъснатия модел при първото облитане, е определено времето на зависване при второто облитане – 100 s. Това е 10 пъти по-дълго време на интегриране от времето при непрекъснатия модел, което намалява неопределеността 3 пъти. В конкретния случай активността на източниците бе определена точно благодарение на предварителната калибровка на системата в същата геометрия и със същите източници. При реални замърсявания би следвало да се спазва геометрията на калибрирането, доколкото е възможно, а точното определяне на активността на замърсеното петно ще зависи и от това, дали замърсяването е от същите радионуклиди, които са използвани при калибрирането.



**Фиг. 5**. Резултати от тестовото измерване под форма на карта, генерирана с помощта на софтуера ArcGis

## 5. ИЗВОДИ И БЪДЕЩА РАБОТА

Първоначалните сравнителни анализи, симулации и първи експериментални резултати от изпълнението на настоящия нов проект показаха, че има сериозни основания да се счита, че дистанционната гама-картография с помощта на безпилотни летателни системи (самолети и квадрокоптери) осигурява необходимата точност и ефективност на подобен род измервания и може да замени в някакъв смисъл скъпо струващи пилотируеми измервания, когато се търсят бързи, евтини и относително по-груби първоначални резултати, особено на места, където подобни експерименти, изпълнени с участие на хора, са трудно-осъществими (тук не се визират само опасни места на аварии и бедствия с радиоактивно замърсяване, а по-скоро трудно достъпни терени вън и във вътрешността на сгради и на далечни разстояния от измервателни лаборатории).

Бъдещата работа на екипа включва провеждане на серия от полеви експерименти с различни БЛС, усъвършенстване на управлението на БЛС с основна мисия гама-картографиране, усъвършенстване на летателния план и обогатяване на мисията чрез включване на допълнително оптично заснемане на терена, точно определяне на относителната височина на полета и други подобни аерокосмически технологии. Екипът има намерение да продължи работата по този обещаващ проект.

**Благодарности.** Изследванията са проведени с частичната финансова поддръжка по Договор 80-10-82/2017 с Фонд НИ на СУ

### REFERENCES

[1] *ICRU, Report* **53**, U.S.A, 01.12.1994.

[2] International Atomic Energy Agency, IAEA, *IAEA-TECDOC-1092*, June 1999.

[3] International Atomic Energy Agency, IAEA, *IAEA-TECDOC-1363*, July 2003.

[4] Mabit, L. and C. Bernard. *Journal of Environmental Radioactivity*, October 2007, **97**, 2-3, 206-219, http://www.sciencedirect.com/science/article/pii/S0265931X07001336

[5] Horsfall, K.R. *AGSO Journal of Australian Geology & Geophysics*, 1997, **17**(2), 23-30.

[6] Iliev, I., I. Pastuhov, V. Gourev. Airborn gamma-mapping, *RAD Conference 2016*, Nish, Serbia, 23-27.05.2016, online available.

[7] V. Terziev, M. Ivanova, D. Sharalieva, A. Pipev, Methodology for realization of measurements on the base of "In situ" method by mobile Gamma-spectrometric equipment for determination of the activity on the objects with different geometric shape and stationary location in the enveloped environment, Manual, Eco Programma Ltd., 2017 (in Bulgarian)

[8] V. Terziev, M. Ivanova, D. Sharalieva, A. Pipev, Methodology for calibration of mobile Gamma-spectrometric equipment, Guidance, Eco Programma Ltd., 2017 (in Bulgarian)

[9] Connor, D. T., P. G. Martin, and T. B. Scott. *International Journal of Remote Sensing*, 2016, **37**(24), 5953-5987. DOI:10.1080/01431161.2016.1252474.

[10] Zavala, M. *PhD Thesis*, Georgia Institute of Technology, May 2016, online available.

[11] Martin, P., J. Moore, J. Fardoulis, O. Payton, and T. Scott, *Remote Sensing*, 2016, **8**(11), [913]. DOI: 10.3390/rs8110913

[12] Šálek, O., M. Matolín, and L. Gryc, *Journal of Environmental Radioactivity*, 2017, **182**, 06 Dec, 101-108.

[13], I. Iliev, P. Dankov, I. Pastuhov, „Analysis of the possibility for utilization of unmanned vehicles and small satellite for remote control of the radiation situation, *Third National Congress on Physical Science*, 29.09-02.10.2016, Sofia, InterExpo Center (in Bulgarian)

[14] Karanth, S. P., V. Shobha, M. A. Sumesh, T. V. Sridevi, K. T. Manjunath, B. Thomas, L. V. Prasad, and M. Viswanathan, *Journal of Small Satellites*, vol. 6, No. 2, 2017, pp. 581–589

[15] Iliev, I., V. Vasilev, M. Mladenova, *Int. Conf. Comprehensive Nuclear-Test-Ban Treaty (CTBT)*, Vienna, Austria, 26–28.06.2017 (to be published in Journal of Physical Science and Application, ISSN 2159-5348)

[16] University Center for air monitoring of Sofia University "St. Kliment Ohridski": https://www.uni-sofia.bg/index.php/bul/universitet_t/centrove/universitetski_cent_r_za_v_zdushno_nablyudenie

[17] Manual of the software for Arduino controllers: https://www.arduino.cc/en/Guide/HomePage

[18] Manual of the software Gamma Vision: http://www.ortec-online.com/-/media/ametekortec/manuals/a66-mnl.pdf;

[19] Manual of the software MAESTRO: http://www.ortec-online.com/-/media/ametekortec/manuals/a65-mnl.pdf;

[20] Manual of the software MicroShield: http://radiationsoftware.com/microshield/;

[21] Manual of the software Mission Planner: http://ardupilot.org/planner/docs/common-mission-planning.html;

[22] Manual of the software Geosoft (package): http://www.geosoft.com/products/target/target-arcgis/; [23] Manual of the software Gamma Design: https://www.gammadesign.com/Features.aspx;

[24] Manual of the software ArcGis (package): http://www.esri.com/arcgis/about-arcgis

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ

Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS

Volume 111

# SEASONAL VARIATIONS OF THE MESOSCALE CIRCULATION IN THE EASTERN BLACK SEA BASIN

GRETA GEORGIEVA, ELISAVETA PENEVA

*Department of Meteorology and Geophysics*

*Грета Георгиева, Елисавета Пенева*. СЕЗОННИ ВАРИАЦИИ НА МЕЗО-МАЩАБНАТА ЦИРКУЛАЦИЯ В ИЗТОЧНИЯ БАСЕЙН НА ЧЕРНО МОРЕ

Комбинирани алтиметрични данни, обработени и дистрибутирани от Морските услуги на Програма Коперник, са анализирани с цел да разкрият сезонната изменчивост на Черноморската мезомащабна циркулация. Зимната и лятната циркулация са с подобни хоризонтални характеристики, но с обратен знак: циклонична аномалия на основното черноморско течение и Батумския вихър през зимата и антициклонична – през лятото. Преходът между тези две състояния е през април–май и октомври–ноември. Хофмюлеровата диаграма на аномалията на свободната повърхност показва разпространение на сезонния сигнал със скорост от порядъка на 1 cm/s.

*Greta Georgieva, Elisaveta Peneva*. SEASONAL VARIATIONS OF THE MESOSCALE CIRCULATION IN THE EASTERN BLACK SEA BASIN

The multi-mission altimeter data for Black Sea Level Anomaly, processed and distributed by Copernicus Marine Environment Monitoring Service, are analyzed in order to reveal the seasonal variability of the Black Sea mesoscale circulation. Winter and summer circulation show similar patterns with opposite sign: cyclonic anomaly of the Rim current and Batumi eddy in winter and anticyclonic – in summer. The transition between these two stages occurs in April-May and October-November. The Hovmoler diagram of the SLA shows evidence of seasonal signal propagation with approximate speed ~ 1 cm/s.

***Keywords:*** mesoscale eddies, surface circulation, Black Sea
***PACS numbers:*** 92.10 ad, 92.10ak

---

*For contact*: Greta Georgieva, Department of Meteorology and Geophysics, Faculty of Physics, Sofia University “St. Kliment Ohridski”, 5 James Bourchier Blvd., Sofia 1164, Bulgaria, Phone: +359 2 8161289, Mobile: +359 887 538 044, E-mail: gretageorgieva7@abv.bg

## 1. INTRODUCTION

The Black Sea is an inland basin with a surface area of about $4.2 \times 10^5$ km$^2$ (zonal and meridional dimensions ~1000 km and ~400 km) and a maximum depth of 2210 m. It is almost completely surrounded by land but not totally isolated from the neighbouring seas and the world oceans. The Bosphorus Strait is the connection with the Marmara Sea (which is connected to the Aegean Sea through Dardanelles) in the southwest. In the north, the Black sea is connected to the shallow Sea of Azov through the Kerch Strait. The largest river discharges (about ~80% of the total river runoff in the sea) are in the north-western part [1], where is the major shelf region. The rest of the basin is much deeper.

The circulation of the Black Sea is characterized by a basin-wide cyclonic boundary gyre known as Rim Current, which is formed by several factors: the curl of the wind stress field and the fresh water discharges from rivers (buoyancy forces), bathymetry and thermohaline fluxes [2]. Within this current, two or more smaller cyclonic cells are formed as depicted in Fig. 1 [4]. They are usually referred as eastern and western gyres. The Rim current is quasi-geostrophic as it engages the surface and several hundred meters water column. In addition to the principal Rim Current, the Black Sea circulation system contains many mesoscale eddies (see Fig. 1), meanders and filaments spread over the basin. The Rim Current separates the cyclonically dominated inner zone from the anticyclonically dominated coastal area [2]. The Danube, Constanta, Kaliakra, Bosphorus, Sakarya, Sinop, Kizilirmak, Batumi, Sukhumi, Caucasus, Kerch, Crimea, Sevastopol eddies reside on the coastal side of the Rim Current zone. They are quasistationary and quasiperiodical structures. This means that they do not exist there for the whole time once they were formed, but more properly they are regions of most probable formation or existence of an anticyclonic circulation. Actually the individual eddies travel slowly along the Rim current.

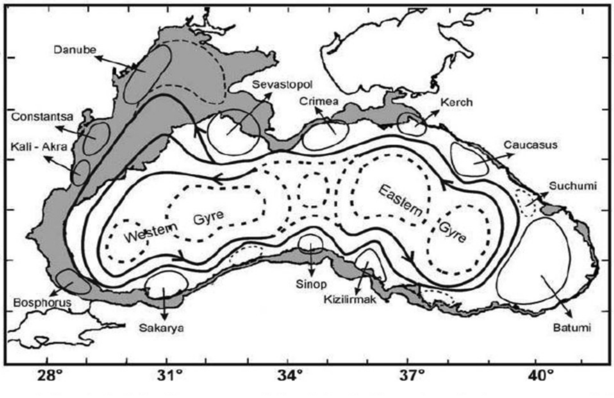


**Fig. 1.** Main features of the Black Sea circulation system [4]

In the Northwestern shelf region of the Black Sea the prevailing circulation is cyclonic with some anticyclonic gyres along the coast (the Danube eddy and the eddies in the vicinities of Odessa, gulf of Kalamita and the Karkinit bay). There is an exeption in cases of strong southern or southwestern winds; then the circulation transforms in anticylonic [3].

In addition to the above-mentioned there are three prominent points of bifurcation of the Rim current  and some areas of convergence. The bifurcations are: (i) near the Bosphor Strait, (ii) southwestward of the Crimea peninsula, (iii) near the entrance to the Kerch-Tamanian shelf. The most significant location of convergence is in the vicinity of cape Kaliakra [3].

The upper layer circulation exhibits significant seasonal and interannual variability. The main driver is the atmospheric circulation variations. According to [3] there are two basic 2D patterns of the atmospheric sea level pressure field: the former is observed for the winter-spring season and the latter for the summer-autumn. In the cold part of the year, due to the positive temperature contrast between sea and land,  a low-pressure system is formed over the Black Sea. During the warm period an anticyclonic circulation is observed over western part of Black Sea, leading to an eastward extension of the Azore maximum. Analysis of climatological data of annual variations of mean wind stress curl over the Black Sea region reveals cyclonic anomalies during the cold period and anticyclonic anomalies during the warm months [14].

In winter months (January, February and March) the Rim Curent is strong and narrow without any significant lateral variations. This is due to the stronger winter winds. The  inner cyclonic zone consists of two well developed gyres (Eastern and Western ones). The stronger winter conditions the stronger the Rim Current. In summer (July, August, September) Rim Current weakens and diffuses. The two-gyres inner pattern trasforms into one composite cyclonic cell. In early autumn (October) the Rim Current becomes hardly noticeable and inner area is completely disintegrated into small cyclonic cells. The reverse process of integration starts in the late autumn (November, December) [4]. The mesoscale features like anticyclonic outer eddies and meanders mostly exist late winter  to the early autumn. For example the Batumi gyre, one of the most persistent cells, forms in the early March and lives to the end of October. The only exception of this rule is the Sukhumi gyre, which forms when the Batumi eddy is over, and lives until it is formed again. Later it is absorbed by the Batumi gyre [4]. The horizontal sizes of the largest well-developed anticyclonic gyres are about 100-150 km; the vertical dimension is up to 1000 m [3].

Although the quasi-stationary mesoscale eddies have been a subject of investigation of many studies [2, 4–9] still lots of uncertainties exist about their evolution and propagation. This paper aims to investigate the formation and evolution of the Batumi eddy during the year using satellite altimeter data. Similar approach is used in [10].

## 2. DATA USED IN THE STUDY

The present paper uses altimeter data reprocessed by Copernicus Marine Environment Monitoring Service (CMEMS) downloaded from http://marine.copernicus.eu. This product is processed by the SL-TAC multimission altimeter data processing system. It processes data from all altimeter missions: Jason-3, HY-2A, Saral/AltiKa, Cryosat-2, Jason-2, Jason-1, T/P, ENVISAT, GFO, ERS1/2. It includes gridded sea surface height anomalies and derived geostrophic current velocities anomalies. The anomalies have been taken with respect to the 20-year means (1993–2012). The used data covers the period from January 1993 to December 2015. During this 23-year period there have been many missions (mentioned above) and they all have been homogenized with respect to a reference mission which is currently OSTM/Jason-2. The spatial resolution is 0.125° / 0.125° (about 10 km) and the time resolution is daily. The maps are calculated with all the satellites available (up to 4 satellites) for each date. The standart corrections have been made and data have been filtered from small scale signals [11].

## 3. SEASONAL VARIATIONS OF THE SEA LEVEL HEIGHT ANOMALY

As we are interested in the horizontal variability, the first step is to exclude the effect of the mean sea level variability. The latter is due to the contribution of water fluxes (river discharges, evaporation, precipitation, strait fluxes) and steric effects. According to [13] this type of variability is about 85% of the total variance of the Black Sea level. The calculated area average daily sea level height anomaly is given in Fig. 2. It reflects the tendency for increase in the last decade of 20th century, followed by relative stability and periodic oscillations. The derived mean is then substracted from actual SLA for each grid point. This correction is based on assumption that the response of sea level to these volume fluxes is almost uniform over the Black Sea area.

The daily maps of the SLA are averaged to obtain monthly mean horizontal maps and seasonal averages. The seasons are considered as follows: winter – January, February and March; spring – April, May and June; summer – July, August and September; and autumn – October, November and December.

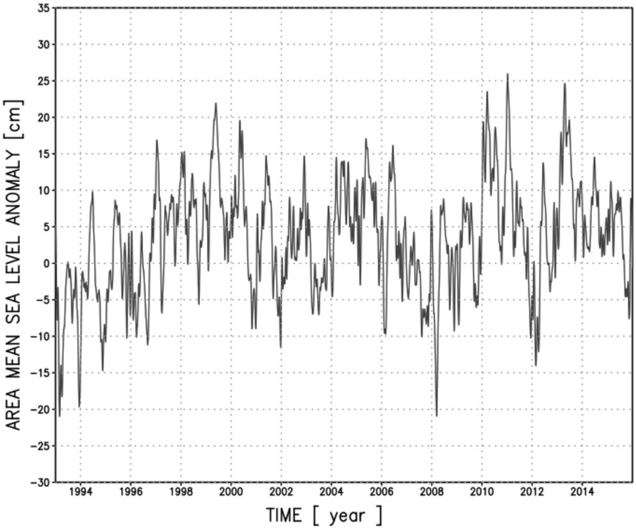


**Fig. 2**. Daily area mean sea level height anomaly [cm] for the period 1993–2015

The monthly mean maps for SLA are shown in Fig. 3 for each month. They reveal a significant intensification of the cyclonic circulation in winter and weakening during summer, this result is in accordance with the earlier studies findings. The anticyclonic eddies gather strength during spring season, achieve maximum severity in summer and later lose momentum. The most interesting feature in the eastern basin is the tripole structure with Batumi eddy in the easternmost end, which changes the anomaly sign seasonally. During the cold part of the year (December to March) Batumi eddy presents negative (cyclonic) anomaly; and in the warm part (June to September) – positive (anticyclonic) anomaly. Furthermore, this is accompanied with the alternating anomalies along the eastern Black Sea coast. In the transition period (spring April–May) a slight positive anomaly replaces the negative one and pushes it north-westward gaining momentum. The same process with the opposite sign occurs in the other transition period (autumn October–November).

The Batumi eddy reveals the most prominent anual amplitude. Acording to [4] the Batumi eddy is absent in winter months. Our analysis shows cyclonic anomaly during winter, so the Batumi eddy (which is considered anticyclonic) almost disappears. In this region we estimated SLA anual amplitude of 16 cm and the anticyclonic eddy is most pronounced in September.

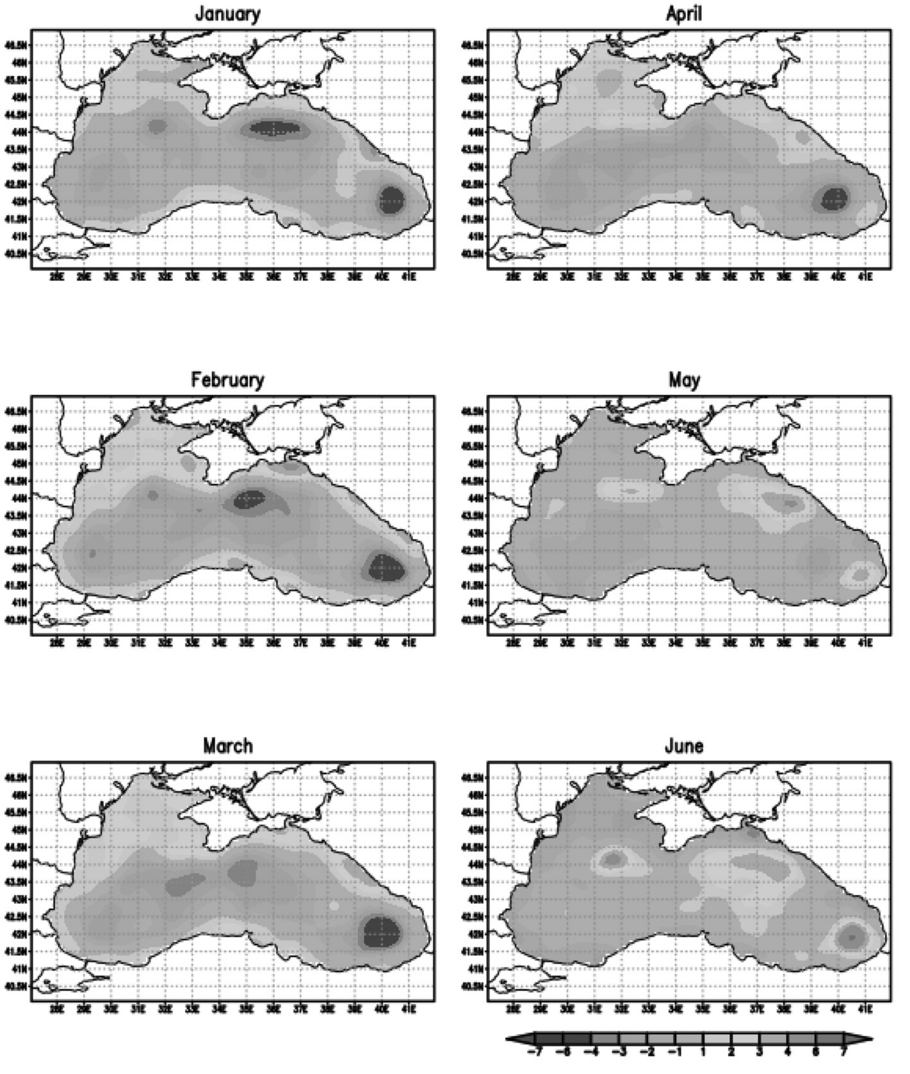


**Fig. 3**. Monthly mean sea level anomaly maps [cm] January to June, derived from averaged altimeter data for the period 1993–2015

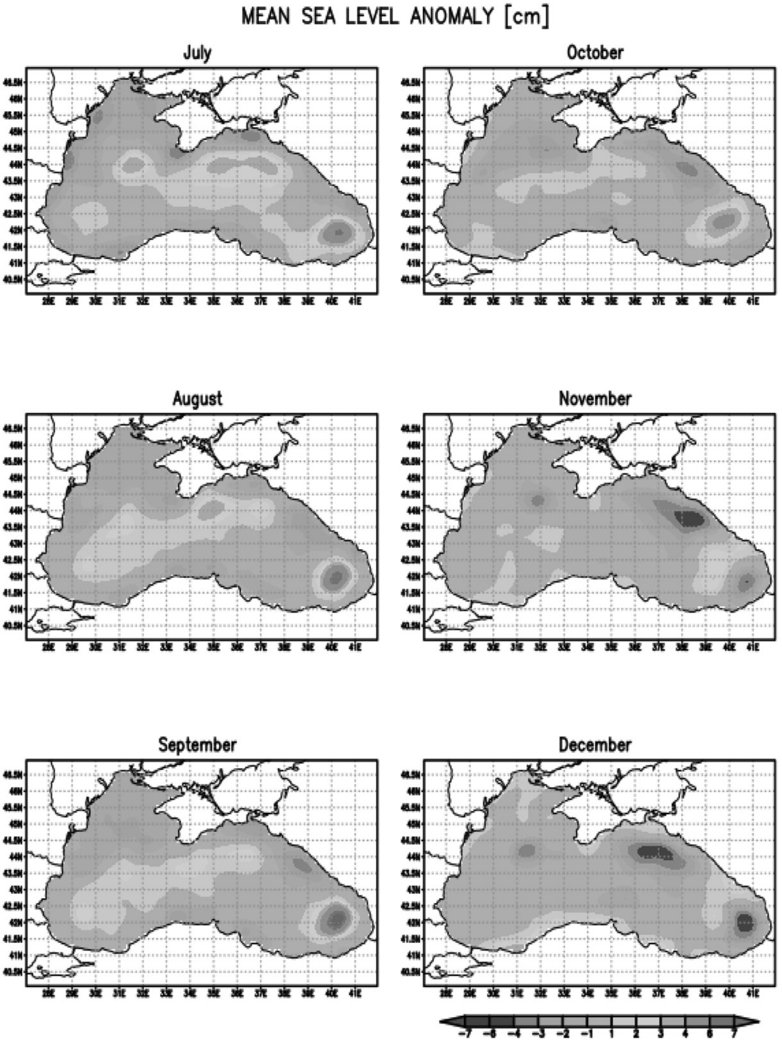


**Fig. 3**. (continued) Monthly mean sea level anomaly maps [cm] July to December, derived from averaged altimeter data for the period 1993–2015

The above described process is very well seen in the seasonal maps of the geostrophic current anomaly (Fig. 4). The geostrophic velocity anomaly is derived from the sea level height anomaly. The Rim current and the anticyclonic eddies along the coast are easily distinguished. Winter and summer circulation show similar patterns with opposite sign: cyclonic anomaly of the Rim current and Batumi eddy in winter and anticyclonic – in summer. The transition between these two stages begins with the appearance of small opposite-sign anomaly near Georgian coast, which slowly pushes the main eddy north-westward, takes its place and grows. The meanders of the Rim current exibit a wave pattern which leads to the conclusion of wave propagation and will be discussed in the next chapter. We have to note that similar process (less intensive) could be observed in the Sevastopol eddy area in the Western Black Sea basin.

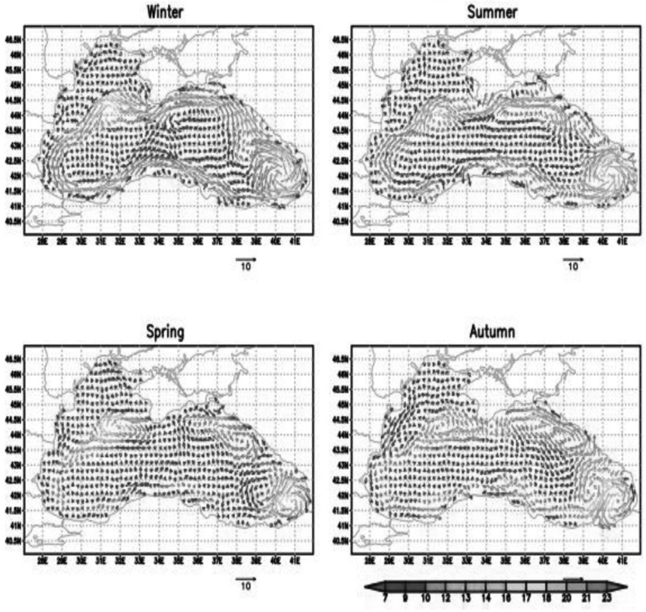


**Fig. 4**. Seasonal mean geostrophic velocity anomaly [cm/s] maps for the period 1993–2015

## 4. SEASONAL SIGNAL PROPAGATION ALONG THE EASTERN BLACK SEA COAST

In order to track the possible wave propagation of the seasonal SLA signal we have constructed a Hovmoller diagram of the monthly SLA fields along the cross-section AB (Fig. 5). The line AB is drawn to cross the eastern basin passing through the approximate centers of the tripole structure, extending from the Batumi area to the Eastern Crimean region.

The Hovmoller diagram for the monthly SLA for the whole regarded period 1993–2015 along the cross-section AB is plotted in Fig. 6 (left). There is an evidence of a propagating pattern from southeast to northwest. At most locations the SLA changes alternatively the sign in time: positive anomaly in summer, negative – in winter. This is especially pronounced in the easternmost part (40E–41E). The right plot in Fig. 6 is a zoom over the period 2012–2015 in order to better visualize the described above process. This allows us to approximately calculate the wave propagation speed – in an order of 1 cm/s.

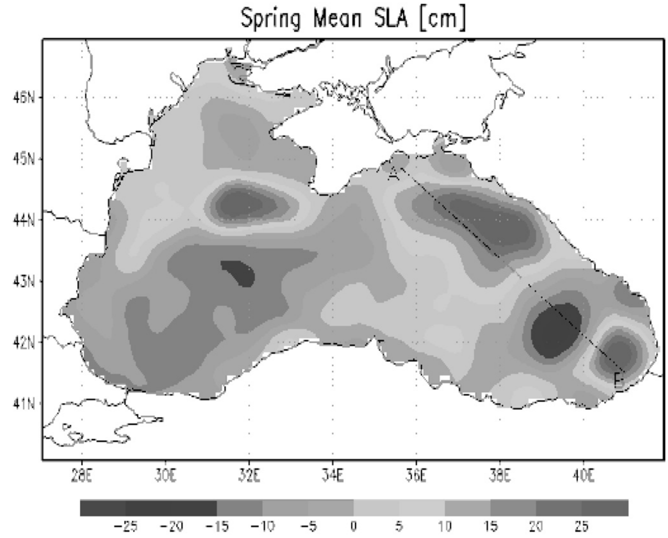


**Fig. 5**. Seasonal mean SLA [cm] for spring and the chosen cross-section AB

In this paper our objective is to show what is observed in the eastern Black Sea using altimeter data and not to search a theoretical explanation and quantification of the process. There are previous studies of the Black Sea circulation based on numerical simulations, which associate the alternating cyclonic and anticyclonic eddies to Rossby waves and basin modes [reference 6, 7, 12]. Our future plans include to further investigate the described above process in order to study the controlling physical mechanisms.

## 5. CONCLUSIONS

The Black Sea circulation reveals seasonal and interannual variability which have been studied for a long time. In winter months (January, February and March) the Rim Curent is strong and narrow without any significant lateral variations, due to the stronger winter winds. The mesoscale features like anticyclonic outer eddies and meanders mostly exist late winter  to the early autumn. Although the quasi-stationary mesoscale eddies have been a subject of investigation of many studies, still lots of uncertainties exist about their evolution and propagation. In this paper we investigate the formation and evolution of eddies in the eastern Black Sea basin during the year using satellite altimeter data.

We use Copernicus Marine Environment Monitoring Service (CMEMS) multi-mission altimeter data product, which combines data from Jason-3, HY-2A, Saral/AltiKa, Cryosat-2, Jason-2, Jason-1, T/P, ENVISAT, GFO, ERS1/2. It includes gridded sea surface height anomalies and derived geostrophic current velocities anomalies.

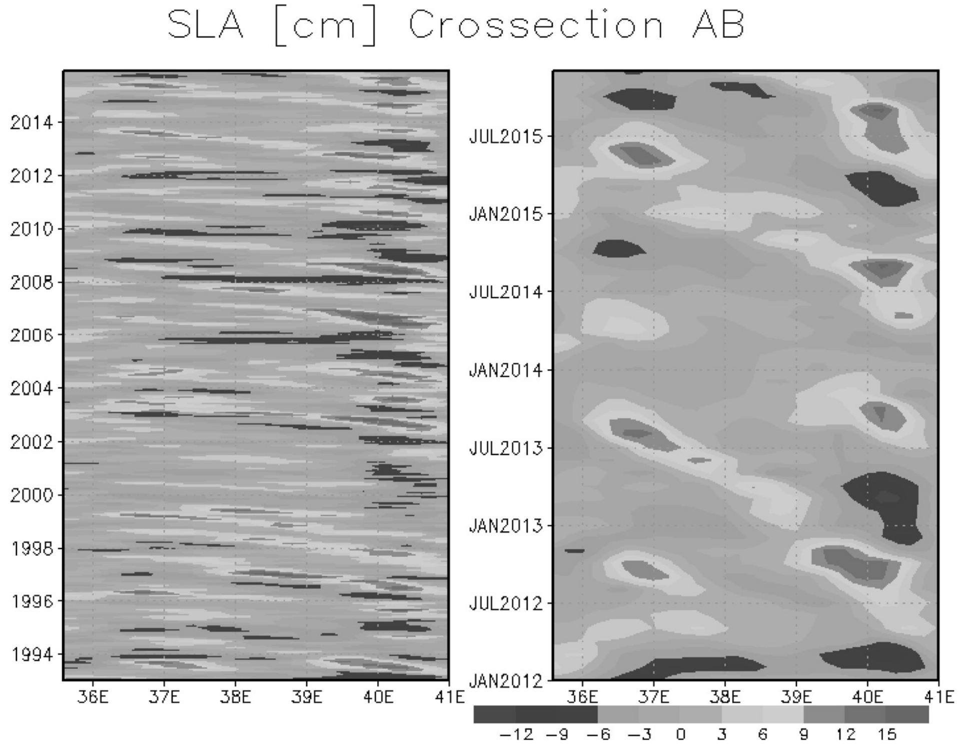


**Fig. 6**. The left plot shows monthly SLA [cm] for the period 1993–2015 along the cross-section AB. The right plot shows the same for the 4-year period 2012–2015 for better visibility

The anomalies have been taken with respect to the 20-year means (1993–2012). The used data covers the period from January 1993 to December 2015 with spatial resolution of 0.125º / 0.125º (about 10 km) and the time resolution is daily.

The first step of the analysis is to subtract the sea level signal associated with the water fluxes and steric effect. This is done in order to focus on the spatial variations. Then the daily SLA maps are averaged to obtain the monthly and seasonal mean maps.

The most interesting feature in the eastern basin is the tripole structure with Batumi eddy in the easternmost end, which changes the anomaly sign seasonally. Winter and summer circulation show similar patterns with opposite sign. During the cold part of the year (December to March) Batumi eddy presents negative (cyclonic) anomaly; and in the warm part (June to September) – positive (anticyclonic) anomaly. The transition between these two stages occurs in April-May and October-November: it begins with the appearance of small opposite-sign anomaly near Georgian coast, which slowly pushes the main eddy north-westward, takes its place and grows. Similar process is observed in the Sevastopol eddy area in the Western Black Sea basin.

We have analyzed the SLA along cross-section passing through the eastern basin. The Hovmoler diagram shows evidence of a propagating seasonal signal in the direction southeast-northwest with approximate speed of an order of 1 cm/s.

Previous studies based on numerical simulations associate such propagating eddies to Rossby waves and basin oscillations. Future investigation is planned in order to reveal the physical mechanism controlling the observed processes.

## REFERENCES

[1] Dobrovolsky, A.D. and Zalogin, B.S. USSR Seas, Moscow State University, 1982, (in Russian).
[2] Oguz, T., P. La Violette, U. Unluata. *J. Geophys. Research*, 1992, **97**, C8, 12569.
[3] Ivanov, V.A., Belokopytov, V.N. Oceanography of the Black Sea, NAS of Ukraine, Marine Hydrophysical Institute, Sevastopol, 2011, (in Russian).
[4] Korotaev, G.K., Oguz, T., Nikiforov, A., Koblinsky, C.J. *J. of Geophys. Research*, 2003, **108,** C4, 3122.
[5] Stanev, E.V., Beckers, J.M. *Journal Marine Systems,* 1999 , **19**, 65.
[6] Stanev, E.V., Rachev, N.H. *Jourmal Marine Systems*, 1999 , **21**, 283.
[7] Stanev, E.V., Staneva, J.V. *Jourmal Marine Systems*, 2000, **24**, 3.
[8] Stanev, E.V., Staneva, J.V. *Dynamics of Atmospheres and Oceans*, 2001 , **33**, 163.
[9] Enriquez, C. E., G. I. Shapiro, A. J. Souza, A. G. Zatsepin. *Ocean Dynamics,* 2005, **55**, 476.
[10] Korotaev, G.K., Saenko, O.A., Koblinsky, C.R. *J. Geophys. Researh* . 2001, **106,** № C1, 917.
[11] Pujol, M.-I., Faugère, Y., Taburet, G., Dupuy, S., Pelloquin, C., Ablain, M., and Picot, N. *Ocean Science*, 2016, **12**, 1067.
[12] Rachev, N.H., Stanev, E.V. *J. Of Physical Ocenography*, 1997, **27**, 1581.
[13] Stanev, E., E. Peneva. *J. Global and Planet. Change*, 2002, **32**, 33.
[14] Stanev, E., Piere-Yves Le Traon, Peneva, E. *J. of Geophys. Research*, 2000, **105**, 17203.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# UNUSUAL PREVALENCE OF COLOR VISION DEFECTS IN CHILDREN FROM A SMALL BULGARIAN TOWN

MILA DRAGOMIROVA, SNEJANA IORDANOVA

*Department of Optics and Spectroscopy*

Мила Драгомирова, Снежана Йорданова. НЕОБИЧАЙНО ВИСОКО РАЗПРО-СТРАНЕНИЕ НА НАРУШЕНИЯ В ЦВЕТНОТО ЗРЕНИЕ ПРИ ДЕЦА ОТ МАЛЪК БЪЛГАРСКИ ГРАД

Цветното възприятие, като една от важните характеристики на зрението, е изследвано по време на скрийнинг на 203 ученици на възраст между 6 и 18 години, живеещи в Сапарева баня. Изследването е проведено чрез псевдоизохроматични карти на Ишихара, тест на Farnsworth D-15 и тест на Lanthony Desaturated D-15. С оглед спестяване на време тестовете са проведени бинокулярно. Според картите на Ишихара измежду 87 момчета и 116 момичета 9.2% и 4.31% съответно имат наследствен цветови дефицит. Данните от статистическата обработка са сравнени с други известни от целия свят. Нито един от изследваните ученици по-рано не е изследван за цветно възприятие. Измежду тях са открити случаи с проблеми в цветното възприятие, които не са разпознати нито от ученика, нито от родителя, нито от учителя.

Mila Dragomirova, Snejana Iordanova. UNUSUAL PREVALENCE OF COLOR VISION DEFECTS IN CHILDREN FROM A SMALL BULGARIAN TOWN

Color perception as one of the important attribute of vision is investigated during the screening session of 203 school students between 6 and 18 years old, from a small town Sapareva Banya. The examination was performed through Ishihara pseudoisochromatic plates, Farnsworth D-15 test and Lanthony Desaturated D-15 test. In the interest of saving time during

---

*For contact:* Snejana Iordanova, Department of Optics and Spectroscopy, Faculty of Physics, Sofia University “St. Kliment Ohridski”, 5 J. Bourchier Blvd., 1164 Sofia, Bulgaria, Phone: +359 2 81 61 237, E-mail: snejana@phys.uni-sofia.bg

the screening the tests have been administered binocularly. According to Ishihara plates among the 87 boys and 116 girls, 9.2% and 4.31% had a congenital color defect, respectively. The statistically processed data from color vision tests have been compared with other known data from all over the world.

All students have never been tested for color perception. Among them there were found cases with lack of student, parental, and teacher recognition of presence of color perception problems.

***Keywords: anomalous*** color vision, pseudoisochromatic plates, arrangement test, Ishihara, Farnsworth D-15, Lanthony Desaturated D-15, school students screening.

***PACS numbers:*** 42.66.Ne

## 1. INTRODUCTION

Color vision, in particular in humans, is the ability to distinguish objects based on the wavelengths of the light they reflect, emit, or transmit. The light level is also important prerequisite to see the color. Normal human color vision, in daylight is trichromatic, based on the presence of three spectrally-distinct types of cone photoreceptors – red, green and blue. Their response to light in the retina depends on the wavelength of peak absorption for the different cone pigments (i.e. 560, 530 and 420 nm). When there is an alteration or absence of one of the receptor photopigments the person is characterized with a color vision deficiency (CVD). The latter could be congenital/inherited CVD or acquired. The first type CVD is nonpatological, incurable and it does not change over time. The most common is the red-green deficiency, which is inherited as an X-chromosome-linked recessive trait. Acquired color deficiency accompanies another condition (e.g. disease, trauma) or it is caused by the side effects of certain drugs, medications or exposure to chemical toxins. Acquired color deficiencies are classified as red-green and blue-yellow [1].

The people with inherited CVDs are classified as anomalous trichromats, dichromats, and monochromats. The anomalous trichromat is a relatively mild form of defective color vision. The terms protanomaly, deuteranomaly and tritanomaly are given when there is a defect in red, green and blue pigments, respectively. The dichromats base their color vision on only two pigments. The class of dichromats characterized by entire absence of green cones, red cones or blue cones is called deuteranopia, protanopia and tritanopia, respectively. Protanomaly and protanopia are referred to protan CVD whereas deuteranomaly and deuteranopia to deutan deficiency. Protan and deutan CVD are often described as red-green deficiency. Tritanomaly and tritanopia are described as the tritan or blue-yellow CVDs.

The lack of systematic early age color vision examination in Bulgaria suggests a large number of people affected by CVD to remain undetected as they simply adapt to the environment to certain extent and also because of unawareness of the disorder. Because of the role color coding plays in instructional materials used during the early school years, the color-deficient student may find some tasks difficult and, as a consequence, may develop a dislike for school and learning. Which in turn hinders the development of the pupil's potential. The student color vision status is also important when choosing a future profession. Currently, there is no treatment for congenital CVD, however the sooner the color vision deficiency can be identified the sooner accommodation can be made to help the child. Considering these aspects, there is an increasing interest across the world [2–6] in view of identifying the prevalence of color vision deficiencies in school children. The study aims to explore these deficiencies in school age children in the small Bulgarian town Sapareva Banya. During the screening session in case of color vision deficiency the parents of the child were informed so that it can be guided properly in his learning process as well as in choosing and adjusting with the suitable profession.

## 2. METHODES AND MATERIALS

### 2.1. PARTICIPANTS

The screening session of school students from Sapareva Banya took place in January 2017. A total of 203 students aged 6–18 years (both males and females) were evaluated for congenital color vision deficiency. The students were selected by purposive sampling method. Only these interested and intending to be evaluated were enrolled in the present study. Investigation of visual acuity preceded color testing.

### 2.2. STIMULI AND PROCEDURE

The screening of all students was done by using Ishihara 38 Plates Edition as according to instruction in the case of large scale examination the test was simplified to an examination of six plates only (№ 1, one of № 2–5 one of the № 6–9 one of № 10–13 one of № 14–17 and one of № 18–21). In the assessment of color appreciation by this short method involving 6 plates only if there was a discrepancy in any of the recordings, a number of plates was used before diagnosing deficiency. The plates were held about 75 cm from the child and tiled so that the plane of paper is at right angle to the line of vision. Time allowed to each plate did not exceed 5 seconds. In the case of subject unable to read numerals, plates 18–24 are used and the winding lines between the two X's are traced with a brush within ten seconds. While the Ishihara's test is designed to assess color vision deficiency of congenital origin the two other used tests Farnsworth D-15 and Lanthony Desaturated D-15, which are arrangement test, in addition give opportunity to diagnose acquired color vision deficiency. Manifestation of the latter is not expected in the studied group but as it is known [6] the color perception in blue-yellow region is changing due to both cognitive development and development of the color vision system. The Farnsworth D-15 test is called "dichotomous" [7] because it was designed to separate subjects into one of two groups: 1) strongly/medium color deficient or 2) mildly color deficient or color normal. The Lanthony D-15 test was designed also to separate patients into one of two groups: 1) mild color deficient or 2) normal color perception. The only difference between these items is the color saturation of the discs. According to the identical administration needs of both arrangement tests they were carried out on a black background to prevent surroundings from affecting the color perception by the students.

*The* tests were conducted at a working distance (25–30 cm) in about 2 minutes. The arrangement tests procedure stated with explanation of the task, according to [1]: "Find a cap from here (indicate the random array of the 15 moveable caps) that looks most like this cap (indicate the reference or pilot cap), and place it next to the fixed cap in the tray. Then find the cap that looks most like the one you already selected, and repeat this until you have placed all of the caps in the tray". Although the researcher checked after instruction that the participant understood the task condition the results of 5 from 150 and 3 of 94 examined students are removed from the Farnsworth D15 and Lanthony D15 analysis, respectively, because they imply poor task comprehension.

The tests administration was organized in a bright room without direct sunlight. In the interest of saving time during the screening all color tests were administered binocularly. The students have been wearing their refractive correction in case of proven need, and there was no tint in the glasses or contact lenses.

## 3. RESULTS AND DISCUSSION

In this study 203 students (boys 87, girls 116) are examined. The screening session started with a questionnaire comprising the following information about students: have they had some eye examination and do they have any prescribed glasses. The initial study of students' visual acuity divided them generally in two groups (Fig. 1). The first one contains 124 students, who have never been to an eye specialist. The results of the current screening showed that 91 of them (44.83%) still do not need glasses, while the other 33 (16.26%) were asked to visit an eye specialist for additional examination for prescription of glasses. The second group includes 79 students, who have been to an eye examination. From them, 48 in number did not need eye correction, while the other 31 needed corrections and carry glasses. The present visual acuity tests revealed that among the students who have attended before an eye specialist, 12 (5.91%) students without glasses need a correction and 20 (9.85%) students with glasses needed a new correction.

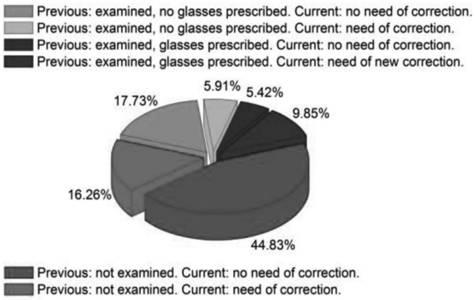


**Fig. 1.** Results from previous eye examinations of children from Sapareva Banya, compared with the current screening

### 3.1. ISHIHARA TEST

The subsequent study of color vision of the students, by Ishihara test, revealed that out of total 203 students, 13 (6.4%) were color deficient (Fig. 2a), according to Ishihara test plates. From them, 8 are boys and the other 5 are girls. The statistical treatment of gender-related outcomes (Fig. 2b) showed that among 87 boys and 116 girls, 8 (9.2%) and 5 (4.31%) were color deficient, respectively.

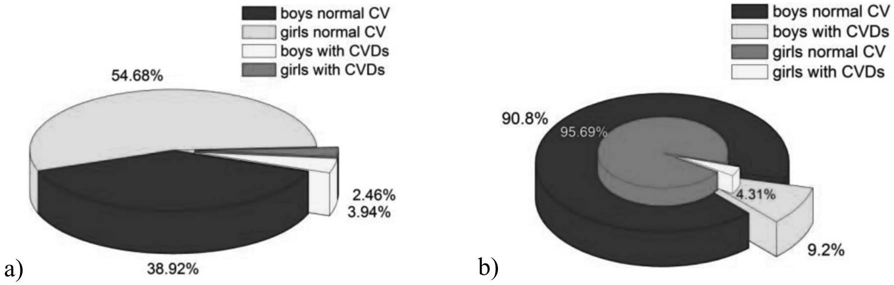


**Fig. 2.** Distribution of girls and boys with normal color vision (CV) and with color vision deficiency in (a) and gender-related percentages in (b)

According to the world statistic [8–10] about 8% of all men and about 0.5% of all women are suffering from congenital CVD. The variability of the latter depends also of the ethnic group. The higher prevalence in male population of CVDs provided the information given in Figure 3. In-depth review of the literature, however, shows that authors report for both higher male Asian (11.05%) [5] and female (0.83%) [4], (2.42%) [5] CVDs. However, no data for such extremely high prevalence of CVD in female population as this established in Sapareva Banya are reported.

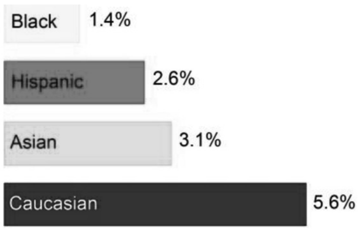


**Fig. 3.** Prevalence of color vision deficiency in boys, by ethnicity [4]

The distribution of number of students with normal color vision and with CVD on age is shown in Figure 4a. Higher number of students with CVDs was established in the most numerous age groups of the 8 year olds. The latter consisted of 3 boys and 1 girl (Fig. 4b) with CVDs. In the groups of 9 and 10 year olds the number of boys and girls with CVD was the same. According to world data for gender-related prevalence of CVD the existence of age groups with only male representatives is not surprising. This is not the case, however of the age groups (see Fig. 4b), 12 and 17 year olds with only female representatives with CVD. The most likely cause was that in these groups the number of boys was 6 and 2 in 12 and 17 year olds, respectively.

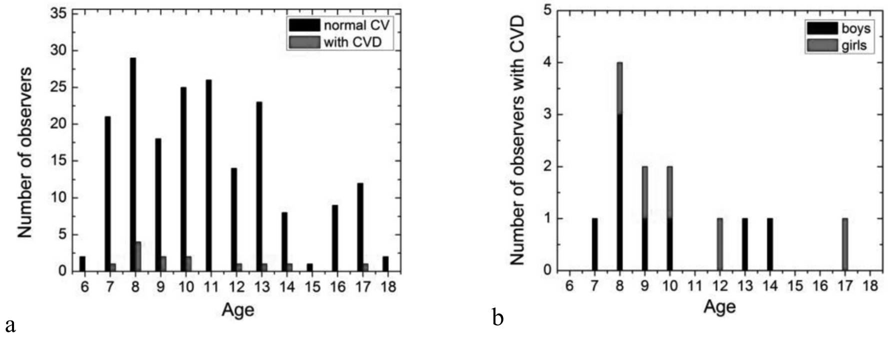


**Fig. 4.** Distribution of the students with normal color vision and with CVD on age in (a) and gender-related number of students in (b)

### 3.2. FARNSWORTH D-15 TEST

**Table 1.** Comparison between Ishihara and Farnsworth D-15 tests for color vision deficiency. Values are number of students

| Color Test | | Pass Ishihara | Fail Ishihara | Total |
|---|---|---|---|---|
| Pass D15 | | 86 | 4 | 90 |
| Fail D15 | red-green | 7 | 7 | 14 |
| | blue-yellow | 23 | 0 | 41 |
| | undefined | 16 | 2 | |
| Total | | 132 | 13 | 145 |

The color vision of 145 students out of 203, surveyed with the Ishihara test, was investigated in addition with the Farnsworth D-15 test. The results of both color vision tests were summarized in Table 1. Out of all 13 students who failed the Ishihara test, 7 failed also on Farnsworth D-15, but 4 successfully passed D-15 and another 2 have made mistakes on D-15 which did not allow determining accurately the CVD. Illustrations of some of the students' results with red-green CVD according to Ishihara and Farnsworth D-15 tests are shown on Figure 5. Typical distribution of caps arrangement of deutan (Fig. 5a) and protan (Fig. 5b) CVD was established in two of boys examined. The result of D-15, shown on Fig. 5c, reveals protanomalous CVD of a girl.

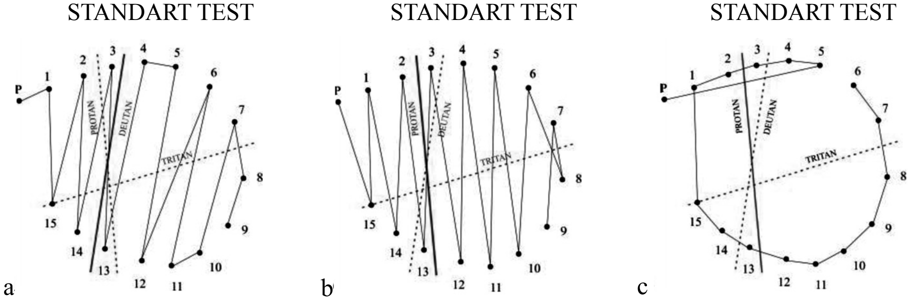


**Fig. 5.** Examples of Farnsworth D-15 test arrangements in the case of deutan (a), protan (b), and protanomalous (c) CVD

Among those who passed successfully Ishihara test plates, 132 in number, 86 passed also D-15 but the other 46 failed on it. The distribution of CVDs between them is shown on Figure 6. Another 7 (15.22%) students had red-green CVD, 23 (50%) – blue-yellow CVD (Fig. 7a) and 16 (34.78%) in number, have made mistakes on D-15 which did not allow determining accurately the CVD (Fig. 7b).

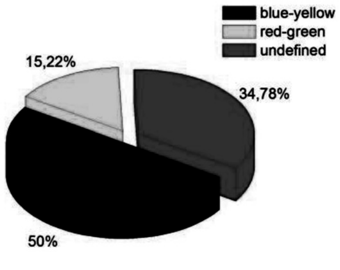


**Fig. 6.** Distribution of the students with color vision deficiencies who failed Farnsworth D-15 test but passed successfully Ishihara test plates

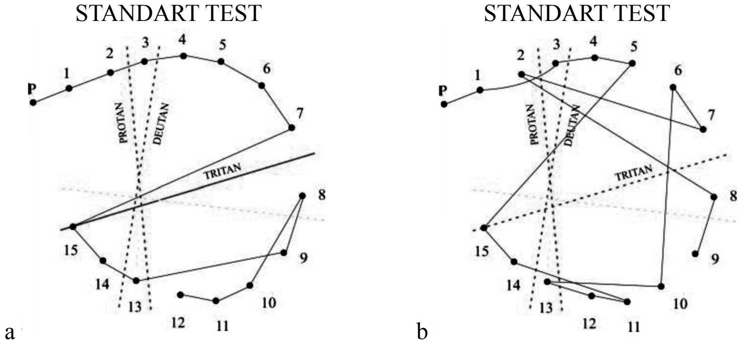


**Fig. 7.** Examples of Farnsworth D-15 test arrangements in the case of blue (a), and undefined (b) CVD

According to Ishihara test (Table 2) 8 boys were with red-green CVD, 4 of them have shown results on D-15 that confirm this CVD, 1 was classified with undefined CVD and another 3 – as such with normal color vision with respect to D-15. The situation in girls was quite different. 5 girls were selected as such with red-green CVD according to Ishihara, 3 of them also failed on D-15, 1 was classified with undefined CVD and another girl passed without any problems D-15. The results of D-15 showed, however, that among those girls who successfully passed Ishihara, 7 failed on it. So, according to Farnsworth D-15 test (Fig. 8) among 62 boys and 83 girls, 4 (6.45%) and 10 (12.05%) were with red-green color vision deficient, respectively. The results of D-15 revealed an extremely high prevalence of red-green CVD in the young female population in Sapareva Banya which even exceeded the world male trend. Age distribution of red-green CVD in young male and female population according to Farnsworth D-15 is given in Figure 9. The results revealed also that the blue-yellow CVDs exceed those of red-green deficiencies. The distribution of the students with those CVDs on age could be traced out in Table 3.

**Table 2.** Comparison between gender-related color vision deficiencies from Ishihara and Farnsworth D-15 tests. Values are number of students

| Color Test | | Boys Ishihara | | Girls Ishihara | | Total |
|---|---|---|---|---|---|---|
| | | Pass | Fail | Pass | Fail | |
| Pass D15 | | 0 | 3 | 0 | 1 | 4 |
| Fail D15 | red-green | 0 | 4 | 7 | 3 | 14 |
| | blue-yellow | 0 | 0 | 0 | 0 | 0 |
| | undefined | 0 | 1 | 0 | 1 | 2 |
| Total | | 8 | | 12 | | 20 |

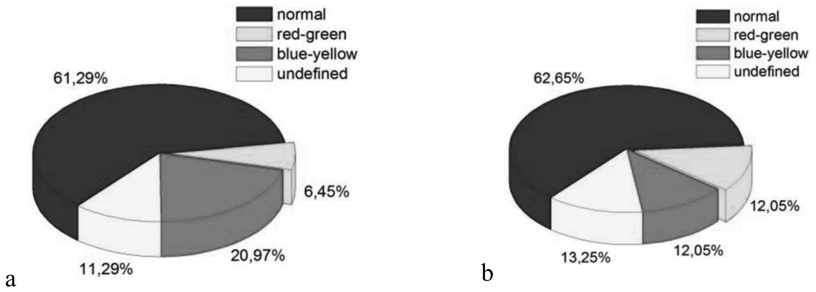


**Fig. 8.** Distribution of boys (a) and girls (b) with normal color vision and with color vision deficiency according to D-15

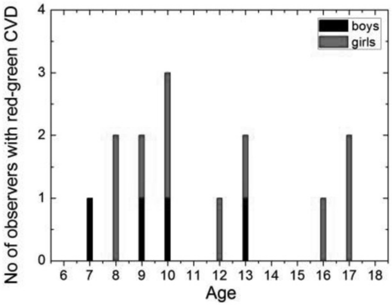


**Fig. 9.** Age distribution of the examined boys and girls with CVDs according to Farnsworth D-15 test

**Table 3.** Distribution of color vision deficiencies on age according to Farnsworth D-15 test

| Age | Red-green | | Blue-yellow | | Undefined | |
|---|---|---|---|---|---|---|
| | Boys | Girls | Boys | Girls | Boys | Girls |
| 7 | 1 | - | 1 | - | - | - |
| 8 | - | 2 | 5 | 3 | 2 | 2 |
| 9 | 1 | 1 | 4 | 1 | - | 2 |
| 10 | 1 | 2 | 1 | 2 | 3 | 2 |
| 11 | - | - | - | 1 | 1 | 1 |
| 12 | - | 1 | 1 | - | 1 | - |
| 13 | 1 | 1 | 1 | - | - | 2 |
| 14 | - | - | - | 1 | - | 1 |
| 16 | - | 1 | - | - | - | - |
| 17 | - | 2 | - | 2 | - | 1 |
| Total | 14 | | 23 | | 18 | |

### 3.3. LANTHONY DESATURATED D-15 TEST

The color vision of 91 students out of 203, surveyed with the Ishihara test, was investigated also with the Lanthony Desaturated D-15 test. The results of both color vision tests are summarized in Table 4. Out of 5 students who failed the Ishihara test, 2 failed also on Lanthony Desaturated (LD) D-15, 2 successfully passed LD D-15 and another 1 made blue-yellow mistakes on LD D-15.

**Table 4.** Comparison between Ishihara and Lanthony Desaturated D-15 tests for color vision deficiency. Values are number of students

| Color Test | | Pass Ishihara | Fail Ishihara | Total |
|---|---|---|---|---|
| Pass LD D15 | | 42 | 2 | 44 |
| Fail LD D15 | red-green | 4 | 2 | 47 |
| | blue-yellow | 32 | 1 | |
| | undefined | 8 | 0 | |
| Total | | 86 | 5 | 91 |

Among those who passed successfully Ishihara test plates, 86 in number, 42 passed also LD D-15 but the other 44 failed on it. The distribution of CVDs between them is shown on Figure 10. Another 4 (9.09%) students had red-green CVD, 32 (72.73%) – blue-yellow CVD and 8 (18.18%) in number, have made mistakes on LD D-15 which did not allow determining accurately the CVD.

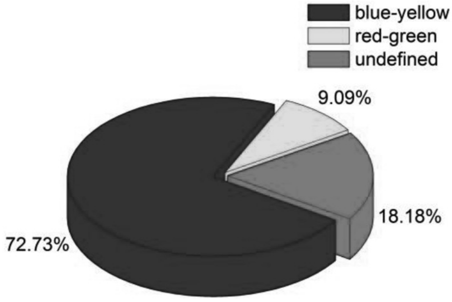


**Fig. 10.** Distribution of the students with color vision deficiencies who failed Lanthony Desaturated D-15 test but passed successfully Ishihara test plates

**Table 5.** Comparison between Farnsworth D-15 and Lanthony Desaturated D-15 tests for color vision deficiency. Values are number of students

| Color Test | | Pass D15 | Fail D15 | | | Total |
|---|---|---|---|---|---|---|
| | | | red-green | blue-yellow | undefined | |
| Pass LD D15 | | 22 | 1 | 1 | 0 | 24 |
| Fail LD D15 | red-green | 0 | 6 | 0 | 0 | 42 |
| | blue-yellow | 20 | 0 | 3 | 5 | |
| | undefined | 5 | 0 | 2 | 1 | |
| Total | | 47 | 19 | | | 66 |

The results of 66 students, to whom both arrangement tests Farnsworth D-15 and Lanthony Desaturated D-15 were performed, are summarized in Table 5. It is important to note that from the 6 students diagnosed with red-green CVD according to D-15 and LD D-15 only 2 of them failed on Ishihara. The other 4 students were part of those (7 in number, see Table 2) which failed on D-15 and have been checked with both arrangement tests. They were all girls. The confirmation of results from two independent color tests for more girls with red-green CVD than those defined according to Ishihara test plates showed that the prevalence of red-green CVD in young female population in Sapareva Banya is at least as much as this indicated from Ishihara test. The results in Table 5 reveal that among those students who passed successfully D-15 almost half of them have shown on LD D-15 the presence of milder blue-yellow CVD. This could be traced out and from the results in Table 1 (D-15) and Table 4 (LD D-15), where 23(15.86%) students of 145 and 33(36.26%) students of 91, respectively, have shown blue-yellow CVDs. The changes in blue-yellow perception on the one hand are associated with acquired CVD due to some systematic or ocular disease, toxicity, or trauma on the other hand are associated with the immature visual system of tested students which have none of the reported symptoms. The trend in the development of the color vision system suggests fewer students with stronger blue-yellow CVD compared to those with milder CVD. This expectation is also confirmed by the research, 15.86% with more pronounced CVD in blue-yellow region versus 36.26% with less pronounced CVD in the same region. According to results from LD D-15 the blue-yellow CVDs exceed those of red-green deficiencies. The distribution of the students with those CVDs on age could be traced out in Table 6. The precise evaluation of CVD distributions over ages needs additional statistical data.

**Table 6.** Distribution of color vision deficiencies on age according to Lanthony Desaturated D-15 test

| | Red-green | | Blue-yellow | | Undefined | |
|---|---|---|---|---|---|---|
| Age | Boys | Girls | Boys | Girls | Boys | Girls |
| 8 | - | - | - | 1 | - | - |
| 9 | - | 1 | - | 1 | 1 | - |
| 10 | - | - | - | 2 | - | 1 |
| 11 | - | - | 4 | 1 | - | 1 |
| 12 | - | 1 | 1 | 4 | 1 | 1 |
| 13 | - | 1 | 6 | 4 | 1 | - |
| 14 | - | - | - | 3 | 1 | - |
| 16 | - | 1 | - | 2 | - | - |
| 17 | - | 2 | - | 3 | - | 1 |
| 18 | - | - | - | 1 | - | - |
| Total | 6 | | 33 | | 8 | |

Most of the students with established during the screening session CVD had not previously recognized its presence. This is somewhat understandable as they were born with it. It was surprising, however, that parents who accompanied their children were really surprised to find out that their child had a color vision deficiency even if it was from the strongest red-green type – deutan and protan (see Figs. 5a,b).

## 4. CONCLUSIONS

In the beginning of 2017, the first of its kind screening session of the color vision of students from a small Bulgarian town was held. The color perception of 203 school students between 6–18 years old was examined through three color tests: Ishihara pseudoisochromatic plates, Farnsworth D-15 test and Lanthony Desaturated D-15 test. According to Ishihara plates among the 87 boys and 116 girls, 8 (9.2%) and 5 (4.31%) had a red-green color vision deficiency, respectively. The world statistic reported about 8% of all men and about 0.5% of all women are suffering from congenital red-green color vision deficiency as the variability of the latter depends also of the ethnic group. The established male percentage in the study correlated with the well-known data, however the extremely high percentage in female population exceeded even the highest found (2.42%) in the literature [5]. According to Farnsworth D-15 test among 62 boys and 83 girls, 4 (6.45%) and 10 (12.05%) were found to have red-green color vision deficient, respectively. The results of D-15 for the girls' problems in color perception even exceeded the world male trend. A part of the girls with CVD established from D-15 girls was tested with the Lanthony Desaturated D-15 test which confirmed the result of D-15. The studies done revealed that the prevalence of the red-green color vision deficiency in female population in Sapareva Banya is not lower than that established by the Ishihara test plates.

The observed high prevalence in the blue-yellow perception deficiency will be subject of further study. At least partially it may be explained by the fact that at school age the cognitive abilities and color vision system are still forming.

### REFERENCES

[1] Benjamin, Borish's Clinical Refraction, Elsevier Inc., 2006.

[2] Ramachandran, N., Wilson, G. A., Wilson, N. *Clin. Exp. Opt.*, 2014, **97**, 499.

[3] Mulusew, A., Yilikal, A *J. Ophtalmol. East. Central and Southern Africa*, August 2013, **17**, 10..

[4] Fareed, M., Anwar, M. A., Afzal, M. *Genes&Diseases*, 2015, **2**, 211.

[5] Sushil, K., Mandira, M. *J. Mol. Genet. Med.*, 2017, **11**, 1000306.

[6] Ling, B. Y., Dain, S. J. *Vis. Neurosci.* 2008, **25**, 441.

[7] Farnsworth D-15 and Lanthony Test Instructions, P/N 910315, Rev 1.7 (05/06).

[8] https://visiontechnology.co/statistics/

[9] https://nei.nih.gov/health/color_blindness/facts_about

[10] http://www.color-blindness.com/2006/04/28/colorblind-population/

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# СИНЯТА СВЕТЛИНА В ЛАЗЕРНИТЕ ТЕХНОЛОГИИ – ВЪЗМОЖНОСТИ И ПРЕДИЗВИКАТЕЛСТВА

НАСКО ГОРУНСКИ, МАЯ ЖЕКОВА

*Лаборатория по лазерна техника, Физически факултет*

*Наско Горунски, Мая Жекова*. СИНЯТА СВЕТЛИНА В ЛАЗЕРНИТЕ ТЕХНОЛОГИИ – ВЪЗМОЖНОСТИ И ПРЕДИЗВИКАТЕЛСТВА

Направено е сравнение на въздействието на синята и инфрачервената светлина върху различни материали. Теоретично са получени температурата на повърхността и нейното изменение в дълбочината на материал, облъчен с лазер. Нагряването в зоната на въздействие е измерено с помощта на инфрачервена камера.

*Nasko Gorunski, Maya Zhekova*. BLUE LIGHT IN LASER TECHNOLOGY – POSSIBILITIES AND CHALLENGES

The blue and infrared light influences on different materials are compared. The temperature on the surface and its change in depth of material are numerically calculated. Measurements of heating of the materials are made with infrared camera.

***Keywords:*** semiconductor lasers, laser diodes, laser technology, industrial applicationson
***PACS numbers:*** 42.55.Px , 42.62.Cf

---

*За контакти*: Наско Горунски, ЛЛТ, Физически факултет, Софийски университет „Св. Климент Охридски“, ул. Галичица 33А, София 1164, телефон: +359 2 8161 751, E-mail: naskog@phys.uni-sofia.bg

# 1. УВОД

В основата на всички съвременни лазерни технологии е взаимодействието на светлината с различните вещества. Механизмите са добре известни – поглъщане, отражение и преминаване. Ефективността на взаимодействие зависи както от източника на светлина (лазер), така и от оптичните параметри на средата. В зависимост от оптичните си свойства материалите се делят на прозрачни, полупрозрачни и непрозрачни. Металите, към които се прилагат най-често лазерните технологии, са непрозрачни за видимата светлина, а вълната се поглъща в много тънък повърхностен слой.

В общия случай разпространението на електромагнитна вълна може да се опише теоретично чрез уравненията на Максуел [1]. Оптичните константи могат да се получат от комплексния израз за диелектричната проницаемост. От инженерна гледна точка за пресмятането на тези константи е добре да се избягва използването на сложната електронна теория. Има нужда от приблизителни формули, които да свържат оптичните параметри с добре известни параметри на материала, например електричната проводимост. Поглъщането на светлината при облъчване с лазер на различни вещества е изследвано в [2–5]. Резултатите показват, че металите поглъщат много по-добре вълна с дължина 500 nm, отколкото тази с дължина 1064 nm, и особено с дължина $10^4$ nm. По-доброто поглъщане би довело до по-бързо нарастване на температурата в метала, което от своя страна би позволило по-ефективна обработка.

Доскоро единствените кохерентни източници в синята част на спектъра бяха газови лазери – аргонови, криптонови и ексимерни, макар че последните по-скоро генерират ултравиолетова светлина. Експлоатацията на всички тези източници е скъпа. Алтернатива на тези лазери относно дължината на вълната са полупроводниковите лазери, базирани на GaN, които генерират светлина с дължина на вълната около 445 nm. Технологичните трудности около производството им се преодоляват и те стават все по-достъпни и надеждни.

В нашето изследване беше използван полупроводников лазер, генериращ на 445 nm, който е съставна част от гравираща система.

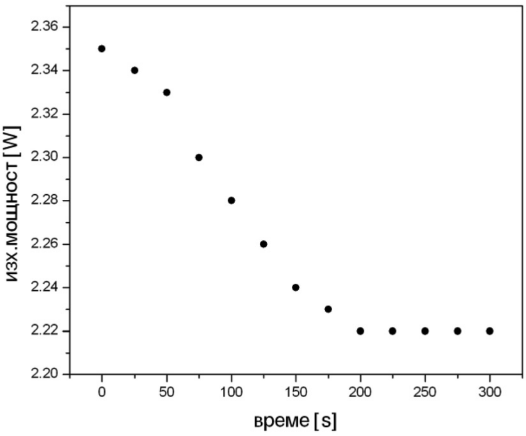
# 2. ИЗСЛЕДВАНЕ НА ОПТИЧНОТО ЛЪЧЕНИЕ

Изследваните характеристики на синия лазер, имащи отношение към възможните технологични приложения са:

1. Дължина на вълната на генерация – 445 nm.

2. Изходна мощност – стабилността на тази характеристика във времето е от съществено значение за лазерните технологии. Промяната или случайните флуктуации биха намалили ефективността на въздействие на светлината върху съответния материал. При измерването на мощността на използвания

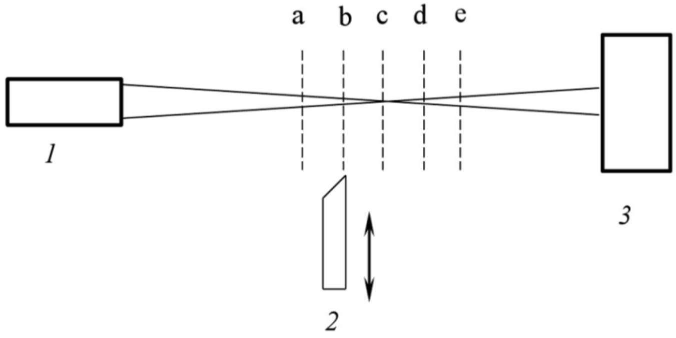
лазер беше установено, че тази, която фигурира в техническите данни като максимална, след включването на лазера пада и след известно време се стабилизира (фиг. 1).



**Фиг. 1.** Изменение на изходната мощност във времето

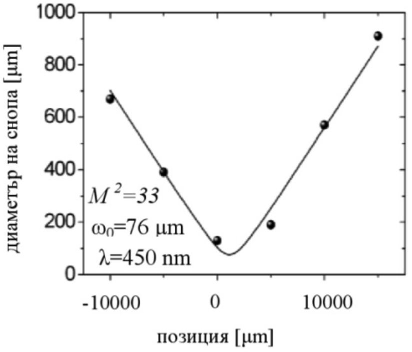
Ефектът най-вероятно се дължи на липсата на добро охлаждане. Характеристиките на полупроводниковите лазери се поддържат постоянни чрез охлаждане с пелтие-елементи, които ефективно отнемат топлината от структурата. В случая диодът се охлажда въздушно и изминава известно време, преди температурата му да се стабилизира.

3. $M^2$-фактор. Определяне на оптичното качество на снопа, т. нар. $M^2$-фактор, има отношение към възможността за фокусиране на лазерния сноп в петно с най-малки размери. Доброто фокусиране повишава качеството на среза при лазерно рязане, дава възможност за фина обработка на облъчвания материал, намалява термичното натоварване на материала около зоната на облъчване. Един от методите за определяне на този параметър включва използването на събирателна леща, с чиято помощ лазерната светлина се фокусира. Определя се размерът на снопа в множество равнини преди и след шийката на снопа. Под размер (диаметър) се разбира разстоянието между точките, в които мощността (в случая) е 90 и 10% от максималната мощност. Данните позволяват да се направи хиперболичен фит и от параметрите на хиперболата се получава $M^2$-факторът. В случая фокусното разстояние е 8 cm, а измервания са направени в 5 равнини. Схемата на опитната постановка е показана на фиг. 2.

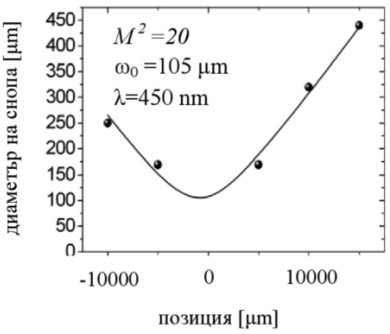


**Фиг. 2.** Схема на експерименталната установка за измерване на $M^2$-фактора на изследвания лазер: *1* – син лазер, *2* – острие, *3* – измерител на мощност; a, b, c, d, e – равнини, в които е измерен размерът на лазерния сноп

Лазерният сноп, излизащ от полупроводниковия лазер, има правоъгълно сечение, което означава, че този фактор ще е различен по двете оси. Резултатите са показани на следващите две фигури (фиг. 3, 4).



**Фиг. 3.** Резултати от измерванията по бързата ос и апроксимацията им с хипербола



**Фиг. 4.** Резултати от измерванията по бавната ос и апроксимацията им с хипербола

Стойностите на този фактор не са малки, но въпреки това се вижда, че около фокуса размерите на снопа са ~100 µm. Предимството на късата вълна е, че при фокусиране диаметърът на снопа във фокуса ще бъде около 2 пъти по-малък в сравнение с този на инфрачервената светлина, фокусирана с леща със същото фокусно разстояние.

## 3. ЧИСЛЕНО ПРЕСМЯТАНЕ НА ТЕМПЕРАТУРАТА И НЕЙНОТО ИЗМЕНЕНИЕ В ДЪЛБОЧИНА ЗА РАЗЛИЧНИ МЕТАЛИ

При облъчване на твърдо тяло със светлина част от нея се поглъща и се превръща в топлина. Тази топлина може да се разглежда като вторичен топлинен източник, който определя както разпределението на температурното поле в материала, така и динамиката на изменението на температурата. Решаването на задачата за нагряването изисква познаване както на оптичните и топлинните свойства на телата, така и на пространственото и времевото разпределение на интензитета на лазерното лъчение. В общия случай задачата няма аналитично, а само числено решение. Разпределението на температурата в тялото $T = T(x,y,z,t)$ се дава с уравнението

$$\nabla^2 T - \frac{1}{k}\frac{\partial T}{\partial t} = -\frac{A(x,y,z,t)}{K},$$

където $A(x,y,z,t)$ е количество топлина, отделена в единица обем за единица време; $k$ – коефициент на температуропроводност [$m^2/s$]; $K$ – коефициент на топлопроводност [W/m.K].

Металите са широко използвани в различни области на човешкия живот, свойствата им са добре проучени и в литературата лесно могат да се намерят данни за оптичните им константи. Ето защо числените пресмятания са направени само за метали и се базират на резултатите, получени в [6, 7].

От особен интерес е изменението на температурата в точката на облъчване. Това изменение беше пресметнато за едни и същи материали под въздействието на две различни дължини на вълната – 445 nm и 1064 nm. Това са дължините на вълните на синия полупроводников лазер и съответно на Nd:YAG лазер, един от най-често използваните технологични инфрачервени лазери.

Данните, които се използват при численото решаване са:

- мощност на лазера – 0.8 W;
- диаметър на снопа във фокуса (за опростяване на задачата приемаме, че петното във фокуса има кръгла форма) – 100 μm;
- време на облъчване – лазерът работи в непрекъснат режим и времето, за което се облъчва материалът, попадащ във фокалното петно, се определя от скоростта, с която се движи лазера. Времето се определя, като диаметърът на снопа се раздели на скоростта на движение. Скоростта може да се променя в границите от 0.0167 до 1.67 cm/s. Времената, които отговарят на скорости на движение 0.0167 cm/s, 0.833 cm/s и 1.67 cm/s, са дадени в табл. 1.

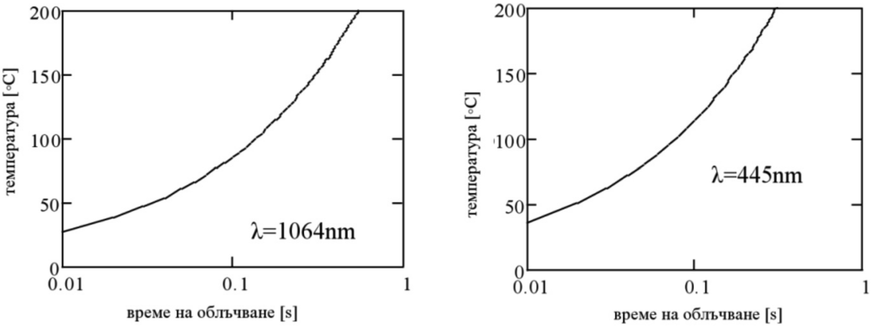
**Таблица 1.** Времена на облъчване при различни скорости на движение на лазерния източник

| Скорост [cm/s] | Време [s] |
|---|---|
| 0.0167 | 0.6 |
| 0.833 | 0.012 |
| 1.67 | 0.006 |

Връзката между температурата и времето на облъчване има вида

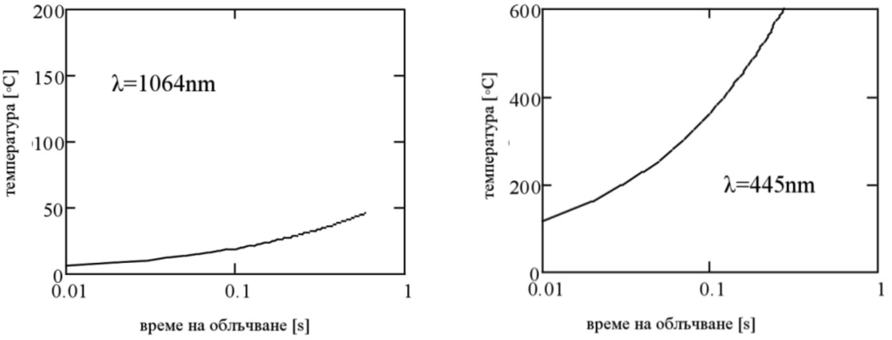
$$T = \frac{2P_a}{r_0^2 K}\sqrt{\frac{kt}{\pi^3}},$$

$P_a$ е погълнатата мощност от веществото; $P_a = (1 - R)P$, където $P$ е изходната мощност на лазера, а $R$ е коефициентът на отражение на веществото за съответната дължина на вълната. При тези допускания и ограничения беше проследено изменението на температурата в точката на облъчване за два метала: алуминий и мед, при облъчване със синя (445 nm) и инфрачервена светлина (1064 nm). Данните за съответните константи са взети от литературата. Резултатите за алуминий за показани на фиг. 5а, б, а за медта – на фиг. 6а, б.



а б

**Фиг. 5.** Изменение на температурата в зоната на облъчване за алуминий при: а) λ = 1064 nm, б) λ = 445 nm



а б

**Фиг. 6.** Изменение на температурата в зоната на облъчване за мед при: а) λ = 1064 nm, б) λ = 445 nm

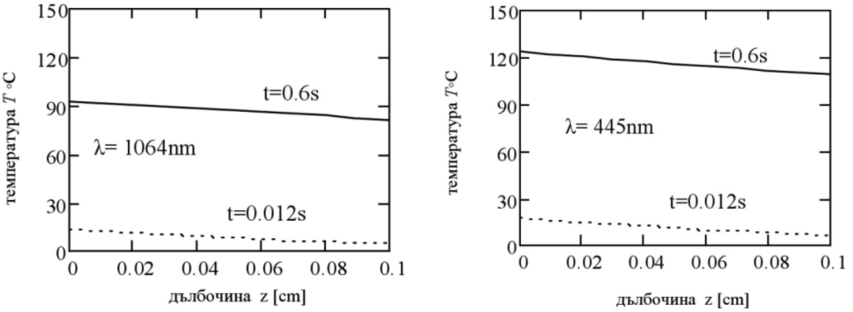
Двата метала показват значително по-добро поглъщане на синята светлина, което води до по-висока температура при едно и също време на облъчване. Може да се очаква, че обработката на метали със светлина от синята част на спектъра ще е по-ефективна от тази с инфрачервена светлина.

От съществено значение за практиката е и изменението на температурата в дълбочина на обработвания детайл. В този случай отново се използва уравнението за топлопроводност със съответните допускания и опростяващи предположения. В случай на повърхнинен източник, полубезкрайна среда и равномерно разпределение на интензитета в равнината на падане температурата *T* на определена дълбочина *z* се намира чрез интегриране по времето чрез формулата:

$$T(z,t) = \frac{1}{K}\sqrt{\frac{k}{\pi}}\int_0^t \frac{Ia}{\sqrt{\tau}} exp\,(\frac{-z^2}{4k\tau})d\tau.$$

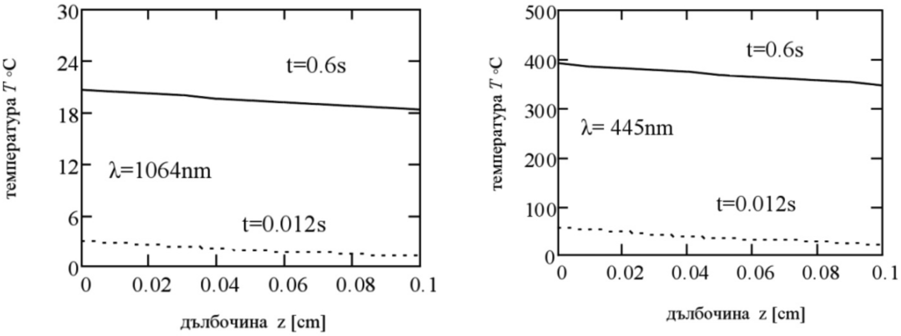
Интегрирането по времето е направено за две времена, които отговарят на скорости на движение 0.0167 cm/s и 0.833 cm/s. Приехме, че разпределението на интензитета във времето е правоъгълно, т.е. падащият интензитет е концентриран в точката, над която се намира лазера, а след като той се премести, светлина в тази точка няма. Пресмятанията отново са направени за две дължини на вълните, дебелина на материала *z* = 0.1 cm и два метала – алуминий и мед.

Получените резултати (алуминий – фиг. 7а,б; мед – фиг. 8а,б) отново показват, че синята светлина се поглъща по-ефективно от инфрачервената. Разликата е особено голяма при медта, която има много голям коефициент на отражение за вълната с дължина 1064 nm. Практиката доказва този факт – лазерното рязане на медни детайли изисква сериозна мощност от инфрачервените лазери. Това е наложително, за да се нагрее материалът и да намалее неговият коефициент на отражение. Този ефект е характерен за металите – намаляване на отражението с повишаване на температурата.



а б

**Фиг. 7.** Изменение на температурата в дълбочина за алуминий при облъчване с: а) λ = 1064 nm, б) λ = 445 nm



а б

**Фиг. 8.** Изменение на температурата в дълбочина за мед при облъчване с: а) λ = 1064 nm, б) λ = 445 nm

Числени премятания на температурата на повърхността и в дълбочина за неметалните материали, в случая – дърво и гума, не са направени поради недостатъчните данни в литературата. Данни за коефициентите на топло-проводност и температуропроводност могат да бъдат открити, но липсват всякакви такива за коефициентите на отражение за използваните дължини на вълните. Допълнително затруднение е фактът, че и двата материала са твърде разнообразни и с различни физични свойства.

## 4. ЕКСПЕРИМЕНТАЛНИ РЕЗУЛТАТИ

При експеримента бяха използвани два лазера – син полупроводников и диодно напомпван Nd:$YVO_4$. Синият полупроводников лазер е модел EL01-2.5W на фирмата EleksMaker, а инфрачервеният е лабораторен прототип. Целта е да се оцени и сравни въздействието на двете дължини на вълната върху различни материали – метали и неметали. Максималната мощност на синия лазер е 2.45 W, а на другия лазер – около 1 W. Измерванията бяха проведени при една и съща изходна мощност – 0.8 W. Мощността на синия лазер беше редуцирана чрез отражателни и абсорбционни филтри. Двата лазера работят в непрекъснат режим. Бяха облъчени алуминиева и медна пластина, гума и дърво. Времената на облъчване са 5 s и 10 s. Температурата в зоната на облъчване е регистрирана чрез инфрачервена камера FLIR i3 с параметри :

- чувствителност към топлина – < 0.15 ºC;
- работен спектрален диапазон – (7.5 ÷ 13) μm;
- обхват на измервана температура: от –20 ºC до + 270 ºC;
- точност: ± 2 ºC.

Регистрираните максимални температури са дадени в табл. 2.

**Таблица 2.** Измерени температури в зоната на въздействие

| Материал | 445 nm | | 1064 nm | |
|---|---|---|---|---|
| | 5 s | 10 s | 5 s | 10 s |
| Al | 21.5 ºC | 21.9 ºC | 25.9 ºC | 27.4 ºC |
| Cu | стайна ºC | 22.6 ºC | 25.6 ºC | 25.9 ºC |
| Дърво | 270 ºC | 238 ºC | 36.7 ºC | 53.3 ºC |
| Гума | 150 ºC | 138 ºC | 77.9 ºC | 133 ºC |

Температурата в стаята с инфрачервения лазер беше 25 ºC, а в тази със синия – 20 ºC. Данните показват, че под действието на синята светлина, температурата на неметалите достига много по-големи стойности, отколкото при инфрачервената. Поглъщането е много по-силно при 445 nm. Вижда се, че с увеличаване времето на облъчване максималната температура в зоната на въздействие при неметалите намалява. Предполагаме, че това се дължи на овъгления горен слой, който започва да отразява част от лъчението и пречи на нарастване на температурата. В потвърждение на това е фактът, че при максимална мощност синият лазер успява да среже лист хартия (за принтер), но не може да реже по-дебели листа. Проблемът може би ще се отстрани, ако срезът се обдухва с газ, който да премахва овъгления слой.

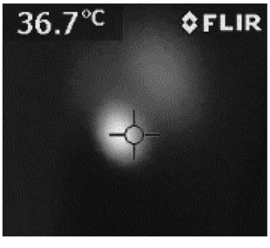
При металите се наблюдава разминаване между теоретично пресметнатата температура в точката на въздействие и тази, измерена с термокамерата при експеримента. Причините според нас са следните:

– Разлика между действителния размер на снопа във фокуса и приетия от нас въз основа на хиперболичния фит размер (фиг. 3, 4). С цел опростяване на пресмятанията приехме, че петното е кръгло с диаметър 100 μm. В действителност то си остава с правоъгълно сечение, площта е голяма и плътността на енергията намалява, съответно и достигнатата температура също намалява.

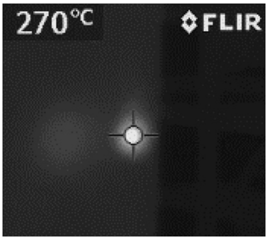
– Времената, за които се облъчва материалът (действителни) при обработката му, са посочени в табл. 1. Те са по-малки от 1 s и за експериментаторът е невъзможно за това време да включи и изключи лазера и да отчете температурата. Необходимо е допълнително оборудване, което не е налично към момента. Затова времената на облъчване са избрани 5 s и 10 s. При тях ефект оказва и топлопроводността на металите, която за алуминий и мед не е малка, а това води до разсейване на топлината и температурата пада.

Експерименталните данни не показват съществена разлика и за двата метала и не може да се каже коя дължина се поглъща по-добре.  Ефектът би се проявил при достатъчно голяма мощност на синия лазер – 100 W. Тази стойност е приблизителна – инфрачервените лазери с подобна мощност режат метали. Повишаването на мощността на лазерните източници, генериращи в синята част на спектъра, би позволило да се използва предимството на синята светлина (по-силно поглъщане) при обработката на металите, които са найчест обект на лазерните технологии. Направени са стъпки в тази посока [8].

На фиг. 9 ($\lambda$=1064 nm) и фиг.10 ($\lambda$=445 nm) са показани резултатите от измерване на температурата с инфрачервената камера за едно и също време за дърво.



**Фиг. 9.** Температура при облъчване с инфрачервена светлина



**Фиг. 10.** Температура при облъчване със синя светлина

## 4. ЗАКЛЮЧЕНИЕ

Теоретичните и експерименталните резултати показват ясно предимствата на синята пред инфрачервената светлина. При сегашните нива на изходна мощност и режим на работа (непрекъснат) сините лазери биха имали приложения в лазерните технологии, които не са свързани с пробиване и рязане, и то на неметали. Повишаването на изходната мощност и реализирането на импулсен режим биха разширили областта на тяхното приложение.

**Благодарности.** Настоящето изследване е осъществено с финансиране от ФНИ на Софийския университет и със съдействието на катедра „Квантова електроника“.

ЛИТЕРАТУРА

[1] M. Born and E. Wolf. Principles of Optics. Oxford, 1986.
[2] Y. Kawahito, N. Matsumoto, Y. Abe, S. Katayama. *Weld. Int.,* 2013, **27**, 129.
[3] Q. Nguyen, C.Y. Yang. *Int. J. Heat Mass Transfer*, 2016, **102**, 1222.
[4] D. Bergström. The absorption of laser light by rough metal surfaces, Ph.D. Thesis (Luneå University of Technology, 2008).
[5] F. Dausinger, J. Shen. *ISIJ International*, 1993, **33**, 9, 925.
[6] Carslaw H.S, Jeager J.C. Conduction of Heat in Solids. London, 1959.
[7] Duley, W.W. Laser Processing and Analysis of Materials. Plenum Press, New York, 1983.
[8] H. Wang, Y. Kawahito, R. Yoshida, Y. Nakashima, K. Shiokawa. *Opt. Lett.*, 2017, **42**, 12, 2251.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЗИЧЕСКИ ФАКУЛТЕТ
Том 111

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHYSICS
Volume 111

# PHOTOPOLARIMETRICAL STUDY OF BLAZAR-TYPE AGN OJ287 IN 2012-2015 WITH THE 2M RCC TELESCOPE AT NAO ROZHEN

VLADIMIR BOZHILOV[1], EVGENI OVCHAROV[1], MILEN MINEV[1,2,3] ,
YORDAN DARAKCHIEV[1], ANGEL DIMITROV[1], STEFAN GEORGIEV[1,2],
MANOL GERUSHIN[1], BORISLAV SPASSOV[1,2], KALINA STOIMENOVA[1]

*[1] Department of Astronomy*
*[2] Institute of Astronomy and National Astronomical Observatory, BAS*
*[3] Institute for Nuclear Researches and Nuclear Energy, BAS*

*Владимир Божилов, Евгени Овчаров, Милен Минев, Йордан Даракчиев, Ангел Димитров, Стефан Георгиев, Манол Герушин, Борислав Спасов, Калина Стоименова.* ФОТОПОЛЯРИМЕТРИЧНО ИЗСЛЕДВАНЕ НА БЛАЗАРА OJ287 В ПЕРИОДА 2012–2015 г. С 2М RCC ТЕЛЕСКОП НА НАО-РОЖЕН

Представяме резултатите от фотополяриметрично изследване на блазара OJ287 в периода ноември 2012–януари 2015 г. Налюденията са извършени с помощта на фокалния редуктор FoReRo-2 на 2-метровия RCC телескоп на Националната астрономична обсерватория (НАО) Рожен. Наблюдаваната промяна в позиционния ъгъл (*P.A.*) отговаря на въртене на равнината на поляризация с 6,23 ±0,05 deg/day. Открита е индикация за корелация между кривата на блясъка на обекта във филтър R и промяната в степента на поляризация.

*Vladimir Bozhilov, Evgeni Ovcharov, Milen Minev, Yordan Darakchiev, Angel Dimitrov, Stefan Georgiev, Manol Gerushin, Borislav Spassov, Kalina Stoimenova.* PHOTOPOLARIMETRICAL STUDY OF BLAZAR-TYPE AGN OJ287 IN 2012–2015 WITH THE 2M RCC TELESCOPE AT NAO ROZHEN

---

*For contact:* Dr. Vladimir Bozhilov, Department of Astronomy, Faculty of Physics, Sofia University “St. Kliment Ohridski”, 5 James Bourchier Blvd., 1164 Sofia, Bulgaria, phone: +359 2 8161413, E-mail: vbozhilov@phys.uni-sofia.bg

We present the results of a photopolarimetrical study of the blazar OJ287 in the period November 2012–January 2015. Observations were conducted using the Focal Reductor FoReRo-2 of the 2-meter RCC telescope of the National Astronomical Observatory (NAO) – Rozhen. The observed change of the position angle (*P.A.*) corresponds to mean rotation of the plane of polarization of 6.23 ±0.05 deg/day, in good agreement with previous measurements. An indication of a correlation between the change of brightness in R-band and the change in the degree of polarization is also observed.

***Keywords***: techniques: polarimetric; galaxies: active; BL Lacertae objects: general; BL Lacertae objects: individual: OJ287; galaxies: nuclei; quasars: general

***PACS numbers***: 98.54.Cm

## 1. INTRODUCTION

Blazars are amongst the most variable and energetic objects in the universe. They are a type of active galactic nuclei (AGN), powered by matter, falling on a supermassive black hole in the center of a large galaxy.

OJ287 is a BL Lac type AGN at $z = 0.306$. It is one of the most widely observed and well-studied objects of this type. Observational data dates back to 1890 and the analysis give a confirmed period of outbursts at intervals of roughly every 11 years [1]. The most prominent theoretical model for OJ287 is the one of a binary black hole (BBH) system in which the smaller black hole crosses the accretion disk of the larger one [2].

OJ287 exhibits some characteristics, which are common for all blazars of such type. However, the periodicity of its outbursts is an unusual behavior [3]. Even the best current models fail to predict some of the outbursts that are observed and polarization behavior might be the key to determine which models work better. The observations of OJ287 show linear polarization at multiple wavelengths [4]. According to polarimetrical studies, significant wavelength dependence in the degree of polarization is also observed [5]. Observations in the optical and radio- wavelengths indicate strong variability in the degree of polarization ($P$) and position angle (*P.A.*) [6].

## 2. OBSERVATIONS

This study presents our original polarimetrical and photometrical observations of OJ287 blazar-type AGN. We gathered and analyzed data from the nights of 17–18 Nov 2012, 13 Jan 2013, 04–05 Apr 2013, 30–31 Mar 2013, 19–21 Oct 2014, 17 Nov 2014, 13–14 Jan 2015. The observations were conducted with the 2-m RCC telescope at the National Astronomical Observatory Rozhen, Bulgaria, equipped with the focal reducer FoReRo2 [7]. The observation data for November 2012 through January 2015 is given in Table 1. Please note, that part of the data in the period November 2012–April 2013 was published in [13] and [14], but the data from 2014–2015 is published here for the first time.

For the observations we use a color splitter that transmits redder than 5800 Å light into the red channel and reflects bluer than 5100 Å into the blue channel of the reducer. We used an optical element with two combined Wollaston prisms that form a single polarizer. The *P.A.* of the prism differ by 45 deg, so we get four polarized beams with orientations at 0, 45, 90 and 135 deg each [13].With this setup and by using the Stokes equations [8], we perform the polarization determinations. Description of the method of measurements and the calculations can be found in [9] and [13].

**Table 1.** OJ287 observational data for 2012 November–2015 January

| Object | JD at midnight – 2456700 [d] | Images | Exposure [s] | Total integration time [s] | Total integration time [s] |
|---|---|---|---|---|---|
| HD 10476 | 49.4 | 30 | 0.2 | 6 | 6 |
| HD 14433 | 49.4 | 30 | 0.5 | 15 | 15 |
| OJ287 | 49.5 | 30 | 60 | 1800 | 1800 |
| OJ287 | 49.5 | 30 | 60 | 1800 | 1800 |
| OJ287 | 49.6 | 30 | 60 | 1800 | 1800 |
| OJ287 | 49.6 | 30 | 60 | 1800 | 1800 |
| OJ287 | 49.6 | 30 | 60 | 1800 | 1800 |
| OJ287 | 49.6 | 20 | 60 | 1200 | 1200 |
| OJ287 | 50.5 | 10 | 60 | 600 | 600 |
| OJ287 | 50.5 | 10 | 60 | 600 | 600 |
| OJ287 | 50.5 | 10 | 60 | 600 | 600 |
| HD 90508 | 106.4 | 10 | 3 | 30 | 30 |
| HD 90508 | 106.5 | 10 | 1 | 10 | 10 |
| HD 43384 | 106.5 | 10 | 1 | 10 | 10 |
| OJ287 | 106.4 | 10 | 300 | 3000 | 3000 |
| OJ287 | 106.5 | 10 | 300 | 3000 | 3000 |
| HD 144287 | 187.5 | 5 | 0.5 | 2.5 | 2.5 |
| HD 144287 | 188.4 | 5 | 0.5 | 2.5 | 2.5 |
| HD 154445 | 187.5 | 5 | 0.1 | 0.5 | 0.5 |
| HD 154445 | 188.5 | 5 | 0.2 | 1 | 1 |
| OJ287 | 187.5 | 3 | 300 | 900 | 900 |
| OJ287 | 188.4 | 15 | 300 | 4500 | 4500 |
| HD 43384 | 546.5 | 5 | 0.1 | 0.5 | 0.5 |
| HD144287 | 546.5 | 5 | 0.2 | 1.0 | 1.0 |
| OJ287 | 546.5 | 5 | 120.0 | 600.0 | 600.0 |
| OJ287 | 547.5 | 25 | 150.0 | 3750.0 | 3750.0 |
| HD 23512 | 749.5 | 10 | 0.5 | 5.0 | 5.0 |
| HD 65583 | 749.5 | 10 | 0.2 | 2.0 | 2.0 |
| OJ287 | 749.5 | 10 | 200.0 | 2000.0 | 2000.0 |
| OJ287 | 750.5 | 10 | 200.0 | 2000.0 | 2000.0 |
| HD 23512 | 778.5 | 10 | 0.2 | 2.0 | 2.0 |
| HD 65583 | 778.5 | 10 | 0.2 | 2.0 | 2.0 |
| OJ287 | 778.5 | 5 | 300.0 | 1500.0 | 1500.0 |
| HD 23512 | 835.5 | 20 | 0.5 | 10.0 | 10.0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HD 65583 | 835.5 | 20 | 0.1 | | 2.0 | 2.0 | |
| OJ287 | 835.5 | 30 | 100.0 | | 3000.0 | 3000.0 | |
| OJ287 | 836.5 | 20 | 80.0 | 1600.0 | | | 1600.0 |
| OJ287 | 836.5 | 10 | 100.0 | 1000.0 | | | 1000.0 |

## 3. DATA REDUCTION

Polarization can be determined using the Stokes parameters *I*, *Q*, *U* and *V*. They fully describe the state of polarization of an arbitrary electromagnetic wave [10]. A complete derivation of these parameters, as well as guides and methods for measuring them, are given in [8], chapters 1.6 and 1.7.

The method for measurement of the Stokes parameters is given by [9]. This method requires the usage of standard stars whose *P.A.* and polarization, are well known. Furthermore, it requires that some stars are with low *P*, and some – with high *P*. In this article, we have used the low-polarization standard stars HD 144287 and HD 65583, and the high-polarization standards HD 43384 and HD 23512.

All images have been reduced using the software package IRAF (Image Reduction and Analysis Facility, [11]). The photometric measurements have been reduced using the widely-known and accepted procedure: correction for noise (thermal and other), "flat field" correction, and median combination of the resulting images. The method, its purpose, application, advantages and drawbacks, are described broadly in many articles, e.g. [12].

Each of the combined images is then subjected to aperture photometry with aperture radius approximately equal to the full width at half maximum (FHWM). The process for finding P and P.A. is described in [9] and [13].

## 4. RESULTS AND DISCUSSION

Measurements of the polarization and the *P.A.*, as well as the number of complete rotations of the plane of polarization *k* (see [13] for elaborate details) are shown in Table 2. Magnitude in R-band and polarization curves is shown simultaneously in Fig. 1. Measurements of the *P.A.* with the applied fitting are shown in Fig. 2**.**

Our results show mean rotation of the plane of polarization of 6.23 ±0.05 deg/day, which is in good agreement with previously measured values [13]. However, further studies are necessary to understand better the physical properties of OJ 287, especially in the periods following the flares of this blazar [15].

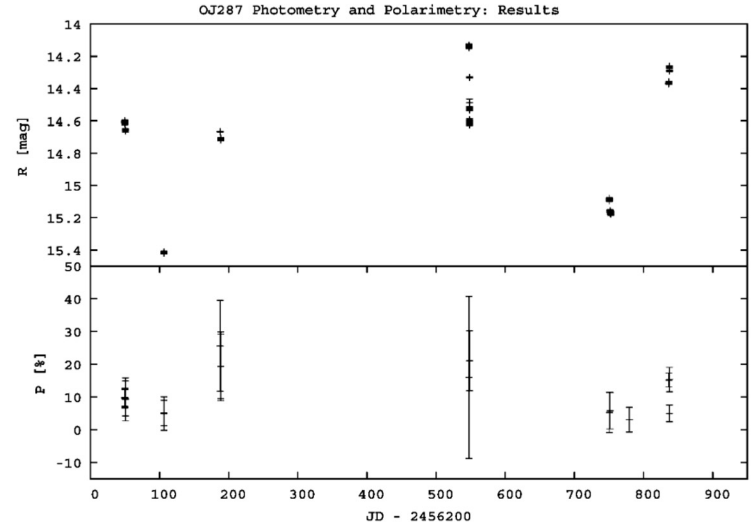


**Fig. 1.** OJ287 Photometry and Polarimetry: Results. An indication for correlation between photometric and polarimetric variability is observed

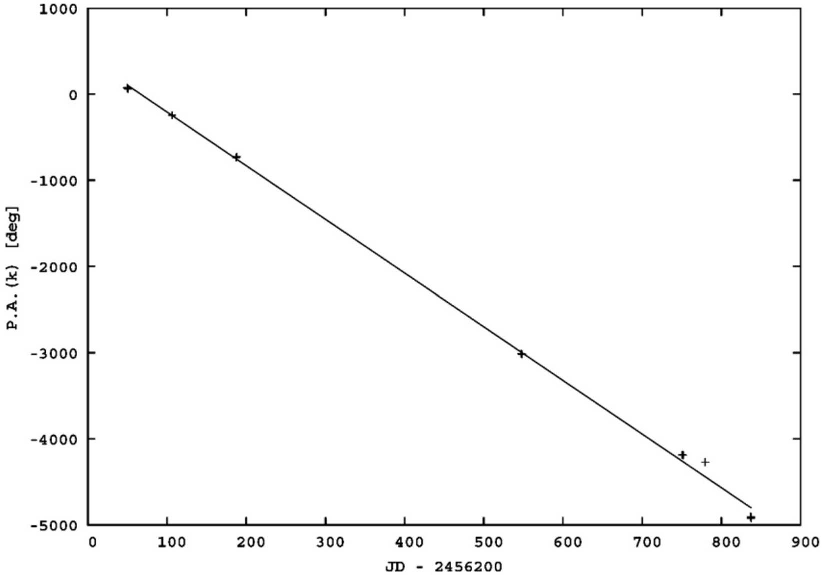


**Fig. 2.** *P.A.* Measurements for OJ 287. Observed *P.A.* change corresponds to rotation of 6.23 (± 0.05) deg/day

**Table 2**. Photopolarimetrical results for Blazar-type AGN OJ287 in the period 2012–2015

| DATE | JD - 2456200 | *P* [%] | Error [%] | *P.A.* [deg] | *P.A.*(*k*) [deg] | Error [deg] | *k* |
|---|---|---|---|---|---|---|---|
| 17 NOV 2012 | 49.495 | 9.80 | 2.72 | 73.29 | 73.29 | 3.15 | 0 |
| | 49.523 | 9.84 | 2.69 | 74.52 | 74.52 | 3.15 | 0 |
| | 49.552 | 9.71 | 2.67 | 74.77 | 74.77 | 3.15 | 0 |
| | 49.588 | 9.89 | 2.76 | 74.62 | 74.62 | 3.15 | 0 |
| | 49.618 | 9.45 | 2.78 | 73.53 | 73.53 | 3.16 | 0 |
| | 49.646 | 9.48 | 2.76 | 72.53 | 72.53 | 3.16 | 0 |
| 18 NOV 2012 | 50.491 | 9.27 | 6.56 | 62.60 | 62.60 | 3.37 | 0 |
| | 50.500 | 9.54 | 5.31 | 64.45 | 64.45 | 3.29 | 0 |
| | 50.510 | 9.53 | 5.34 | 64.17 | 64.17 | 3.29 | 0 |
| 13 JAN 2013 | 106.364 | 4.93 | 5.10 | 117.33 | -242.67 | 3.20 | 1 |
| | 106.496 | 5.07 | 3.91 | 117.91 | -242.09 | 3.07 | 1 |
| 04 APR 2013 | 187.448 | 25.59 | 13.90 | 163.65 | -736.35 | 1.53 | 2.5 |
| 05 APR 2013 | 188.348 | 19.37 | 10.52 | 172.34 | -727.66 | 1.53 | 2.5 |
| | 188.385 | 19.35 | 9.91 | 172.37 | -727.63 | 1.52 | 2.5 |
| 30 MAR 2014 | 547.390 | 15.95 | 24.74 | 46.72 | -3013.28 | 6.52 | 8.5 |
| 31 MAR 2014 | 548.337 | 21.1 | 9.15 | 42.04 | -3017.96 | 5.94 | 8.5 |
| 19 OCT 2014 | 750.576 | 5.28 | 6.15 | 132.74 | -4187.26 | 3.58 | 12 |
| 20 OCT 2014 | 751.590 | 5.81 | 5.57 | 130.1 | -4189.90 | 3.47 | 12 |
| 17 NOV 2014 | 779.524 | 3.09 | 3.77 | 48.74 | -4271.26 | 2.04 | 12 |
| 13 JAN 2015 | 836.558 | 15.18 | 2.1 | 135 | -4905.00 | 2.41 | 14 |
| 14 JAN 2015 | 837.359 | 4.99 | 2.56 | 116.84 | -4923.16 | 2.59 | 14 |
| 15 JAN 2015 | 837.673 | 15.34 | 3.72 | 132.82 | -4907.18 | 2.46 | 14 |

## 5. CONCLUSIONS

Photopolarimetrical studies of the blazar OJ287 are a powerful tool to determine between different plausible theoretical models. It is particularly important to perform such measurements during outbursts when the characteristic double-peaked flares could be separated [15]. The first peak is known to be of thermal origin while the second one could be due to either polarized synchrotron radiation or to unpolarized / low-polarized bremsstrahlung (braking radiation).

In this article we extend the polarimetrical and photometrical observations we began in the period of November 2012 through April 2013, described in [13] and [14]. We add more recent data obtained in the period from March 2014 through January 2015 in an effort to refine our results for the measurement of the rotation of the plane of polarization of OJ287. From the variation of the P.A. we were able to derive mean rotation of the plane of polarization of 6.23(± 0.05) deg/day, compared to the result from the previous data set of 5.80 (± 0.03) deg/day [13]. The combined photopolarimetrical observations from the whole period suggest significant correlation between polarization and brightness in R-band which is an expected result in some of the theoretical models [13].

Further polarimetrical observations of this object are strongly encouraged, since such results are scarce, but can serve as an important constrain to the theoretical models.

**Acknowledgements.** The analysis of the gathered data and the preparation of this publication was partially supported by project „Time domain astrophysics of selected active galactic nuclei and Local group stars“, funded by Contract DN18/10 with the National Science Fund of Bulgaria. Part of the work was done by students, participating in the lecture course „Preparation of Scientific Articles in Astronomy“, read in the Department of Astronomy, Faculty of Physics, Sofia University “St. Kliment Ohridski”

### REFERENCES

[1] Sillanpaa, A., Haarala, S., Valtonen, M. J., Sundelius, B., Byrd, G. G. *Astrophysical Journal, Part 1,* 1988, **325**, Feb. 15, 628.

[2] Lehto, Harry J., Valtonen, Mauri J. *Astrophysical Journal*, 1996, **460**, 207.

[3] Takalo, L. O. *Vistas in Astronomy*, 1994, **38**, 1, 77.

[4] Kinman, J.D. et al. *Astronomical Journal*, 1974, **79**, 349.

# ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“ ФИЗИЧЕСКИ ФАКУЛТЕТ

**Обхват на списанието**. *Годишникът* включва всички научноизследователски направления във Физическия факултет. Той се издава веднъж годишно. Публикуването в него е безплатно.

Изисквания към статиите. Статиите се изпращат по електронната поща на главния редактор (elfa@phys.uni-sofia.bg). Представянето включва текста, написан на Word в DOC формат, с включени фигури (вж. Template във web-страницата на Годишника), фигурите в отделни файлове, както и придружаващо писмо с електронните адреси на трима потенциални рецензенти.

Изисква се ръкописът да не е бил и да не бъде публикуван в никое друго издание. Всички ръкописи ще бъдат рецензирани. Авторите ще бъдат уведомени при одобряване на статиите за публикуване. Редакторите си запазват правото да редактират ръкописите, когато е нужно, и да връщат тези, които не отговарят на изискванията и обхвата на списанието.

Авторите отстъпват авторските права върху ръкописа на Физическия факултет на Софийския университет „Св. Климент Охридски». Това включва и правата за адаптиране и оформяне на статията с цел използване на компютърни програми и системи, необходими при отпечатването.

**Подготовка на ръкописа**. Желателно е ръкописите да не са по-големи от 20 стр. Те трябва да са напечатани едностранно на листове А4 и с достатъчно широки полета. Ръкописът трябва да има следната структура: заглавие, автор(и), месторабота, абстракт на английски, абстракт на български, PACS номер, ключови думи (на англ.), основен текст, благодарности, допълнения, литература (вж. Template). Името, пълен пощенски адрес, телефонен и факс-номер и електронен адрес на автора за кореспонденция трябва да са изписани на първата страница като бележка под черта.

*Фигурите* трябва да са поставени в текста, близо до първото им споменаване. Те трябва да са с високо качество (резолюция не по-малко от 300 dpi) и се представят в отделни файлове в EPS формат (за векторните изображения – чернобели чертежи и графики) и JPG или TIF формат (за растерните изображения – снимки, рисунки). Всички цветни фигури трябва да са конвертирани в чернобели.

*Таблиците* да са с минимален брой разграничителни линии, да са поместени в текста, близо до техния коментар и отделени с известно разстояние отгоре и отдолу.

*Забележките под черта* да са минимален брой, кратки и последователно номерирани.

*Литературата* да се цитира в квадратни скоби, например [3], [1, 3], [5, с./p. 98], [12, гл./Ch. 2.11], като номерирането е последователно, по реда на цитирането.

*Примери за оформяне списъка на литературата*:

[1] Haake, F. Quantum Signatures of Chaos. Berlin, 1991.

[2] Berlad, G. I., A. P. Dar, G. M. Eilam. *Phys. Rev. D*, 1980, **22**, 7, 1547.

[3] DeWitt-Morette, C. In: *NATO ASI Series B: Physics*, 1997, **361**, 51.

За по-подробна информация относно подготовката на ръкописа, моля, контактувайте с редакторите на адрес annuaire@phys.uni-sofia.bg

# ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
# FACULTY OF PHYSICS

**Scope of the journal** . The Annuaire covers all areas of research at the Faculty of Physics. It is published once a year. There is no page charge.

**Submission of papers** . The papers are to be submitted by e-mail to the Editorin-Chief (elfa@phys.uni-sofia.bg). The submission of a paper includes the paper in Word DOC format with inserted figures (see Template), the figures as separate files, and a cover letter with the e-mail addresses of three potential Referees.

The submission of a paper implies that it has not been published previously, nor has been submitted for publication elsewhere. All papers will be subjected to a peer review. The authors will be notified of the acceptance of the paper. The Editors reserve the right to edit papers, where necessary, and refuse papers, which do not satisfy conditions of standard or contents.

The acceptance of a paper implies that the authors transfer the copyright to the Faculty of Physics of St. Kliment Ohridski University of Sofia including the right to adapt the paper for uses in conjunction with computer systems and programs, and also reproduction or publication in machine-readable form and incorporation in retrieval systems.

**Manuscript preparation** . The manuscripts should be no longer than 20 pages. The manuscripts must be typed on one side of A4 paper with wide margins.

The manuscript should have the following structure: Title, Author(s), Affiliation(s), Abstract in English, Abstract in Bulgarian, PACS codes and keywords, Main text, Acknowledgements, Appendices, References.

The name, complete postal address, telephone and fax numbers and the e-mail address of the corresponding author should be given as a footnote on the first page of the manuscript.

*Figures* must be inserted in the manuscript near their first reference with captions just below the figures. High-quality figures (at least 300 dpi) must be provided in the submission as separate files in EPS format (for black and white figures) and JPG or TIF format (for grey scale figures). Any color figure should be converted into a grey scale one before submission.

*Tables* should be with a minimum number of dividing lines and should be inserted in the text near their first reference.

*Footnotes* should be brief and kept to a minimum, and must be numbered consecutively.

*References* must be consecutively numbered in the text using square brackets, e.g., [3], [1, 3], [5, p. 98], [12, Ch. 2.11], etc. They must be listed by number in the Reference list.

*Examples:*

[1] Haake, F. Quantum Signatures of Chaos. Berlin, 1991.

[2] Berlad, G. I., A. P. Dar, G. M. Eilam. *Phys. Rev. D*, 1980, **22,** 7, 1547.

[3] DeWitt-Morette, C. In: *NATO ASI Series B: Physics*, 1997, **361**, 51.

For detailed information about manuscript preparation, please, contact the Editors at annuaire@phys.uni-sofia.bg.