Professor Adam Marian Dziewoński
(1936-2016)

OBITUARY

Professor Adam Marian Dziewoński, outstanding seismologist and geophysicist, Professor of Harvard University, member of the Polish Academy of Sciences and the United States National Academy of Sciences, awarded of the Crafoord Prize in Geosciences and the William Bowie medal, passed away on 1 March 2016 in Cambridge, Massachusetts, USA.

Adam Marian Dziewoński was born on 15 November 1936 in Lwów. After completing his high school education in Wrocław, he enrolled at the Faculty of Mathematics and Physics, University of Warsaw. Still an undergraduate student, he went to Vietnam to conduct measurements of the Earth's magnetic field in the Cha-Pa Observatory, in the framework of the II International Geophysical Year (1958-1959). He graduated from the University of Warsaw, with major in geophysics, writing a master's thesis on daily magnetic activity variations in equatorial zone. In the years 1960-1965 he was employed at the Department (now Institute) of Geophysics, Polish Academy of Sciences, in Warsaw. He received his doctor's degree in the field of seismology from the Academy of Mining and Metallurgy in Cracow on the basis of the thesis: "The problems of multiple reflections in synthetic seismograms". After years, this Scientific Institution honored him with the title of "Doctor Honoris Causa".

In 1965 he moved to the United States, where he continued his scientific career, first at the Southwest Center for Advanced Studies, then at the University of Texas in Dallas, and since 1972 at Harvard University in Cambridge. The activity of Adam Dziewoński covered a wide range of topics in seismology, relating to the earthquake focal mechanisms and the Earth's in-

terior structure. His first works were about the structure of the upper mantle inferred on the basis of surface wave dispersion.

Adam Dziewoński was one of giants of science; his greatest achievements concerned: (1) studies of the Earth's free oscillations and inner core, (2) development of the Preliminary Reference Earth Model, PREM, (3) studies on earthquake mechanisms based on centroid moment tensor inversion, and (4) global seismic tomography. His accomplishments in these four areas are the milestones in seismology; each of them alone would suffice to call its author a giant.

In the epoch of "analog seismology", the digitization and then analysis of seismograms from the great Alaska earthquake of 1964 enabled him to identify and correct the modes of free oscillations of the Earth and consequently to demonstrate that the inner core is solid.

In 1981, Adam Dziewoński, jointly with Professor Don L. Anderson, developed the Earth's structure model, named the Preliminary Reference Earth Model, PREM, which has been and still is the basic reference for other models of the structure and geodynamic processes. After nearly 40 years from its publication, the PREM model continues to be widely used, with some 200 citations per year (making up a total of over 5000).

Adam Dziewoński is the author of a method for studying the earthquake mechanisms on the basis of seismic wave recordings (inversion of the centroid moment tensor). The Harvard Seismology Group that has been led by him is now publishing results of earthquake mechanism studies for several hundred greatest earthquakes every year.

Adam Dziewoński invented the global seismic tomography, which yields a three-dimensional model of the Earth's interior structure based on millions of measurements of seismic wave travel-times between the source and the seismic station. Taking advantage of the differences in seismic wave propagation velocities, it is possible to map the Earth's interior. For instance, it is possible to observe the geometry of cooler lithospheric plates submerging into warmer Earth's mantle, down to the boundary with the liquid outer core (layer D″). This is essential in the study of dynamic processes responsible for the drift of continents, earthquakes, recent tectonic movements, and volcanism.

Adam Dziewoński was very strongly involved in organizing modern seismological stations network, indispensable for obtaining the highest-quality observational materials, giving insight into the deep structure of the Earth. For many years, he has been Chairman of the Executive Committee of the Incorporated Research Institutions of Seismology (IRIS), a large consortium of universities and research institutes in the field of seismology and Earth interior research.

Recognizing the outstanding contribution of Adam Dziewoński to our understanding of the Earth's interior, the Royal Swedish Academy of Sciences honored him with the Crafoord Prize in 1998. The Crafoord Prize, funded in 1980, is awarded every year, on a rotating scheme, in one of the five disciplines that are not covered by the Nobel Prize: Mathematics, Astronomy, Geosciences, Biosciences, and Rheumatology. It was first granted in 1982 in Mathematics. The 1998 Crafoord Prize, in Geosciences, was awarded by Adam Dziewoński jointly with his close collaborator, Professor Don L. Anderson (California Institute of Technology, Pasadena, USA), for their fundamental contribution to the knowledge of the structure and processes in the Earth's interior. The Crafoord Prize is being granted by the same Committee as the Nobel Prize, and it is also presented by the King of Sweden.

Adam Dziewoński had been keeping close, friendly contact with the Polish geophysical community. He was member of Advisory Board of *Acta Geophysica*.

Marek Grad
Institute of Geophysics,
Faculty of Physics,
University of Warsaw

Aleksander Guterch
Institute of Geophysics,
Polish Academy of Sciences

On Data Interpolation
at Three Croatian Repeat Stations by Using
the Spherical Elementary Currents Systems Method

Eugen VUJIĆ

Faculty of Geodesy, University of Zagreb, Zagreb, Croatia
e-mail: eugvujic@gfz.hr

A b s t r a c t

A general mathematical tool for expanding vector systems on a sphere into basis functions, spherical elementary current system (SECS) method, was applied for separation of the geomagnetic field variations into external and internal parts, over a limited region of central and southeastern Europe. The registered variations at three Croatian repeat stations were compared to the variations estimated by the SECS method using the variations from the different sets of observatories. The results of the SECS method were also compared to a simple assumption that the variations at repeat station are equal to those at particular observatory. The relevance of this comparison was to get an insight about the possibility of using the SECS method for estimating the geomagnetic field variations over Croatia. The guidelines for the application of the SECS method for the purpose of reducing repeat station data were also given.

Key words: geomagnetic repeat station, geomagnetic observatory, geomagnetic data interpolation, spherical elementary currents system.

1. INTRODUCTION

The geomagnetic field measured at the Earth's surface is a superposition of several contributions. The main field originates from the electrically conductive fluid motions in the outer core, and the lesser part of the geomagnetic

field originates from the magnetized rocks in the crust and upper mantle (the lithospheric field). There are also contributions of the ionospheric and magnetospheric currents (*i.e.*, the external field) and their induced effects in the lithosphere. The geomagnetic field is continuously monitored at geomagnetic observatories and periodically on the repeat stations network (Mandea and Purucker 2005).

The external field variations (and associated induced effects) at some repeat stations and at surrounding observatories can be prominently different, and this occurs, *e.g.*, when these sites are quite far away (especially in latitude), and/or the lithospheric electrical conductivity differs under these sites (Korte and Thébault 2007). The time variations at some point can be interpolated by using the data recorded at relatively nearby magnetometers, up to few hundreds of kilometers, by using the simple assumption that these variations (about some arbitrary baselines) are the same at those sites. Such a method is commonly used in repeat stations data reduction procedure (Newitt *et al.* 1996).

Several methods, based on spectral decomposition of magnetic potential in spherical or rectangular geometry (Amm and Viljanen 1999 and references therein), were developed in order to model the spatial distribution of geomagnetic time variations at the Earth's surface. In this paper the attention is focused on spherical elementary current systems (SECS) method, introduced by Amm (1997). This method for field continuation from ground to the ionosphere is not based on spectral decomposition. It expands the measured ground geomagnetic time variations into a sum of the magnetic field contributions of spherical elementary current systems placed in the ionosphere (Amm 1997, Amm and Viljanen 1999, Pulkkinen *et al.* 2003b), or in the ionosphere and at some depth inside the Earth (Pulkkinen *et al.* 2003a). The centers of these elementary current systems (*i.e.*, "poles") can be placed freely, such that their locations are most suitable with respect to the type of magnetic variations to be analysed or to the density of measurement sites (Amm and Viljanen 1999). The amplitudes of elementary systems (scaling factors) are represented by the scaling factor matrix (Pulkkinen *et al.* 2003a), and it is the solution to be determined by the data from measurement sites.

Some examples of using the SECS method are: modelling of Cowling effect (Amm 1997, Amm and Viljanen 1999, Amm *et al.* 2013), modelling of some ionospheric events and their electrical current configurations, the magnetic field effects of which were recorded by BEAR and/or IMAGE magnetometer arrays (Pulkkinen *et al.* 2003a, b; Vanhamäki *et al.* 2003, Apatenkov *et al.* 2004), determining of ionospheric currents from CHAMP satellite measurements (Juusola *et al.* 2006), data interpolation at some observatories in Europe and North America (McLay and Beggan 2010), and deriving the

geomagnetically induced electric field at the Earth's surface by using the data from BEAR and IMAGE magnetometer arrays (Vanhamäki *et al.* 2013).

The prime aim of this work was to inspect an implementation of the SECS method for the estimation of the centered geomagnetic field time series over Croatia. The usage of centered data means that the variations' baselines are the mean values of geomagnetic components over some interval. The other variations with different baselines could also be used for testing the SECS method, and some of those baselines are the quiet night values and the annual mean values. Their accuracies at Croatian repeat stations were not satisfactory for this case study. The estimations derived by the SECS method were also compared to the corresponding estimations derived by a simple method, which assumes that the centered data at a repeat station are equal to the centered data registered at the relatively close observatory. Implicitly, the relevance of this comparison was to inspect an alternative (improved) method for the reduction of repeat station data.

2. METHODS AND DATA

2.1 Spherical elementary currents systems method

Two types of spherical elementary sheet currents have been defined by Amm (1997), one being divergence-free and the other curl-free (Amm and Viljanen 1999). In the case of ground disturbance continuation, it is necessary to consider only the divergent-free elementary systems (Amm and Viljanen 1999, Juusola *et al.* 2006). Their poles are placed in the ionosphere of radius R_i and below the Earth's surface at radius R_g (Pulkkinen *et al.* 2003a), respectively, both in geocentric reference frame. The definition of one elementary sheet currents system, in a spherical coordinate system (r', θ', φ') in which the pole of the elementary system is at $\theta' = 0$, is (Amm and Viljanen 1999, McLay and Beggan 2010):

$$J_c(\theta) = \left(I_{0,c}/4\pi R_c\right) \cdot \cot(\theta/2) \cdot \hat{\varphi}' , \qquad (1)$$

where index $c = i, g$ (i stands for the ionosphere and g for the ground), $I_{0,c}$ is scaling factor of the elementary system, radii R_i and R_g are defined as 110 km above the Earth's surface and 100 km below it, respectively (Amm and Viljanen 1999, McLay and Beggan 2010). In fact, there are two infinitely thin horizontal current layers, the first in the ionosphere and the second inside the Earth, since any divergence-free current system can be composed by superposition of elementary current systems (Pulkkinen *et al.* 2003a). The magnetic effect of these two layers at some point on the Earth's surface is (in geocentric frame):

$$\boldsymbol{B}(r,\theta,\varphi,t) = \sum_{m=1}^{M} I_{m(0,i)}(t) \cdot T_i\left(r,\theta,\varphi,R_i,\vartheta_m,\varphi_m\right)$$

$$+ \sum_{s=1}^{S} I_{s(0,g)}(t) \cdot T_g\left(r,\theta,\varphi,R_g,\vartheta_s,\varphi_s\right) , \tag{2}$$

where the scaling factors I are the functions of time t, $T_{i,g}$ are geometric parts related to the external and internal part ($R_g < r < R_i$) of magnetic field produced by each elementary current system located at (R_i, θ_m, φ_m) and (R_g, θ_s, φ_s), while M and S are the numbers of poles related to these two current layers, respectively. The expressions for $T_{i,g}$ are given by Amm and Viljanen (1999). The linear system of equations given by Eq. 2 can be written in a matrix form as (McLay and Beggan 2010):

$$\mathbf{B} = \mathbf{T} \cdot \mathbf{I} , \tag{3}$$

A procedure of deriving $\mathbf{I}$ can be taken, *e.g.*, under an assumption that the standard deviations of the errors in measurements are equal at each measurement point at the Earth's surface (Pulkkinen *et al.* 2003a), which can be a reasonable assumption for geomagnetic observatories. The matrix $\mathbf{T}$, of the order 3N by (M + S), where N is a number of measurement points, is then of the form (Pulkkinen *et al.* 2003a):

$$\begin{bmatrix} & & \cdots & & \\ T_{i,\theta}^{k,1} & T_{i,\theta}^{k,M} & T_{g,\theta}^{k,1} & T_{g,\theta}^{k,S} \\ T_{i,\varphi}^{k,1} & \cdots T_{i,\varphi}^{k,M} & T_{g,\varphi}^{k,1} & \cdots T_{g,\varphi}^{k,S} \\ T_{i,r}^{k,1} & T_{i,r}^{k,M} & T_{g,r}^{k,1} & T_{g,r}^{k,S} \\ & & \cdots & & \end{bmatrix} , \tag{4}$$

where k refers to kth measurement point, and, *e.g.*, the element $T_{i,\theta}^{k,1}$ is the geometric part related to the magnetic field of internal (i) origin, produced at kth point in polar direction (θ) of geocentric frame, and by superscripts (1, ..., M) and (1, ..., S) are denoted the elementary systems, respectively. Further, the matrix $\mathbf{B}$, of the order 3N by 1, can be derived from the geomagnetic field variations (B_θ, B_φ, B_r) measured at the points (1, ..., N) on the Earth's surface (in geocentric frame), and in this case it is equal to a transpose of:

$$\left(B_\theta^1\ B_\varphi^1\ B_r^1 ... B_\theta^k\ B_\varphi^k\ B_r^k ... B_\theta^N\ B_\varphi^N\ B_r^N \right) . \tag{5}$$

After deriving matrices **T** and **B**, matrix **I** can be calculated. This system of equations is highly underdetermined (Amm and Viljanen 1999, Pulkkinen *et al.* 2003a, McLay and Beggan 2010), since the number of unknowns $(M + S)$ is in general much greater than the number of measurements at the Earth's surface (3N), at a particular time moment (one minute in this case). The inversion of matrix **T** can be performed by its singular value decomposition (Press *et al.* 2001), and after that the matrix **I** can be calculated at each minute. The usual procedure is that the stabilization in inversion of matrix **T** is done, by choosing the threshold ε for singular values related to different basis vectors of the decomposition, and if larger ε is choosen, the solutions for **I** will, in general, be smoother (Amm and Viljanen 1999, Pulkkinen *et al.* 2003a).

Matrix **I** can be derived, *e.g.*, by using two networks of poles of elementary current systems and field values from observatories, the latter being distributed over some region on the Earth's surface (McLay and Beggan 2010). Once matrix **I** is known, the magnetic field variation at some point on the Earth's surface can be calculated by using Eq. 3, but with matrix **T** calculated for that position and given elementary current systems networks.

These estimated variations can be compared to the measured one at the repeat station, in order to derive root-mean-square (rms) error, defined as $\mathrm{rms}\,(\mathrm{error}) = (\sum \mathrm{error}^2/n)^{0.5}$, where n is the length of one-minute time-series used for testing the interpolation by SECS method, and the maximal absolute error $(\max(|\mathrm{error}|))$ of interpolation. These two parameters were obtained for the errors of estimated centered data, and the estimations were derived, respectively: by the SECS method and by a simple method (an assumption that the centered data at repeat station are equal to the simultaneous centered data at particular observatory).

Further, the SECS method could also be used for the purpose of reducing repeat station data to their annual mean values. Provided that $E(t)$ is the absolute measurement of geomagnetic component E at the instant t, $\mathrm{var}(t)$ is the corresponding variation of that component with respect to its annual mean value, and if $\mathrm{var}(t)$ can be estimated by the SECS method according to Eq. 3, then the annual mean value at repeat station can be determined from:

$$E^{\mathrm{annual\,mean}} = E(t) - \mathrm{var}(t) \ . \tag{6}$$

The simultaneous geomagnetic components' variations at the observatories, which are necessary as the input values in the SECS method (for deriving the matrix **I**), are derived as: $E_{\mathrm{O}}(t) - E_{\mathrm{O}}^{\mathrm{annual\,mean}}$, where subscript O is for observatory. If a simple method is used for data reduction, one has (Newitt *et al.* 1996):

$$E^{\text{annual mean}} = E(t) - \left[E_O(t) - E_O^{\text{annual mean}} \right].$$ (7)

2.2 Data used

During July 2010 the Overhauser effect dIdD (deflected Inclination deflected Declination) variometer was installed successively on three Croatian repeat stations (Fig. 1), Krbavsko Polje (KRBP), Sinjsko Polje (SINP), and Pala-gruža (PALA), in a frame of the joint Croatian-Hungarian project (Csontos *et al.* 2012). More on the Croatian geomagnetic networks can be found in Brkić (2013). The one-minute values of X (north), Y (east), and Z (vertical) components were calculated after the instrument calibration. Due to some unexpected artificial disturbances, several intervals of data for each station were dropped. In Table 1, the usable intervals larger than five hours are displayed, and these twelve time series (four per repeat station) are denoted with letters a, b, ..., l. The baselines of these variations were taken as the mean values of components over each time interval. The total field time series from KRBP, SINP, and PALA are displayed in Fig. 1, and twelve time series (a, b, ..., l) are also denoted.

In this case study, the one-minute values from ten European geomagnetic observatories (Fig. 1) were used for interpolation of the intervals given in Table 1: Budkov (BDV), Black Forest (BFO), Castello Tesino (CTS), Fürstenfeldbruck (FUR), Grocka (GCK), Hurbanovo (HRB), Nagycenk (NCK), Panagjurishte (PAG), Surlari (SUA), and Tihany (THY). The one-minute data were obtained from at Intermagnet Web Site (http://www.intermagnet.org). These observatories were selected because of their proxi-
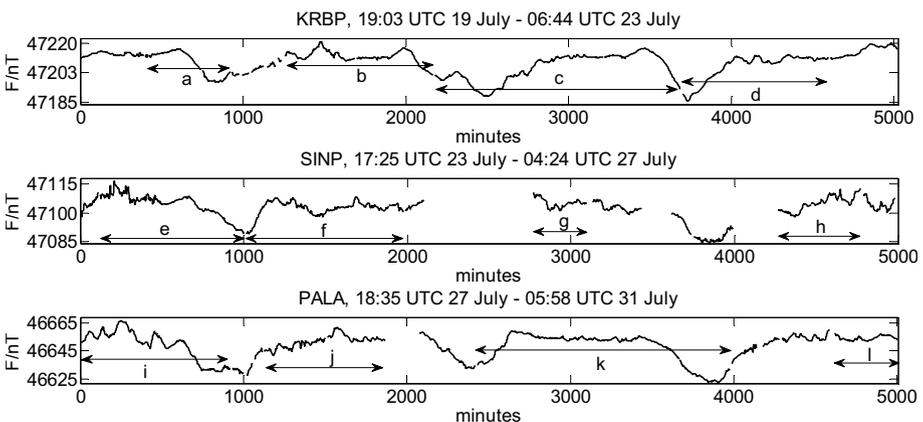
Fig. 1. The total field time series from KRBP (top), SINP (middle), and PALA (bottom). The durations of twelve time series (a, b, ..., l) are also denoted.

The durations of twelve (a, b, ..., l) time series used in this study
which were recorded at three Croatian repeat stations

Repeat station	Time series	Time interval [UTC]	Duration [min]	Kp
KRBP	a	20 July, 02:03-10:28	503	1-2
	b	20 July, 15:57 – 21 July, 07:08	903	1-2
	c	21 July, 07:24 – 22 July, 08:28	1501	0-2
	d	22 July, 08:44-23:37	894	1
SINP	e	23 July, 19:26 – 24 July, 10:09	881	0-3
	f	24 July, 10:20 – 25 July, 02:20	949	0-2
	g	25 July, 15:36-21:04	326	1
	h	26 July, 16:34 – 27 July, 00:56	500	1-3
PALA	i	27 July, 18:35 – 28 July, 09:18	884	2-4
	j	28 July, 13:23 – 29 July, 01:38	725	2-3
	k	29 July, 11:06 – 30 July, 12:48	1540	1-2
	l	30 July, 23:28 – 31 July, 05:58	391	1-2

Note: The activity index Kp is also given (obtained from http://swdcwww.
kugi.kyoto-u.ac.jp).

The distances (in km) of three Croatian repeat stations (KRBP, SINP, and PALA)
from ten geomagnetic observatories

km	BDV	BFO	CTS	FUR	GCK	HRB	NCK	PAG	SUA	THY
KRBP	505	691	347	512	407	406	340	728	839	304
PALA	764	905	548	751	441	628	584	650	844	518
SINP	637	828	478	653	344	483	443	622	771	374

mity to Croatian repeat stations, and a few of them (FUR, GCK, and THY)
were used previously in data reduction of Croatian repeat stations surveys
(Vujić *et al.* 2011). The distances of these observatories from Croatian repeat
stations are given in Table 2. One can notice that there is no observatory for
data interpolation southwest and south of repeat stations.

Considering the SECS method (*i.e.*, deriving of matrix **I**) and data inter-
polation at repeat stations, four sets of observatories were used, respectively:
set A (CTS, GCK, HRB, NCK and THY), set B (BDV, CTS, GCK, PAG
and THY), set C = set A + BDV + PAG (seven observatories), and set D
consists of all ten observatories. Set A contains five closest observatories to
the repeat stations, and set B consists of the observatories that constitute ap-
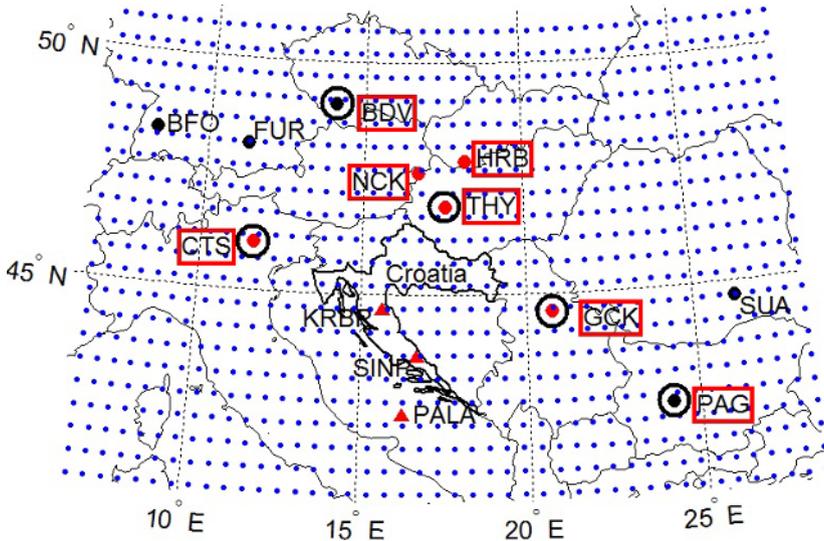
Fig. 2. The locations of ten geomagnetic observatories (black and red dots) used for data estimation at three Croatian repeat stations, and the locations of those repeat stations (red triangles) inside a spatial domain of the SECS's poles. The SECS pole grid is marked with blue dots (with 0.5° spacing). The observatories from set A (CTS, GCK, HRB, NCK, and THY) are marked by red dots, the observatories from set B (BDV, CTS, GCK, PAG, and THY) are marked with black circles, and the observatories from set C (set A + set B) are denoted by red quadrangles.

proximately equidistant spatial distribution of observatories around the repeat stations. The data of the Croatian stations themselves were not used for the SECS method. Beside those sets, two networks of poles of elementary currents systems were used, and their spatial domain is also shown in Fig. 2. These two networks were rectangular ones with spacing δ. Further, as none of the three Croation stations is situated inside the hull made up by the observatories, one can talk about extrapolation, not interpolation, in a spatial sense. The term interpolation refers to the interpolation of the one-minute time series.

3.   RESULTS AND DISCUSSIONS
The first step in this case study was to find the optimal values of parameters ε and δ. The values of rms (error) and (max (|error|), *i.e.*, error parameters, were derived for repeat stations time series by using the observatory sets (A, B ,C, and D), with the values for ε and δ in the ranges $\varepsilon = 0.01\text{-}0.1$, as suggested in Amm and Viljanen (1999), and $\delta = 0.2°\text{-}0.8°$, respectively. The following conclusions arised from these analyses: (i) the errors parameters
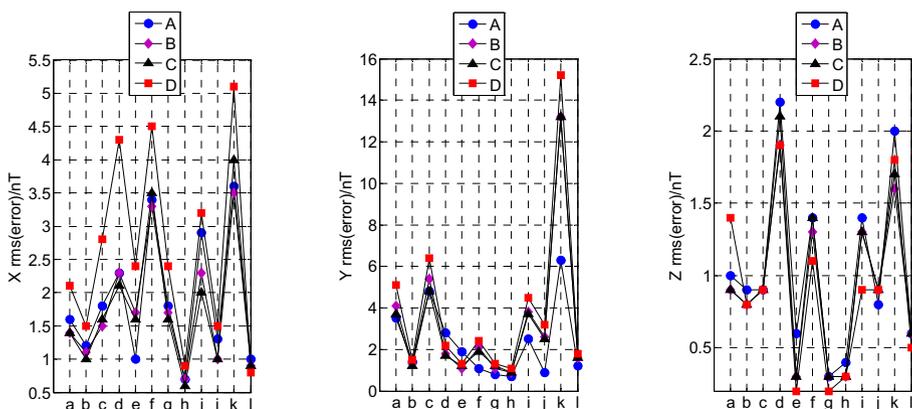
Fig. 3a. The root-men-square errors of SECS method in geomagnetic components for given time intervals (a, b, ..., l), derived by using different observatory sets (A, B, C, and D) for data interpolation, with $\delta = 0.5°$ and $\varepsilon = 0.1$.
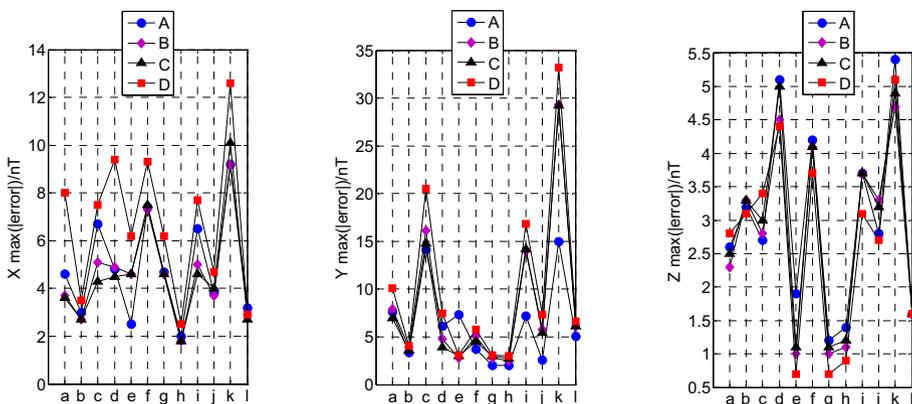
Fig. 3b. Maximal absolute errors of SECS method in geomagnetic components for given time intervals (a, b, ..., l), derived by using different observatory sets (A, B, C, and D) for data interpolation, with $\delta = 0.5°$ and $\varepsilon = 0.1$.

are almost not sensitive to δ, for a given set of observatories and ε, but slightly better values are for $\delta = 0.5°$; (ii) the error parameters were sensitive to ε for Y, but were stable for X and Z, and the minimal values were for $\varepsilon = 0.1$; (iii) set C was the most suitable for KRBP, while set A for SINP and PALA, for geomagnetic variations interpolation; (iv) set D has given the worst results in errors parameters. The errors parameters derived with different observatory sets and values of $\delta = 0.5°$ and $\varepsilon = 0.1$ are displayed in Figs. 3a and 3b, respectively. On average, they have the lowest values in X for KRBP, and in Y and Z for SINP, and in general they are the smallest in Z and highest in Y.

As mentioned, none of these repeat stations is situated inside the hull made up by the observatories. In Pulkkinen *et al.* (2003a) it was derived that at the edges of the magnetometer array, and at the regions where the two-dimensionality of the array is lost, errors of data interpolation may become larger. One can expect the results from the Croatian repeat stations to be different provided that the stations are situated inside the hull. The same procedure was performed also when the data from a particular station were included in the SECS method. In this case, the errors were much lower, as can be expected.

The sets A or C gave the best results for those three repeat stations, on average. If the sets of observatories are ordered with respect to their mean distances from KRBP and SINP stations, one obtains: A, C, B, and D, *i.e.*, A is the closest set, and D is the farthest one. The same order for station PALA is: A, B, C, and D. Further, a magnetic field produced by the elementary system decreases with the polar angle, as measured from its pole (Amm and Viljanen 1999). It can be presumed that the scaling factors of the elementary systems which are closer in latitude to the observatories will be more accurately determined by inversion process using the data from those observatories, than the scaling factors of the elementary systems that are farther from the observatories. On the other hand, a magnetic field at repeat station that is closer to those observatories or is situated inside the hull of those observatories, will be more accurately determined. These could be the reasons why a spatial interpolation is dominant over extrapolation. Based on these simple arguments, one can qualitatively explain why the set A could be the most suitable for SINP and PALA stations, and the set D the worst choice for all three stations.

The set C comprises two sets, A and B, so it could be expected that the results for the set C are at least the same as for either of them, since generally adding information should never worsen the output. It can be presumed that this is not the case for stations SINP and PALA due to a spatial extrapolation. The observatories BDV and PAG are probably relatively far away from them, and it could be possible that the elementary systems in the vicinity of those observatories produce some unfavorable total contribution (noise) at SINP and PALA. The total contribution from the elementary systems in the vicinity of those two observatories could be probably favorable at KRBP station, since it is possible that BDV is close enough and PAG is far enough from that station.

Also, it was checked whether the SECS method or a simple method (by particular observatory) better estimates the centered data (variations) at repeat stations. The results, in terms of the minimal values of errors parameters, are presented in Table 3. One can notice that the results are about the same for both methods of interpolation, and none of them is dominant. The

The results of analysis whether the SECS method or some observatory (Obs.)
better explains the centered data at a particular repeat station

Set or observatory	rms (error)			max (|error|)		
Time series	X	Y	Z	X	Y	Z
a	SECS	Obs.	SECS	Obs.	Obs.	SECS
b	Obs.	SECS	SECS	Obs.	SECS	SECS
c	Obs./SECS	Obs.	SECS	Obs./SECS	Obs.	SECS
d	Obs.	SECS	Obs.	SECS	SECS	Obs.
e	SECS	Obs.	Obs./SECS	SECS	Obs.	SECS
f	Obs.	Obs./SECS	SECS	Obs.	Obs.	SECS
g	Obs.	Obs./SECS	Obs./SECS	Obs.	Obs./SECS	Obs./SECS
h	Obs./SECS	SECS	Obs./SECS	Obs./SECS	SECS	Obs./SECS
i	Obs.	Obs.	SECS	Obs.	Obs.	SECS
j	Obs.	SECS	SECS	Obs.	SECS	SECS
k	Obs.	Obs.	Obs.	Obs.	Obs.	SECS
l	Obs./SECS	SECS	SECS	Obs.	Obs.	SECS

Note: The multiple results for some time series mean that the minimal values of errors' parameters were achieved in both cases.

SECS method was generally better for Z component estimations, and the method with particular observatory was generally better for X component, while for Y component they were about the same. In Table 4 there are displayed relatively high values of linear correlation coefficients between original and reconstructed (interpolated by SECS method) time variations of X, Y, and Z components at repeat stations. Table 5 gives error parameters for some characteristic variations (quiet-night and larger parts of diurnal variation). These results could be an indication that the quiet-night values could be relatively accurately estimated.

As the examples, the next two figures display the time variations of geomagnetic components for some particular time intervals, together with the simultaneous reconstructed series derived with the SECS method. In Fig. 4 are the results for major part of diurnal variation and a smaller part of quiet-night interval at KRBP station. The maximal absolute errors were 4.5, 3.9, and 5.0 nT in X, Y, and Z, respectively. In Fig. 5 are displayed time variations that include bay disturbance connected with geomagnetic pulsations that begun at 20:32 UTC on 27 July (Dr. H.J. Linthe, Niemegk Observatory, private communication). The maximal absolute errors for data in Fig. 5 were

Table 4

The linear correlation coefficients (*CC*) between original
and reconstructed (interpolated by SECS method) time variations
of *X, Y,* and *Z* components at repeat stations

Time series	CC_X	CC_Y	CC_Z
a	0.9886	0.9810	0.9801
b	0.9880	0.9965	0.9149
c	0.9844	0.9730	0.9936
d	0.9900	0.9910	0.9547
e	0.9934	0.9907	0.9797
f	0.9300	0.9936	0.9614
g	0.9396	0.9353	0.9870
h	0.9955	0.9876	0.8951
i	0.9861	0.9876	0.9412
j	0.9887	0.9972	0.7050
k	0.8828	0.9965	0.9898
l	0.9194	0.9831	0.9211

Note: In the case of KRBP variations, observatories set C was
used, and in cases of SINP and PALA, set A was used.

Table 5

The root-mean-square and maximal absolute errors for quiet-night variations
(first four rows), and for the parts of diurnal variations (last two rows),
of geomagnetic components at repeat stations

Variation	Station	rms (error)			max (|error|)		
		X [nT]	*Y* [nT]	*Z* [nT]	*X* [nT]	*Y* [nT]	*Z* [nT]
20 July, 22:00 – 21 July, 01:00 UTC	KRBP	0.1	0.2	0.2	0.4	0.5	0.6
21 July, 22:00 – 22 July, 01:00 UTC	KRBP	0.2	0.3	0.2	0.5	0.7	0.6
24 July, 22:00 – 25 July, 01:00 UTC	SINP	0.5	0.3	0.2	1.1	0.9	0.5
29 July, 22:00 – 30 July, 01:00 UTC	PALA	0.3	0.3	0.2	1.0	0.9	0.4
21 July, 07:20-20:00 UTC	KRBP	1.0	4.5	1.1	2.4	8.2	3.1
22 July, 08:44-20:00 UTC	KRBP	2.3	1.5	2.2	4.4	3.4	4.9

Note: The used observatories sets are the same as for the results in previous table.

3.3, 1.9, and 1.3 nT in *X, Y,* and *Z,* respectively. One can notice that there is
no systematic under-estimations or over-estimations of variations in those
two cases, and that *X* and *Y* components time series are very highly corre-
lated.
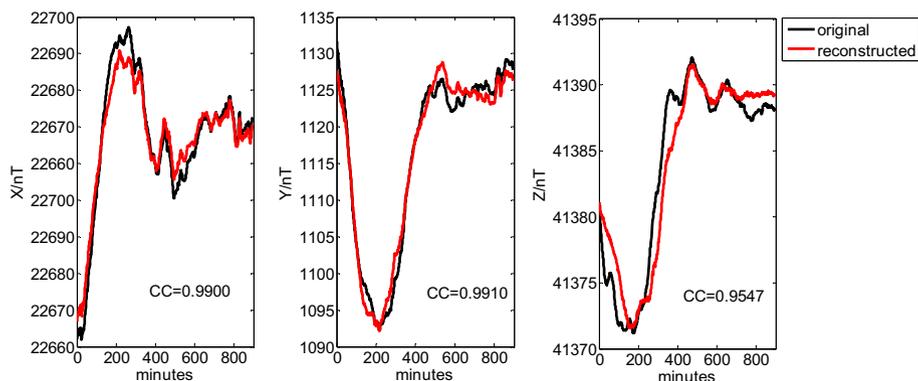
Fig. 4. The time variations of geomagnetic components at KRBP station during the time interval 22 July, 08:44-23:37 UTC. The observatories set C was used for data reconstruction.
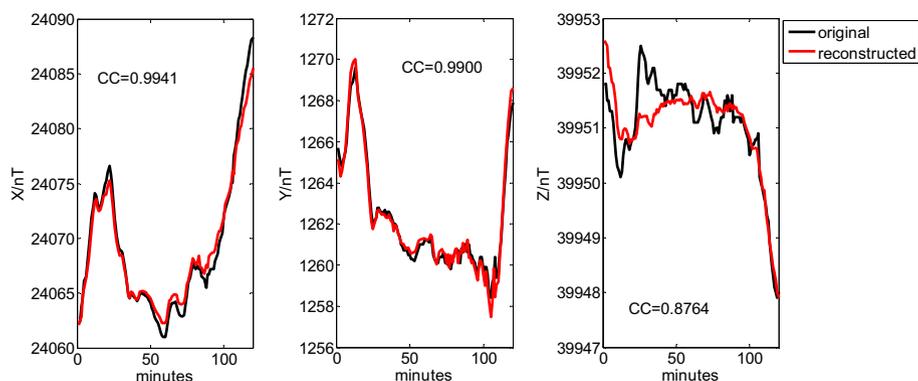
Fig. 5. The time variations of geomagnetic components at PALA station during the time interval 27 July, 20:32-22:32 UTC. The observatories set A was used for data reconstruction.

4. CONCLUSIONS

The method of spherical elementary current systems was implemented for interpolation of geomagnetic time variations at three Croatian repeat stations for the first time. The data from different sets of geomagnetic observatories were used for interpolation of geomagnetic time variations at the positions of repeat stations. The time variations in this case study were the centered time series of geomagnetic components. It was possible to find which method parameters and sets of observatories describe the variations at stations in a best manner, *i.e.*, to have the smallest root-mean and maximal absolute errors. They were in general smallest in *Z* component and highest in *Y* component.

This method was compared to the results of a simple assumption that the time variations of the geomagnetic components are the same at the particular observatory and the repeat station. These two methods gave about the same results; on average, SECS method gave better results in Z component, and simple method in X component, while in Y the results were about the same. The time series from three stations were statistically relatively small samples, and for more consistent conclusions it is necessary to have longer time series of quiet as well as disturbed conditions. Also it could be interesting to see such results if variometers could be installed simultaneously in southern and southwestern part of spatial domain, since for determining current scale factors in SECS method there were no data from that region.

The relevance of the described comparison was to implicitly present an alternative method for the reduction of repeat station data. Since the annual mean values of geomagnetic components at Croatian repeat stations have quite lower accuracies than the corresponding values at observatories, they were not used as variations' baselines in this case study. However, the instructions for the usage of the SECS method for the purpose of reducing repeat station data were also given, in the case when an influence of the differences of secular variations between a repeat station and the observatories can be neglected. The application of this method for the purpose of reducing repeat station data from Croatia can be expected in future.

Considering overall message/recommendation to someone who may face a similar problem in a different region, one can presume that there are two approaches to determine the (optimal) SECS method parameters: without an *in situ* variometer, or to perform an empirical analysis by using a variometer data. The former approach is possible only if the time variations of geomagnetic components can be modeled over the region of interest, prior to an application of the SECS method. However, one can also determine those parameters by finding the minimum misfit between the SECS method predictions and know observations, which are recorded by a variometer. Those conclusions have to be based on an adequate amount of the time series of interest (*e.g.*, the diurnal variations for data reduction).

A spatial distribution of the observatories around the points of interest is also important, as derived here and in the previous studies. Those observatories are necessary for deriving a scalar factors' matrix. It is advisable to inspect which combination of the observatories gives the best solution for the time series of interest. Furthermore, as derived in the cited previous studies, it will be more efficient if the points of interest are inside of the hull made up by the observatories, *i.e.*, if the data at those points are interpolated in a spatial sense, rather than extrapolated (which was the case in this exercise).

R e f e r e n c e s

Amm, O. (1997), Ionospheric elementary current systems in spherical coordinates and their application, *J. Geomagn. Geoelectr.* **49**, 7, 947-955, DOI: 10.5636/jgg.49.947.

Amm, O., and A. Viljanen (1999), Ionospheric disturbance magnetic field continuation from the ground to the ionosphere using spherical elementary current systems, *Earth Planets Space* **51**, 6, 431-440, DOI: 10.1186/BF03352247.

Amm, O., R. Fujii, H. Vanhamäki, A. Yoshikawa, and A. Ieda (2013), General solution for calculating polarization electric fields in the auroral ionosphere and application examples, *J. Geophys. Res.* **118**, 5, 2428-2437, DOI: 10.1002/jgra.50254.

Apatenkov, S.V., V.A. Sergeev, R. Pirjola, and A. Viljanen (2004), Evaluation of the geometry of ionospheric current systems related to rapid geomagnetic variations, *Ann. Geophys.* **22**, 1, 63-72, DOI: 10.5194/angeo-22-63-2004.

Brkić, M. (ed.) (2013), *Basic Geomagnetic Network of the Republic of Croatia 2004 – 2012, with Geomagnetic Field Maps for 2009.5 Epoch*, Croatian State Geodetic Administration, Zagreb (in Croatian).

Csontos, A., D. Šugar, M. Brkić, P. Kovács, and L. Hegymegi (2012), How to control a temporary DIDD based observatory in the field? *Ann. Geophys.* **55**, 6, 1085-1094, DOI: 10.4401/ag-5447.

Juusola, L., O. Amm, and A. Viljanen (2006), One-dimensional spherical elementary current systems and their use for determining ionospheric currents from satellite measurements, *Earth Planets. Space* **58**, 5, 667-678, DOI: 10.1186/BF03351964.

Korte, M., and E. Thébault (2007), Geomagnetic repeat station crustal biases and vectorial anomaly maps for Germany, *Geophys. J. Int.* **170**, 1, 81-92, DOI: 10.1111/j.1365-246X.2007.03387.x.

Mandea, M., and M. Purucker (2005), Observing, modeling and interpreting magnetic fields of the solid Earth, *Surv. Geophys.* **26**, 4, 415-459, DOI: 10.1007/s10712-005-3857-x.

McLay, S.A., and C.D. Beggan (2010), Interpolation of externally-caused magnetic fields over large sparse arrays using Spherical Elementary Current Systems, *Ann. Geophys.* **28**, 9, 1795-1805, DOI: 10.5194/angeo-28-1795-2010.

Newitt, L.R., C.E. Barton, and J. Bitterly (1996), *Guide for Magnetic Repeat Station Surveys*, International Association of Geomagnetism and Aeronomy, Boulder.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (2001), *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, Cambridge University Press, New York.

Pulkkinen, A., O. Amm, A. Viljanen, and BEAR Working Group (2003a), Separation of the geomagnetic variation field on the ground into external and internal parts using the spherical elementary current system method, *Earth Planets Space* **55**, 3, 117-129, DOI: 10.1186/BF03351739.

Pulkkinen, A., O. Amm, A. Viljanen, and BEAR Working Group (2003b), Ionospheric equivalent current distributions determined with the method of spherical elementary current systems, *J. Geophys. Res.* **108**, A2, 1053, DOI: 10.1029/2001JA005085.

Vanhamäki, H., O. Amm, and A. Viljanen (2003), One-dimensional upward continuation of the ground magnetic field disturbance using spherical elementary current systems, *Earth Planets Space* **55**, 10, 613-625, DOI: 10.1186/BF03352468.

Vanhamäki, H., A. Viljanen, R. Pirjola, and O. Amm (2013), Deriving the geomagnetically induced electric field at the Earth's surface from the time derivative of the vertical magnetic field, *Earth Planets Space* **65**, 9, 997-1006, DOI: 10.5047/eps.2013.03.013.

Vujić, E., M. Brkić, and N. Radović (2011), Analysis of Croatian geomagnetic repeat station surveys in 2004 and 2007, *Stud. Geophys. Geod.* **55**, 4, 737-753, DOI: 10.1007/s11200-010-9037-6.

Site-specific Probabilistic Seismic Hazard Map of Himachal Pradesh, India.
Part I. Site-specific Ground Motion Relations

Prabhu MUTHUGANEISAN and S.T.G. RAGHUKANTH

Department of Civil Engineering, Indian Institute of Technology Madras,
Chennai, India; e-mail: raghukanth@iitm.ac.in

Abstract

This article presents four regional site-specific ground motion relations developed for the state of Himachal Pradesh in northwest Himalaya, situated in a seismically active region. These relations are developed from synthetic free surface ground motion databases obtained from a calibrated stochastic seismological model considering the characteristic properties of this specific region. The adopted methodology incorporates the site effects characterised through active MASW tests conducted in 22 important cities. The estimated ground motion levels from the developed relations are found to be in reasonable agreement with the recorded data.

Key words: PSHA, GMPE, seismic hazard, site coefficients, Himalaya.

1. INTRODUCTION

Himachal Pradesh (HP), the study region, is a northern state in India situated in the western part of the seismically active Himalayan region. This region has experienced several major earthquakes in the past, namely, the Kangra earthquake (*Mw* 7.8, 1905) with its epicenter in the Kangra region, which resulted in extensive damages to infrastructure and human life. The seismic zone map of India presented in IS:1893 (2002) has assigned this region as

Zone IV (severe) and V (very severe) with damage intensity of VIII and IX, respectively. Despite the natural hazard posed to this region, the demand posed by tourism and topology has made the land a prime commodity leading to congested development and bunching of population in the major cities and towns. Additionally, the rapid population increase has resulted in several unplanned urban development and unscientific road-cuts in this hilly region. In addition, the state has been promoting several important projects such as hydroelectric power projects, tunnels, roads, *etc*. Without consideration of the appropriate seismic hazard of this region, these construction designs face the risk of being vulnerable during a probable seismic event larger than the considered design basis earthquake, thereby risking the human life and the investments.

An important aspect of hazard estimation is the availability of a reliable ground motion relation. But, the lack of reliable ground motion database for Himalayan region, which is complete for all magnitude, distance, and frequency ranges, is a constraint on the development of such regional ground motion relations. The NDMA (2011) report circumvented this issue by developing ground motion relations specific to the Indian subcontinent, based on synthetic ground motion database generated by considering the regional geology and seismic source characteristics. However, these relations were presented for class A sites, which are less common in the young mountains of Himachal Pradesh. This became evident from the Multi-channel analysis of Surface Waves (MASW) tests conducted at 73 locations spread across Himachal Pradesh. The results of the MASW tests, which are presented in the Section 2 of this article, indicates that none of the test sites were of class A type, instead most of these sites are of class C and D with a couple of class B sites. This necessitates the need to extend the applicability of the NDMA (2011) ground motion relations to other prevailing site conditions in Himachal Pradesh while incorporating the regional seismological characteristics.

Hence, this article presents four site- and region-specific ground motion relations developed for Himachal Pradesh incorporating the regional seismic characteristics and local site effects. The predictions of these proposed ground motion relations are validated with recorded data from the study region. In the companion paper (Muthuganeisan and Raghukanth 2016), the developed ground motion relations have been used to estimate the site-specific hazard of the state of Himachal Pradesh within a probabilistic framework.

2. SITE CHARACTERIZATION USING MASW TEST

The salient feature of this study is incorporating the local site effects prevailing in HP while estimating the seismic hazard as it is well known that the site conditions significantly alter the ground motion characteristics, resulting

in the amplification of the ground motion in certain frequency ranges. The geographical setting of the HP, aided by the extreme variation in the altitude and climatic conditions, clearly indicates that there is a large spatial variation in the site characteristics across the region. Therefore, the hazard maps produced by NDMA (2011) which are valid for A type sites or hard rock level cannot be used directly. In this study, the dynamic properties of the soil strata has been characterized by developing shear wave velocity profiles of the top soil for a depth of 30 m, which is a good indicator of the elastic soil response (IBC 2009).

In this study, the Multi-channel Analysis of Surface Waves (MASW) technique (Park *et al.* 1999), which is an efficient non-destructive technique, was chosen for estimating the shear wave velocity profile up to 30 m of the soil strata. The active source consist of a sledge hammer, weighing 8 kg, with an automatic trigger struck against a metal plate placed firmly on the ground surface. The striking action generates Rayleigh type surface waves, which are recorded by an array of receivers, called geophones, placed at known distances. The variation of the shear wave velocity across the depth can be found by analyzing these waves in the frequency wave number domain.

Figure 1a shows a schematic view of the test setup and shot location. The experimental setup consists of a 24-channel Geode of ultra-light exploration seismograph class from Geometrics with the following properties: 24-bit, ultra-high resolution 20 kHz bandwidth (8 to 0.02 ms sampling), low distortion (0.0005%), low noise (0.2 µV), stacking accuracy (1/32 of sample interval). The tests were performed with a linear spread of 4.5 Hz frequency geophones with fixtures made for both soft and hard surface. The spread length of the geophones varied from 16 to 72 m with the spacing interval ranging from 1 to 5 m. Sampling interval of 0.125 ms and a record length of 5 s was found to adequately capture the waveforms. Multiple shot gathers were stacked to improve the signal to noise ratio of the data. Also, the geophones located far from the source were set to highest gain to improve the strength of the signal. The geophones are placed in equally spaced linear array and can detect the vertical component of the ground motion caused due to the source. The data from all the geophones are collected by the Geode and are sent as digital records to a computer. A sample waveform of the acquired data with the shot location at –2 m away from the first geophone is shown in Fig. 1b. Since this waveform corresponds to the shot position 1, it can be seen that the first geophone has received the first arrival and the travel time to reach the other geophones can be observed from the arrival of the first breaks of other geophone readings. The signals are then filtered to remove any ambient noise. These signals are then transformed into the frequency-
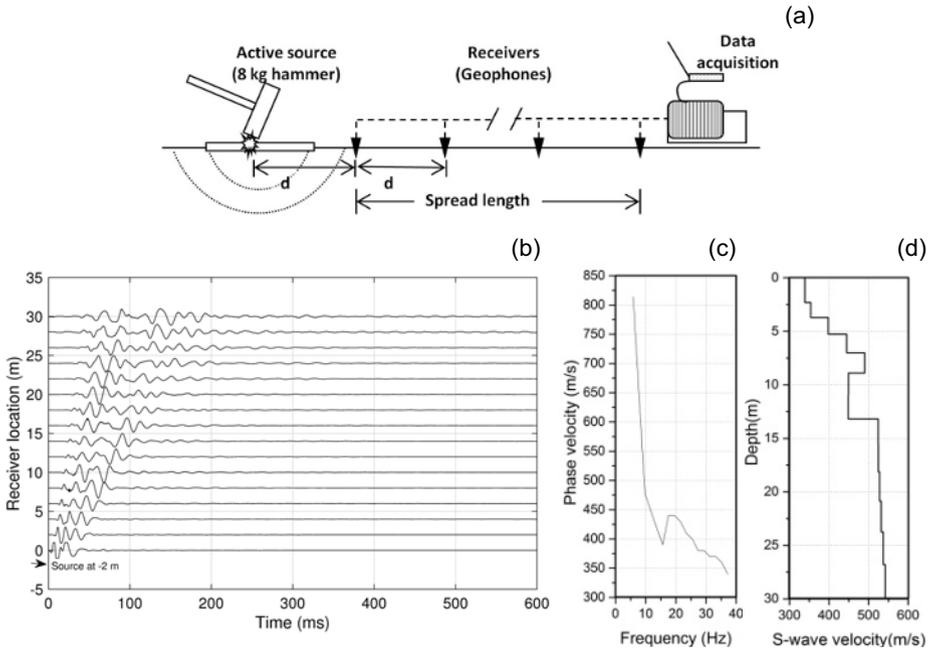
Fig. 1: (a) Schematic view of MASW test, (b) typical waveform of the acquired data. Shot gather at Kullu (KL-4) test site when the source is at –2 m away from the first receiver, (c) the corresponding dispersion curve, (d) the corresponding shear wave velocity profile.
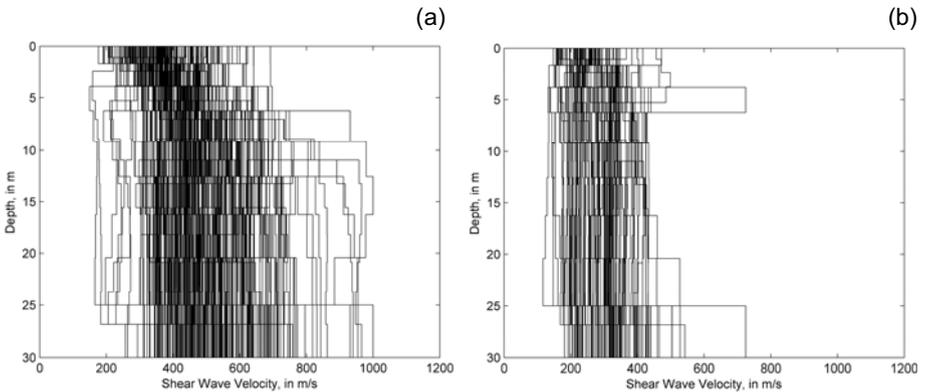

Fig. 2. Vs_{30} profile of all the: (a) 50 class C sites, and (b) 21 class D sites.

phase velocity domain to obtain the dispersion curve, as shown in Fig. 1c, which would reveal the dominant phase velocities *versus* frequency and is an indicator of the variation of the body wave velocities with the soil depth. Using inversion method, the variation of the shear wave velocity across the soil

Site averaged Vs_{30} values and IBC (2009) site classification of all the test sites
in the 22 cities measured by MASW tests

Site id	City name	Location	Average Vs_{30} [m/s]	Soil type
B1		31.34° N 76.76° E	483	C
B2		31.32° N 76.76° E	405	C
B3	Bilaspur*	31.33° N 76.77° E	659	C
B4		31.34° N 76.76° E	626	C
B5		31.34° N 76.76° E	391	C
C1		32.55° N 76.13° E	542	C
C2		32.54° N 76.13° E	326	D
C3	Chamba*	32.55° N 76.13° E	343	D
C4		32.56° N 76.12° E	600	C
C5		32.56° N 76.13° E	291	D
D1	Dankar	32.09° N 78.22° E	413	C
DH1		32.20° N 76.32° E	473	C
DH2		32.20° N 76.34° E	460	C
DH3	Dhramshala*	32.21° N 76.32° E	437	C
DH4		32.22° N 76.33° E	597	C
DH5		32.21° N 76.34° E	432	C
H1		31.69° N 76.52° E	449	C
H2		31.71° N 76.53° E	637	C
H3	Hamirpur*	31.67° N 76.53° E	502	C
H4		31.69° N 76.51° E	227	D
H5		31.68° N 76.54° E	561	C
KP1	Kalpa	31.54° N 78.26° E	394	C
KG1	Kangra fort	32.09° N 76.26° E	853	B
KZ1	Kaza	32.23° N 78.07° E	699	C
KE1	Kee	32.30° N 78.01° E	691	C
KY1	Keylong*	32.57° N 77.03° E	493	C
KY2		32.57° N 77.04° E	300	D
KL1		31.94° N 77.11° E	407	C
KL2		31.96° N 77.11° E	521	C
KL3	Kullu*	31.96° N 77.11° E	467	C
KL4		31.96° N 77.11° E	495	C
KL5		31.96° N 77.12° E	392	C
ML1		32.23° N 77.19° E	416	C
ML2		32.25° N 77.18° E	514	C
ML3	Manali	32.24° N 77.19° E	479	C
ML4		32.25° N 77.19° E	569	C

Table 1 (continuation)

Site id	City name	Location	Average Vs_{30} [m/s]	Soil type
MD1	Mandi*	31.72° N 76.93° E	240	D
MD2		31.71° N 76.94° E	539	C
MD3		31.68° N 76.92° E	352	D
MD4		31.70° N 76.95° E	406	C
NH1	Nahan*	30.56° N 77.30° E	486	C
NH2		30.56° N 77.29° E	407	C
NH3		30.57° N 77.30° E	484	C
NK1	Nako	31.89° N 78.63° E	390	C
RP1	Reckong Peo*	31.54° N 78.27° E	454	C
RP2		31.54° N 78.27° E	348	D
RA1	Rampur	31.44° N 77.63° E	313	D
SL1	Shimla*[+]	31.10° N 77.15° E	525	C
SL2		31.10° N 77.11° E	254	D
SL3		31.09° N 77.14° E	351	D
SL4		31.09° N 77.14° E	779	B
SL5		31.09° N 77.15° E	367	C
SL6		31.11° N 77.16° E	346	D
SL7		31.12° N 77.18° E	352	D
SL8		31.08° N 77.17° E	378	C
SL9		31.10° N 77.17° E	472	C
SL10		31.08° N 77.17° E	296	D
SL11		31.08° N 77.18° E	329	D
SL12		31.09° N 77.21° E	541	C
SL13		31.12° N 77.22° E	494	C
SO1	Solan*	30.92° N 77.09° E	417	C
SO2		30.91° N 77.10° E	469	C
SO3		30.90° N 77.09° E	447	C
SO4		30.89° N 77.07° E	520	C
SO5		30.92° N 77.09° E	452	C
SU1	Sundernagar	31.53° N 76.91° E	379	C
SU2		31.52° N 76.88° E	317	D
T1	Tabo	32.09° N 78.38° E	542	C
U1	Una*	31.47° N 76.27° E	204	D
U2		31.46° N 76.26° E	266	D
U3		31.48° N 76.27° E	183	D
U4		31.47° N 76.28° E	270	D
U5		31.52° N 76.29° E	237	D

Explanations: [+] – capital, * – district headquarters.

layers is obtained by comparing with theoretical dispersion curves iteratively in a least square sense. Thus, a 1D shear wave velocity profile of the middle of the receiver spread is obtained from each shot gather, as shown in Fig. 1d.

The MASW tests were performed at 73 sites located in 22 important cities, including all the 12 district headquarters, of Himachal Pradesh. The tests were performed in open grounds located within the city limits and proximal to very important structures such as hospitals, administrative buildings, bridges, heliports, schools, *etc*. Due to the high sensitivity of the geophones, these tests were performed at minimum traffic noise level. These tests resulted in a collection of 73 one-dimensional Vs_{30} profiles. The test sites were then classified according to the IBC (2009) classification based on the Vs_{30} values, which were calculated as

$$Vs_{30} = \frac{30}{\sum_{i=1}^{n} (d_i / v_i)} , \qquad (1)$$

where n is the number of layers in the top 30 m, d_i and v_i are the thickness [m] and the shear wave velocity [m/s] of the ith layer, respectively. The shear wave velocity profiles of C and D type sites are shown in Fig. 2. The
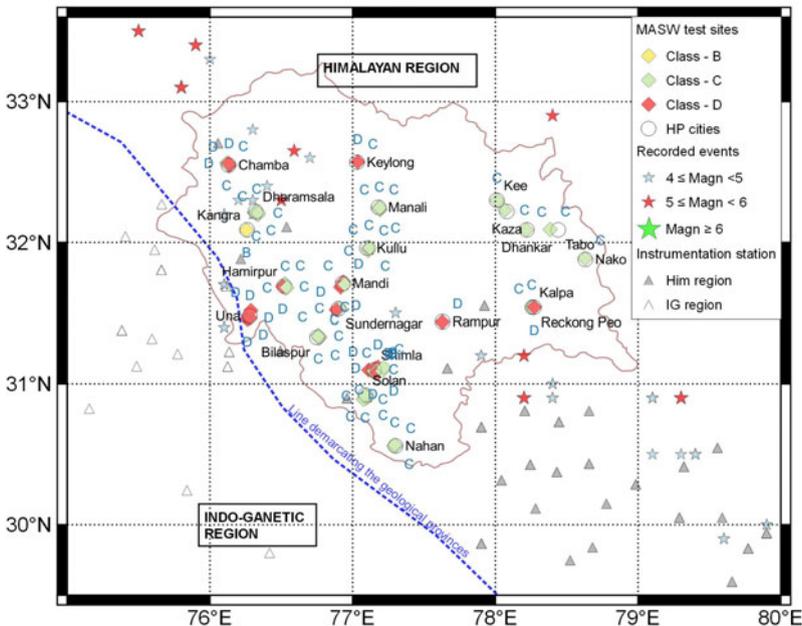
Fig. 3. Location map of the MASW test sites shown along with the site classification based on Vs_{30} values. Also are shown the epicentral location of the recorded events used for GMPE validation along with the locations of the instrumentation stations.

geographic location of the test sites along with the measured Vs_{30} values and the site classification as per IBC (2009) are listed in Table 1.

It can be seen from Table 1 that 50 sites were of class C (360 m/s < Vs_{30} ≤ 760 m/s), 21 sites were of class D (180 m/s < Vs_{30} ≤ 360 m/s), and only 2 sites were of class B (760 m/s < Vs_{30} ≤ 1500 m/s). The spatial distribution of the site classes is indicated in Fig. 3. It can be observed that the test sites inthe Indo-Gangetic region are mainly class D sites whereas the shear wave velocities increase in the higher altitude Himalayan regions where the test sites were mainly of class C. Since the major part of the study area falls under either site class C or D, the existing hazard maps of this region, which are available only for class A sites, cannot be used directly without considering the local site effects. Hence, the site response function have been estimated using these shear wave velocity profiles and are used for developing ground motion relations specific to the class C and D site conditions found in this region.

3.   GROUND MOTION RELATIONS

The reliability of the estimation of seismic hazard at a site depends on how well the uncertainties associated with the source to site path are considered while estimating the ground motion level at the site. These uncertainties associated with the medium characteristics are handled through ground motion relations, which describe the average or other moments of the random hazard parameter (namely, median peak ground acceleration) in terms of magnitude and distance. In this study, the regional properties such as lithographic structure, quality factor (damping); source parameters such as dip angle, focal depth, *etc*., and soil amplification effects are taken into consideration while developing GMPEs specific to this region. In order to develop these relations, an extensive database of ground motion measures (PGA and Sa(T) values in this case) for different combinations of magnitude, distance, site conditions, source mechanism, *etc*., is required. However, the instrumental database for Himalayan and Indo-Gangetic region is sparse (shown in Fig. 4) and is insufficient to derive regional attenuation relationships. Hence, the ground motion relations for such regions have to be based on simulated ground motions, instead of a few past recordings. Stochastic seismological model originally proposed by Boore (1983) is a viable alternative and has been used worldwide for deriving attenuation relationships in regions lacking strong motion data (Atkinson and Boore 1995, Hwang and Huo 1997, Toro *et al.* 1997). The methodology of generating the synthetic ground motion database incorporating the local site effects, using the shear wave velocity profiles developed in the previous section, and developing regional site-specific ground motion relations for the study are discussed below.
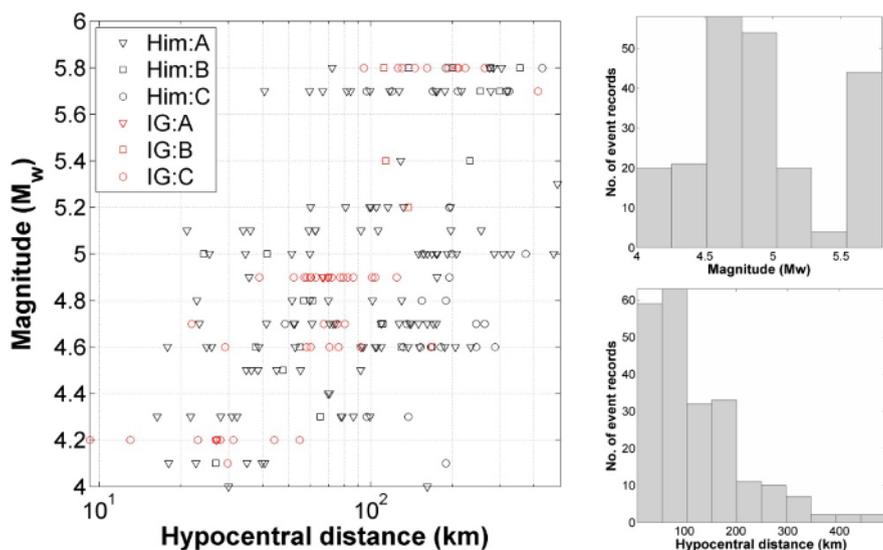
Fig. 4. Distribution of ground motion record based on magnitude, distance, geology, and site class ("Him" refers to Himalayan region and "IG" refers to Indo-Gangetic region).

3.1 Stochastic seismological model

In this study, the synthetic ground motion spectra specific to the study region are generated using a finite source model of Motazedian and Atkinson (2005), which extends the point source model of Boore (1983), with the controlling parameters selected based on regional characteristics. These input parameters, namely quality factor, focal depth, and stress drop vary from region to region, depending on the seismo-tectonic setup and geology. The Indian subcontinent is broadly divided into seven geological regions in NDMA (2011) with differing quality factors to estimate ground motions. Among these seven regions, Himalaya and Indo-Gangetic plains control the hazard at Himachal Pradesh whose important regional parameters are listed in Table 2. These regional parameters have been used in the stochastic seismological model, which will be discussed further here.

In this method, a grid of point sources, each of which denotes a sub-fault, represents a rectangular fault plane. The Fourier amplitude spectrum of ground motion $A(r, f)$ due to the jth subfault at a site is derived from the point source seismological model, expressed as

$$A_j(r, f) = C \cdot H_j \cdot S_j(f) \cdot F(f) \cdot D(r, f) \cdot P(f) , \qquad (2)$$

where C is a scaling factor, $S_j(f)$ is the source spectral function, $D(r, f)$ is the diminution function characterizing the quality of the region, $P(f)$ is a filter

Table 2
Uncertainties in earthquake model parameters

Region	Stress drop $\Delta\sigma$ [bars]	Dip [°]	Focal depth [km]	Reference
Himalaya	50-200	2°-30°	5-40	Kayal (2008)
Indo-Gangetic Plain	50-200	10°-80°	5-40	Kayal (2008)

function, $F(f)$ is the site dependent function that modifies the bed rock motion in the vertical direction, and H_j is a scaling factor used for conserving the energy of high-frequency spectral level of sub-faults. Modification between bedrock and A type site is a linear problem in one-dimension and hence for such sites, amplification function $F(f)$ in Eq. 2 can be directly found by the quarter-wavelength method of Boore and Joyner (1997).

In this study, the principal source model $S_j(f)$ for a jth source, developed by Brune (1970), is taken as a function of the seismic moment (M_{0j}) and the corner frequency f_{0j} is given below as follows:

$$S_j(f) = (2\pi f)^2 \frac{M_{0j}}{\left[1 + \left(f/f_{0j}\right)^2\right]} . \tag{3}$$

The three important seismic source parameters, namely M_{0j}, f_{0j}, and the stress drop $(\Delta\sigma)$, are related through

$$f_{0j} = 4.9 \times 10^6 \left(N_{Rj}\right)^{-1/3} N^{1/3} V_s \left(\frac{\Delta\sigma}{M_{0j}}\right)^{1/3} , \tag{4}$$

where V_s stands for the shear wave velocity in the source region, corresponding to bedrock conditions, and N_{Rj} is the cumulative number of ruptured sub-faults by the time rupture reaches the jth subfault. The shear wave velocity and density at the focal depth are fixed at 3.6 km/s and 2900 kg/m³, respectively, for both the active regions corresponding to compressed hard granite (Singh $et\ al.$ 1999). The regional velocity model used for both the active regions is the one reported by NDMA (2011). The spatial modifying function $D(f)$ is given by

$$D(r, f) = G \exp\left(\frac{-\pi fr}{V_s Q(f)}\right) , \tag{5}$$

where G is the geometric attenuation factor. The exponential term denotes anelastic attenuation with hypocentral distance r and the quality factor Q. The spatial spread of the ground motion depends sensitively on the quality

factor of the local region. The quality factor for the Himalayan region has been reported as $253f^{0.8}$ by Singh *et al.* (2004) after analysing the ground motion from two major earthquakes that occurred in this region, *viz.*, the 1991 Uttarkashi earthquake and the 1999 Chamoli earthquake. Likewise, Mohanty *et al.* (2009) have reported the quality factor for Indo-Gangetic plain as $142f^{1.04}$ based on the strong motion data of 55 earthquakes of magnitude 1.5 to 4.2 acquired from Delhi telemetric network. The spatial modifier of Eq. 5 consists of a general geometrical attenuation term G which is taken as (Singh *et al.* 1999)

$$G = \begin{cases} 1/r \; , & \text{for } r \leq 100 \\ 1/10\sqrt{r}, & \text{for } r > 100 \end{cases}. \tag{6}$$

The constant C of Eq. 2 is

$$C = \frac{\langle R_{\theta\varphi} \rangle \sqrt{2}}{4\pi\rho V_s^3} \;, \tag{7}$$

where ρ is the material density at the focal depth. The *S*-wave radiation coefficient ($<R_{\theta\phi}>$) in the above Eq. 7 varies randomly within particular intervals and is taken to be in the interval 0.48-0.64 (Boore and Boatwright 1984). The coefficient $\sqrt{2}$ in the above equation arises as the product of the free surface amplification and partitioning of energy in orthogonal directions. The scaling factor for *j*th sub-fault, H_j, based on the squared acceleration spectrum is taken as (Boore 2009)

$$H_j = \sqrt{N} \frac{f_0^2}{f_{0j}^2} \;, \tag{8}$$

where f_0 is the corner frequency at the end of the rupture which can be obtained by substituting $N_R(t) = N$ in Eq. 8. The filter function $P_j(f)$ is taken as

$$P_j(f) = \frac{\sqrt{N}}{H_j} \frac{1 + \left(f/f_{0j}\right)^2}{1 + \left(f/f_0\right)^2} \;. \tag{9}$$

The moment of *j*th subfault is computed from the slip distribution as follows

$$M_{0j} = \frac{M_0 D_j}{\sum_{j=1}^{N} D_j} \;. \tag{10}$$

Here, D_j is the average final slip acting on the *j*th subfault and M_0 is the total seismic moment on the fault. To further account for earthquake rupture, Motazedian and Atkinson (2005) introduced the concept of pulsing area where

the cumulative number of active subfaults (N_{Rj}) increases with time after the initiation of rupture and becomes constant at some fixed percentage of the total rupture area. This parameter determines the number of active subfaults during the rupture of the jth subfault. As a result, N_{Rj} subfaults will be considered for computing the corner frequency f_{0j} given in Eq. 4. The above is a general finite source model expressed in the frequency domain, valid for any region if only the various controlling parameters can be selected suitably. In addition to the above parameters, the pulsing percentage area required in the simulation is varied from 25 to 75%.

3.2 Sample ground motion

From the above discussion, it is seen that once the stress drop and quality factors for the region are known, then the Fourier amplitude spectrum of ground acceleration for a given magnitude (Mw) and hypocentral distance (r) can be expressed within the limitations of a finite source seismological model. An ensemble of accelerograms can be retrieved from Eq. 2 in three steps. First, a Gaussian stationary white noise sample of length equal to the strong motion duration (Boore and Atkinson 1987)

$$T = \left(1/f_c + 0.05r\right) \tag{11}$$

is simulated for each subfault. This sample is multiplied by a suitable non-stationary modulating function suggested by Saragoni and Hart (1973). The simulated sample is Fourier transformed into the frequency domain. This Fourier spectrum is normalized by its root mean-square value and multiplied by the seismological source-path-site function $A(f)$ of Eq. 2. The resulting function is transformed back into the time domain to get a sample of the ground motion accelerogram for each subfault. The simulated acceleration time histories for all the sub-faults are summed up with a time delay of Δt_j that accounts for the rupture propagation, to obtain the final ground acceleration expressed as

$$a(t) = \sum_{j=1}^{N} a_j\left(t + \Delta t_j\right). \tag{12}$$

This way, a synthetic database of ground motion using the regional properties has been simulated for all the necessary magnitude and distance ranges. The spectral values of acceleration is then computed from response spectra corresponding to 5% damping.

3.3 Site amplification function

In addition to the uncertainties in the magnitudes and epicentral distances, it is also important to consider the variations in the soil profiles for simulating

the spectral accelerations at the ground surface. It is observed from Table 1 that the test sites in HP can be classified into class C and D. The general approach of finite source seismological model described above can be extended to these two site classes with the help of depth sections sampled from the region with known Vs_{30} values. The modification of ground motion between the bedrock and the surface of the class A sites is a linear problem in one dimension. However, in the case of class B, C, and D type profiles, the soil layering, viscoelastic properties, and nonlinear effects are important. Therefore, the modification of ground motion between the class A and the overlying soil can be treated as a one-dimensional nonlinear problem and the site amplification can be directly found by means of a nonlinear site response analysis. The equivalent linear site response analysis method of Idriss and Sun (1992) implemented in the SHAKE91 software has been used in this study. The main assumption in this method is that the response of a soil deposit to an upward propagating shear waves from the underlying rock formation to the overlying soil mass is analysed to obtain the frequency response of the site dependent function $F(f)$. Therefore, the acceleration time histories are first generated for class A rock profiles from the seismological model discussed earlier and are used as the input ground motion at basement for the class C and D type profiles. Thus, for a specific site, modification of the bedrock results is possible by considering the dynamic characteristics of the soil strata. The nonlinear behaviour of the soil, characterized in terms of reduction of shear modulus and increase of damping ratio with shear strain, can be accounted by using equivalent linear soil properties. In this study, the standard curves for shear modulus reduction *versus* shear strain curves and damping ratio *versus* shear strain curves proposed by Vucetic and Dobry (1991) for different soil types are used. The frequency response function (FRF) defined as the surface response for a rock level sinusoidal motion of unit amplitude is estimated for all the 73 shear wave velocity profiles. These functions for both class C and D sites are shown in Fig. 5. Also, it can be seen from Fig. 6 that the Vs_{30} values of majority of these sites lie within 330-550 m/s range. This can be attributed to the fact that these test sites, essentially located within urban centres in these mountainous regions, had been prepared by slope levelling and land filling for urban uses. Hence, only moderate differences in the site response functions are observed in certain frequency ranges.

Hence, the acceleration time histories in the synthetic database generated for class A rock profiles can be modified to surface ground motions by using the evaluated class C and D site response function. These functions shall be used to generate a database of synthetic surface level response spectra as discussed in the following section.
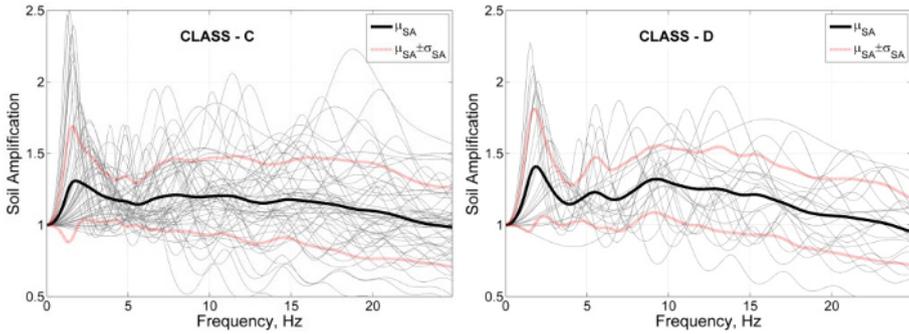
Fig. 5. Mean frequency response function of site amplification for the 50 class C sites and 21 class D sites.
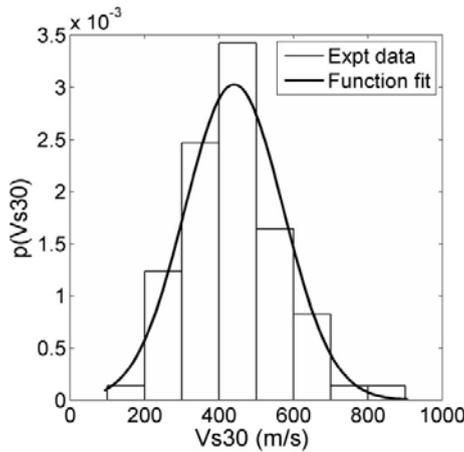
Fig. 6. Distribution of the measured Vs_{30} values.

3.4 Synthetic ground motion database

In this study, horizontal components of the spectral acceleration values have been simulated for moment magnitude (Mw) ranging from 4 to 8.5 in increments of 0.5 units, at 20 values of hypocentral distances ranging from 1 to 500 km. To capture the finiteness of the source, the ground motions are also simulated for eight azimuths ranging from 0° to 315° in increments of 45°. Thus, a total number of 160 distance samples are considered for each magnitude. In all, there are 1600 pairs of magnitudes and distances. Since the stress drop, focal depth, dip, radiation coefficient, and pulsing percentage area are random variables, we have included the uncertainty arising out of these parameters also. It can be observed from Table 1 that there are 50 class C sites and 21 class D sites in HP. The uncertainty associated with the local site amplification has to be included in the ground motion relations. Fifty

samples of stress drop, focal depth, dip, radiation coefficient, pulsing percentage area, and shear wave velocity profiles are generated. These samples are combined using the Latin Hypercube sampling technique (Iman and Conover 1980) to select fifty random seismic parameters sets for each magnitude value. The acceleration time histories are first simulated for class A site using the amplification function reported in NDMA (2011). The simulated time histories combined with the developed shear wave velocity profiles are used as input to SHAKE91 to obtain the surface level response spectrum. Thus, a database of 80 000 PGA and Sa samples corresponding to 1600 simulated earthquake events are generated assuming 5% damping. This synthetic database is developed separately for both Himalayan and Indo-Gangetic regions for both class C and D type site conditions.

3.5 Ground motion prediction equations (GMPE)

The study region comprises both Himalaya and Indo-Gangetic (IG) plain; the differences in the faulting mechanisms in these two regions result in different characteristic ground motions. Therefore, it becomes necessary to develop separate relations for both class C and D sites in these two geological regions. Hence, the simulated surface-level response spectra for both class C and D sites of both the geological regions are used to develop the ground motion relations. Using the simulated database of 80 000 samples of surface level ground motion, the ground motion prediction equation (GMPE) reported by the NDMA (2011) can be updated to predict the surface level hazard measure in the Himachal Pradesh region. The functional form of the GMPE is adopted from NDMA (2011) as

$$\ln\left(\frac{S_a}{g}\right) = c_1 + c_2 M + c_3 M^2 + c_4 r + c_5 \ln\left(r + c_6 e^{c_7 M}\right) + c_8 \log(r) f_0 + \ln(\varepsilon)$$. (13)
$$f_0 = \max\left(\ln(r/1\,00), 0\right)$$

Here, S_a is the spectral acceleration, M is the moment magnitude, and r is the hypocentral distance in km. This form of the attenuation accounts for geometrical spreading, anelastic attenuation and magnitude saturation. Hypocentral distance has been used as the distance measure since reliable estimates of other distance measures that depend on the rupture characteristics are not available for this region. The coefficients $(c_1, …, c_8)$ are evaluated from the simulated database of 80 000 samples using a two-step stratified regression analysis procedure proposed by Joyner and Boore (1981). These coefficients for the class C and D sites in both the Himalayan and IG region are reported in Tables 3 to 6 for 28 different time periods. The sufficiency of the functional form of the presented GMPE to represent the simulated data was tested by examining the trend of the residuals. Figure 7 shows the distribu-

Coefficients of the GMPE for class C sites in the Himalayan region

Period [s]	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	$\sigma(\varepsilon)$
0	−4.7738	1.5004	−0.0204	−0.0048	−1.4216	0.0465	0.8748	0.0912	0.3864
0.01	−4.7739	1.5004	−0.0204	−0.0048	−1.4216	0.0465	0.8748	0.0912	0.3863
0.015	−4.3856	1.4517	−0.0166	−0.0048	−1.4447	0.0526	0.8614	0.0822	0.4077
0.02	−3.9633	1.3831	−0.0117	−0.0048	−1.4506	0.0512	0.8648	0.0732	0.4246
0.03	−3.3644	1.2550	−0.0029	−0.0050	−1.4188	0.0385	0.8975	0.0661	0.4248
0.04	−3.2107	1.1993	0.0002	−0.0053	−1.3663	0.0265	0.9381	0.0721	0.4138
0.05	−3.1912	1.1821	0.0007	−0.0056	−1.3237	0.0189	0.9734	0.0805	0.4050
0.06	−3.2716	1.1903	−0.0007	−0.0057	−1.2895	0.0154	0.9926	0.0870	0.3962
0.075	−3.5114	1.2357	−0.0046	−0.0057	−1.2687	0.0141	0.9988	0.0922	0.3877
0.09	−3.7079	1.2796	−0.0082	−0.0057	−1.2572	0.0140	0.9956	0.0960	0.3836
0.1	−3.9358	1.3384	−0.0128	−0.0056	−1.2546	0.0155	0.9808	0.0987	0.3792
0.15	−5.1125	1.6079	−0.0328	−0.0054	−1.2308	0.0186	0.9500	0.1028	0.3688
0.2	−6.2581	1.8845	−0.0531	−0.0052	−1.2201	0.0251	0.9083	0.1013	0.3641
0.3	−8.5491	2.4630	−0.0947	−0.0049	−1.2018	0.0419	0.8384	0.1008	0.3623
0.4	−10.4802	2.9546	−0.1292	−0.0046	−1.1930	0.0632	0.7866	0.0988	0.3628
0.5	−12.1393	3.3690	−0.1578	−0.0045	−1.1906	0.0800	0.7565	0.0964	0.3609
0.6	−13.8153	3.7663	−0.1852	−0.0043	−1.1786	0.0906	0.7363	0.0953	0.3620
0.7	−14.8718	4.0333	−0.2024	−0.0042	−1.1982	0.1174	0.7123	0.0951	0.3591
0.75	−15.4146	4.1593	−0.2106	−0.0041	−1.2007	0.1232	0.7060	0.0941	0.3587
0.8	−15.9659	4.2939	−0.2190	−0.0041	−1.2118	0.1323	0.6998	0.0978	0.3564
0.9	−17.0514	4.5410	−0.2348	−0.0040	−1.2160	0.1491	0.6892	0.0996	0.3539
1	−17.5645	4.6512	−0.2416	−0.0039	−1.2171	0.1562	0.6857	0.0975	0.3519
1.2	−19.1139	4.9647	−0.2602	−0.0038	−1.2145	0.1519	0.6889	0.0987	0.3423
1.5	−20.4006	5.2107	−0.2722	−0.0037	−1.2349	0.1624	0.6888	0.1004	0.3309
2	−21.8815	5.4498	−0.2809	−0.0035	−1.2588	0.1626	0.6999	0.1000	0.3180
2.5	−22.7389	5.5038	−0.2787	−0.0034	−1.2442	0.1228	0.7327	0.1017	0.3128
3	−23.2798	5.4962	−0.2700	−0.0033	−1.2701	0.1139	0.7527	0.1059	0.3086
4	−23.5964	5.3489	−0.2483	−0.0032	−1.2979	0.1001	0.7809	0.1065	0.3093
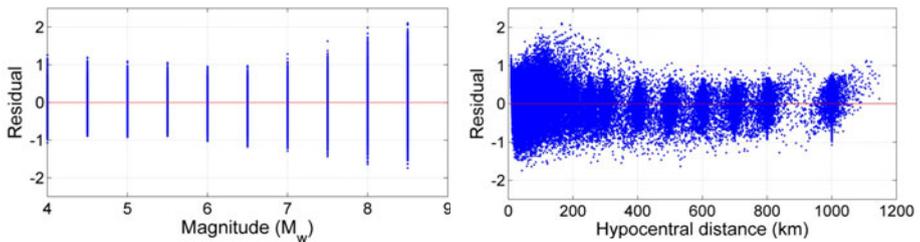
Fig. 7. Residuals [= log(simulated PGA) − log(predicted PGA)] plotted as a function of predictor variables M and R.

Coefficients of the GMPE for class D sites in the Himalayan region

Period [s]	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	$\sigma(\varepsilon)$
0	–4.5304	1.4623	–0.0175	–0.0049	–1.4392	0.0365	0.8948	0.0985	0.3881
0.01	–4.5308	1.4624	–0.0175	–0.0049	–1.4392	0.0366	0.8947	0.0985	0.3881
0.015	–4.1180	1.4086	–0.0134	–0.0049	–1.4658	0.0388	0.8877	0.0920	0.4027
0.02	–3.7021	1.3428	–0.0087	–0.0049	–1.4735	0.0379	0.8899	0.0841	0.4161
0.03	–3.1443	1.2262	–0.0008	–0.0051	–1.4423	0.0299	0.9171	0.0759	0.4176
0.04	–2.9824	1.1712	0.0023	–0.0054	–1.3910	0.0216	0.9527	0.0807	0.4097
0.05	–2.9733	1.1591	0.0025	–0.0056	–1.3503	0.0164	0.9806	0.0884	0.4033
0.06	–3.0671	1.1712	0.0007	–0.0058	–1.3091	0.0130	1.0029	0.0934	0.3964
0.075	–3.2796	1.2130	–0.0029	–0.0058	–1.2844	0.0119	1.0099	0.0982	0.3909
0.09	–3.4798	1.2585	–0.0066	–0.0058	–1.2750	0.0119	1.0068	0.1015	0.3872
0.1	–3.7006	1.3144	–0.0110	–0.0057	–1.2695	0.0130	0.9928	0.1034	0.3835
0.15	–4.9730	1.5972	–0.0319	–0.0054	–1.2472	0.0156	0.9636	0.1059	0.3750
0.2	–6.1367	1.8725	–0.0521	–0.0052	–1.2287	0.0200	0.9277	0.1026	0.3711
0.3	–8.5262	2.4675	–0.0948	–0.0049	–1.2106	0.0330	0.8597	0.1025	0.3700
0.4	–10.4727	2.9671	–0.1300	–0.0047	–1.1957	0.0471	0.8125	0.1002	0.3713
0.5	–12.1248	3.3786	–0.1584	–0.0044	–1.1870	0.0575	0.7864	0.0964	0.3698
0.6	–13.8033	3.7733	–0.1856	–0.0043	–1.1740	0.0666	0.7629	0.0948	0.3699
0.7	–14.8673	4.0347	–0.2024	–0.0041	–1.1916	0.0844	0.7419	0.0936	0.3668
0.75	–15.4015	4.1591	–0.2106	–0.0041	–1.1962	0.0909	0.7333	0.0934	0.3664
0.8	–15.9497	4.2928	–0.2189	–0.0041	–1.2086	0.0995	0.7248	0.0968	0.3636
0.9	–17.0185	4.5281	–0.2338	–0.0040	–1.2102	0.1098	0.7169	0.0984	0.3612
1	–17.5410	4.6380	–0.2405	–0.0039	–1.2095	0.1122	0.7166	0.0968	0.3587
1.2	–19.0391	4.9414	–0.2582	–0.0038	–1.2133	0.1120	0.7183	0.0986	0.3489
1.5	–20.2986	5.1796	–0.2697	–0.0037	–1.2349	0.1217	0.7160	0.1005	0.3380
2	–21.7272	5.4074	–0.2774	–0.0035	–1.2636	0.1270	0.7236	0.1018	0.3251
2.5	–22.5444	5.4556	–0.2747	–0.0034	–1.2543	0.0989	0.7538	0.1038	0.3204
3	–23.0657	5.4404	–0.2654	–0.0033	–1.2791	0.0915	0.7739	0.1081	0.3166
4	–23.3445	5.2846	–0.2430	–0.0032	–1.3093	0.0819	0.8013	0.1096	0.3174

tion of residuals (difference between the logarithm of the simulated and pre-dicted PGA values) with respect to predictor variables. Since no visible trend or bias can be observed, the present functional form has been found to suffi-ciently capture the simulated data. These developed GMPEs can be used to predict the horizontal component of the acceleration values within the study region for a given M and R value.

Table 5

Coefficients of the GMPE for class C sites in the Indo-Gangetic region

Period [s]	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	$\sigma(\varepsilon)$
0	−4.3477	1.2807	−0.0087	−0.0064	−1.2940	0.0084	1.0489	0.1234	0.3930
0.01	−4.3478	1.2805	−0.0087	−0.0064	−1.2938	0.0084	1.0491	0.1234	0.3929
0.015	−3.9582	1.2157	−0.0042	−0.0063	−1.2957	0.0077	1.0595	0.1224	0.4352
0.02	−3.5931	1.1596	−0.0004	−0.0062	−1.2917	0.0072	1.0684	0.1228	0.4514
0.03	−3.1401	1.1051	0.0032	−0.0062	−1.2830	0.0071	1.0705	0.1204	0.4289
0.04	−3.0207	1.1033	0.0032	−0.0062	−1.2733	0.0070	1.0722	0.1185	0.4079
0.05	−3.0180	1.1225	0.0020	−0.0063	−1.2642	0.0067	1.0758	0.1186	0.3960
0.06	−3.1225	1.1567	−0.0007	−0.0063	−1.2528	0.0067	1.0739	0.1160	0.3869
0.075	−3.3532	1.2057	−0.0043	−0.0063	−1.2427	0.0067	1.0719	0.1156	0.3806
0.09	−3.5506	1.2520	−0.0078	−0.0064	−1.2367	0.0071	1.0630	0.1151	0.3772
0.1	−3.7684	1.3094	−0.0118	−0.0064	−1.2395	0.0086	1.0398	0.115	0.3749
0.15	−4.9516	1.5779	−0.0311	−0.0065	−1.2213	0.0108	1.0066	0.1134	0.3709
0.2	−6.1330	1.8563	−0.0512	−0.0065	−1.2076	0.0142	0.9678	0.1089	0.3703
0.3	−8.4797	2.4413	−0.0928	−0.0065	−1.1871	0.0239	0.8986	0.1059	0.3748
0.4	−10.4117	2.9257	−0.1264	−0.0066	−1.1766	0.0365	0.8467	0.1029	0.3785
0.5	−12.0609	3.3261	−0.1538	−0.0066	−1.1689	0.0434	0.8233	0.0997	0.3793
0.6	−13.7506	3.7191	−0.1808	−0.0066	−1.1520	0.0477	0.8059	0.0974	0.3813
0.7	−14.7700	3.9596	−0.1960	−0.0066	−1.1624	0.0567	0.7902	0.0955	0.3791
0.75	−15.2966	4.0758	−0.2036	−0.0066	−1.1613	0.0577	0.7866	0.0946	0.3798
0.8	−15.8293	4.2001	−0.2113	−0.0066	−1.1700	0.0615	0.7806	0.0978	0.3783
0.9	−16.8659	4.4149	−0.2247	−0.0067	−1.1629	0.0606	0.7828	0.0982	0.376
1	−17.3530	4.5072	−0.2303	−0.0066	−1.1569	0.0575	0.7889	0.0956	0.3741
1.2	−18.7600	4.7687	−0.2451	−0.0067	−1.1497	0.0490	0.8061	0.0976	0.3658
1.5	−19.9378	4.9550	−0.2532	−0.0067	−1.1508	0.0392	0.8361	0.0980	0.3554
2	−21.2123	5.0947	−0.2551	−0.0066	−1.1485	0.0248	0.8976	0.0962	0.3431
2.5	−21.8921	5.0838	−0.2484	−0.0066	−1.1250	0.0113	0.9888	0.0962	0.3388
3	−22.2566	4.9833	−0.2334	−0.0066	−1.1222	0.0046	1.1077	0.0965	0.3350
4	−22.2759	4.6884	−0.2015	−0.0066	−1.1090	0.0006	1.3616	0.0935	0.3321

4. VALIDATION OF THE DEVELOPED GMPES

The reliability and the limitation of the developed GMPE can be understood by comparing its predictions with that of the recorded data. The event records (within the considered seismogenic zones) are taken from the Indian national strong motion instrumentation network (Mittal *et al.* 2012) available at the PESMOS database (http://pesmos.in). From the database, the events with magnitude $M \geq 4$ recorded by stations located within HP within a 500 km region around HP were considered. A total of 221 records with mag-

Coefficients of the GMPE for class D sites in the Indo-Gangetic region

Period [s]	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	$\sigma(\varepsilon)$
0	−4.3343	1.2847	−0.0090	−0.0064	−1.2983	0.0086	1.0439	0.1238	0.3941
0.01	−4.3345	1.7285	−0.0090	−0.0064	−1.2982	0.0086	1.0441	0.1238	0.3940
0.015	−3.9775	1.2241	−0.0047	−0.0063	−1.3000	0.0081	1.0529	0.1226	0.4275
0.02	−3.6236	1.1680	−0.0008	−0.0063	−1.2965	0.0076	1.0612	0.1230	0.4427
0.03	−3.1667	1.1100	0.0030	−0.0062	−1.2878	0.0074	1.0647	0.1203	0.4256
0.04	−3.0243	1.1055	0.0031	−0.0062	−1.2788	0.0073	1.0659	0.1187	0.4080
0.05	−3.0117	1.1228	0.0020	−0.0063	−1.2702	0.0069	1.0706	0.1193	0.3979
0.06	−3.0916	1.1569	−0.0006	−0.0063	−1.2569	0.0068	1.0699	0.1164	0.3903
0.075	−3.2895	1.2047	−0.0042	−0.0063	−1.2450	0.0068	1.0686	0.1158	0.3849
0.09	−3.4831	1.2501	−0.0076	−0.0064	−1.2402	0.0071	1.0604	0.1155	0.3821
0.1	−3.6939	1.3068	−0.0116	−0.0064	−1.2408	0.0085	1.0390	0.1149	0.3802
0.15	−4.9172	1.5764	−0.0310	−0.0065	−1.2254	0.0106	1.0068	0.1138	0.3769
0.2	−6.0931	1.8523	−0.0509	−0.0065	−1.2077	0.0135	0.9713	0.1088	0.3764
0.3	−8.4950	2.4461	−0.0931	−0.0066	−1.1901	0.0226	0.9023	0.1066	0.3807
0.4	−10.4342	2.9399	−0.1275	−0.0066	−1.1754	0.0343	0.8500	0.1030	0.3845
0.5	−12.0682	3.3399	−0.1549	−0.0066	−1.1650	0.0406	0.8268	0.0991	0.3852
0.6	−13.7375	3.7257	−0.1814	−0.0066	−1.1487	0.0438	0.8114	0.0970	0.3871
0.7	−14.7482	3.9608	−0.1962	−0.0066	−1.1602	0.0519	0.7969	0.0953	0.3846
0.75	−15.2670	4.0734	−0.2035	−0.0066	−1.1597	0.0528	0.7935	0.0945	0.3852
0.8	−15.7935	4.1946	−0.2110	−0.0066	−1.1684	0.0564	0.7872	0.0977	0.3835
0.9	−16.8245	4.4071	−0.2243	−0.0067	−1.1622	0.0559	0.7887	0.0983	0.3809
1	−17.3089	4.4987	−0.2298	−0.0066	−1.1567	0.0533	0.7944	0.0960	0.3789
1.2	−18.6974	4.7546	−0.2442	−0.0067	−1.1504	0.0455	0.8116	0.0979	0.3704
1.5	−19.8625	4.9382	−0.2520	−0.0067	−1.1526	0.0371	0.8395	0.0985	0.3601
2	−21.1173	5.0736	−0.2536	−0.0066	−1.1519	0.0240	0.8987	0.0973	0.3477
2.5	−21.7819	5.0600	−0.2467	−0.0066	−1.1299	0.0112	0.9874	0.0974	0.3432
3	−22.1434	4.9604	−0.2319	−0.0066	−1.1280	0.0048	1.0999	0.0977	0.3393
4	−22.1664	4.6684	−0.2002	−0.0066	−1.1153	0.0007	1.3416	0.0949	0.3362

nitude and hypocentral distance ranging from *Mw* 4 to 5.8 and 9.25 to 489 km that fall within the study region. Of the total records, the numbers of events reported in the Himalayan and IG region were 165 and 56, respectively. It should be noted that the site classifications of the strong motion stations have been reported based on the overall geology of the region and physical description of the near surface materials and are not based on actual *in situ* characterization (Mittal *et al.* 2012). In the Himalayan region, recorded data were reported at 117, 20, and 28 numbers of class A, B, and C

sites. While the numbers of data records reported at class A, B, and C sites of the Indo-Gangetic region were 4, 3, and 49. None of the above records were reported at class D sites. The histograms of the magnitude and distance of the recorded data are shown in Fig. 4.

In the absence of proper site characterization and lack of sufficient data for all the site classes, it has been decided to perform the residual analysis by considering all the recorded dataset in both the Himalayan and IG region to be of class C type. The residuals were estimated as the difference between the spectral values of the recorded ground motion and the predicted values for the same reported magnitude and hypocentral distance using the developed ground motion relations. The mean and standard deviation of the residuals estimated for different spectral values, ranging from 0.01 to 4 s time period, are plotted in Fig. 8 for the two geological regions. It can be seen that there is a reasonable agreement between the GMPE predictions and the observed data. It can also be observed that the GMPE predicts higher ground motion in time periods below 0.1 s and lower ground motion in the long pe-
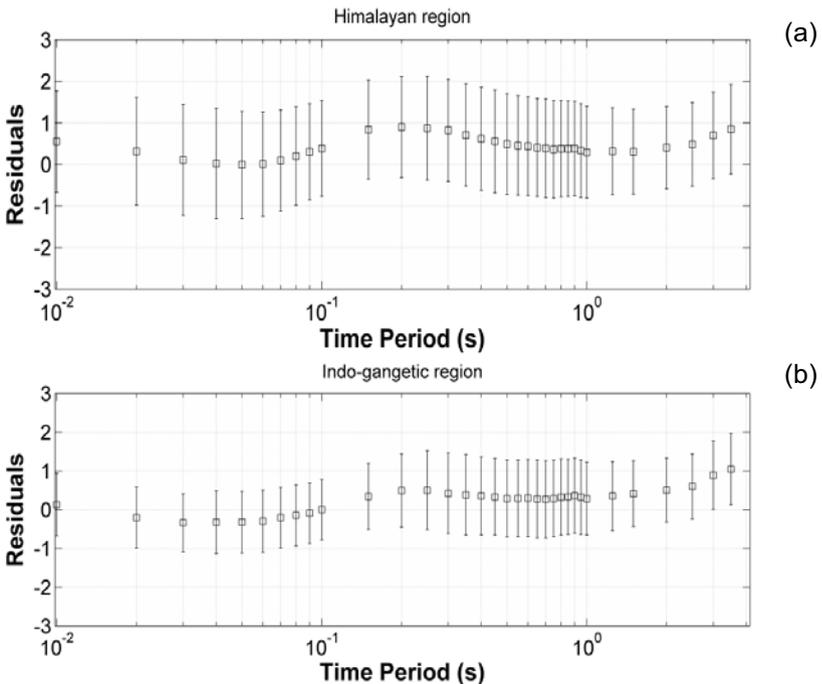
Fig. 8. Mean and standard deviation of the residuals depicted as a function of time period for the Himalayan (a) and the Indo-Gangetic region (b). The residuals are computed as the difference between the logarithm of the spectral values of the observed data and that of the predicted data.
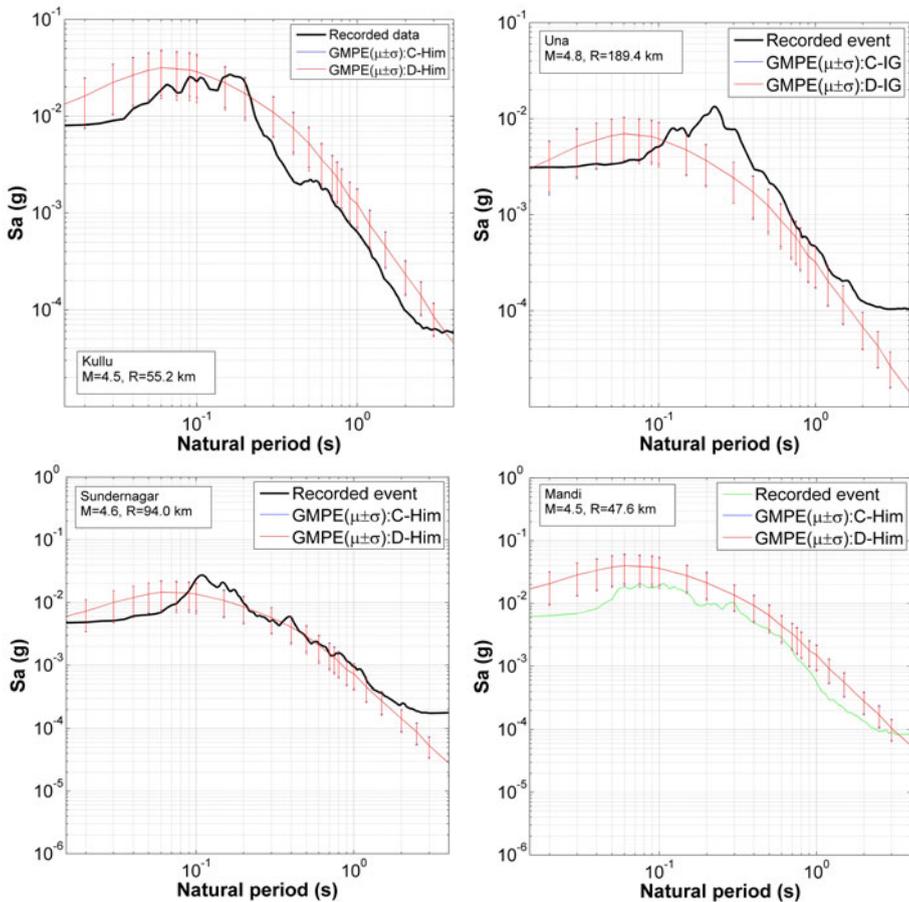
Fig. 9. Comparison of the spectral acceleration (5% damping) of observed events with the GMPE prediction for some of the test cities.

riod range of above 1.5 s. The comparison between the 5% damped response spectrum of recorded data at some of the instrumented stations located at 4 test cities and the spectral values predicted by the developed ground motion relations are shown in Fig. 9. It can be seen that there is a good agreement between the predicted and the recorded data with very close PGA predictions at some stations.

In order to highlight the importance of developing site-specific ground motion relations, the prediction of the GMPEs developed in this study are compared with that of the NDMA (2011) developed for class A site conditions as shown in Fig. 10. The comparison made for a moderate as well as a
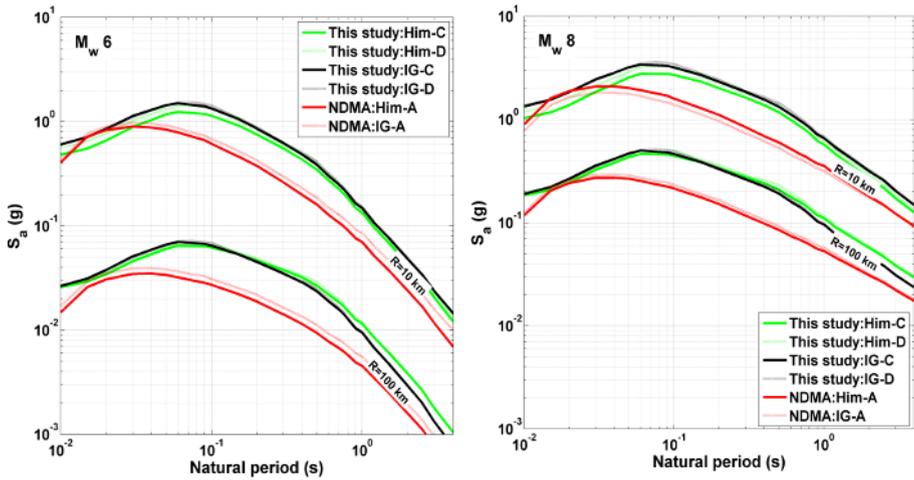
Fig. 10. Comparison of the predicted spectral values of the developed GMPEs (for class C and D sites in the Himalayan and IG region) with GMPE reported by NDMA (2011; applicable only for class A sites). The spectra are shown for $M_w = 6, 8$ and hypocentral distance of $R = 10, 100$ km.

large magnitude event at near and far distance clearly shows the amplification effects on the spectral acceleration values for class C and D sites.

Furthermore, the importance of developing region-specific ground motion relations is highlighted by comparing the developed GMPEs with other ground motion relations developed based on NGA West-2 database for active shallow crustal regions. The spectral acceleration values, for time periods ranging from 0.01 to 4 s, are predicted for two different magnitude values and hypocentral distances and are shown in Fig. 11. It can be observed from Fig. 11 that the NGA West-2 GMPEs compares well in the time period ranges above 0.2 s but underestimates the ground motion for periods lower than 0.2 s. However, Fig. 11 shows that the ground motion observed in this region exhibits high frequency content for time period less than 0.2 s which are predicted reasonably well by the GMPEs presented in this study. Hence, this highlights the need for the regional specific GMPEs such as the ones developed in this study to account for the regional geology, source parameters, and local site effects.

5. SUMMARY AND CONCLUSIONS

Despite being a seismically active region, the state of Himachal Pradesh (HP) lacks a robust seismic hazard map based on regional ground motion relation except for the NDMA (2011) relations developed for class A sites. The investigation of the site characteristics, by MASW tests, at 73 locations in
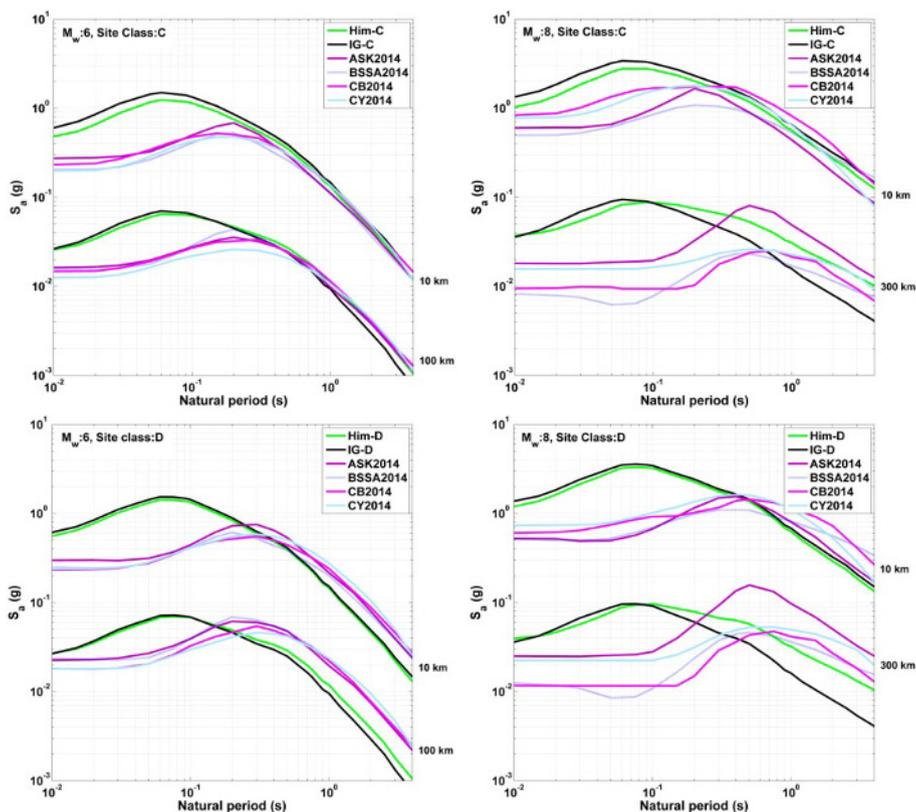
Fig. 11. Comparison of the predicted spectral values of the developed GMPEs with some of the GMPE based on NGA West 2 Data. The spectra are shown for the combination of $M_w = 6$ and $R = 10, 100$ km; $M_w = 8$ and $R = 10, 300$ km; for class C and D sites. The Vs_{30} values for the "NGA West 2" GMPEs were taken as 560 and 270 m/s for class C and D, respectively. Note: ASK2014 – Abrahamson *et al.* (2014), BSSA2014 – Boore *et al.* (2014), CB2014 – Campbell and Bozorgnia (2014), CY2014 – Chiou and Young (2014).

22 important cities situated across the state of HP indicated that the test sites are mainly of class C and D type. This highlights the need to develop regional ground motion relations considering the regional geology and the local site effects.

The development of such ground motion relations require detailed earthquake catalogue for this region. But such a catalogue is not available due to the scarcity of the existing ground motion records. Instead, a simulated database of bedrock ground motions has been generated using the widely adopted stochastic seismological model considering the characteristic features of the source and medium parameters of the two geological regions. The soil

profiles developed in this study are then used to scale the ground motion simulated for class A sites into the widely prevalent class C and D site responses using an equivalent linear site response analysis. As a result, four synthetic databases of surface ground motions has been generated for both the observed site classes and the two geological regions considering the uncertainties associated with the regional and source parameters.

Using these surface ground motion databases, new ground motion relations for class C and D sites have been developed for the two geological regions. The predictions of these ground motion relations are found to be in reasonable agreement with the observed data. Also, the comparison of the predictions with the existing ground motion relations for class A site highlights the need for site-specific ground motion relations. In addition to that, the comparison with the NGA West-2 ground motion relations highlights the importance of developing regional ground motion relations.

In the present scenario with the limited amount of ground motion data available for this seismically active region, the developed ground motion relations are not only both regional and site-specific but also complements the class A ground motion relations of NDMA (2011). With these ground motion relations, the probable site-specific seismic hazard for this region can be evaluated for practical engineering purposes. Therefore, a detailed probabilistic seismic hazard analysis of this region has been carried out and the results are presented in the companion paper for practical engineering purposes.

R e f e r e n c e s

Abrahamson, N.A., W.J. Silva, and R. Kamai (2014), Summary of the ASK14 ground motion relation for active crustal regions, *Earthq. Spectra* **30**, 3, 1025-1055, DOI: 10.1193/070913EQS198M.

Atkinson, G.M., and D.M. Boore (1995), Ground-motion relations for eastern North America, *Bull. Seismol. Soc. Am*. **85**, 1, 17-30.

Boore, D.M. (1983), Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra, *Bull. Seismol. Soc. Am.* **73**, 6A, 1865-1894.

Boore, D.M. (2009), Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM, *Bull. Seismol. Soc. Am*. **99**, 6, 3202-3216, DOI: 10.1785/0120090056.

Boore, D.M., and G.M. Atkinson (1987), Stochastic prediction of ground motion and spectral response parameters at hard-rock sites in eastern North America, *Bull. Seismol. Soc. Am*. **77**, 2, 440-467.

Boore, D.M., and J. Boatwright (1984), Average body-wave radiation coefficients, *Bull. Seismol. Soc. Am*. **74**, 5, 1615-1621.

Boore, D.M., and W.B. Joyner (1997), Site amplifications for generic rock sites, *Bull. Seismol. Soc. Am*. **87**, 2, 327-341.

Boore, D.M., J.P. Stewart, E. Seyhan, and G.M. Atkinson (2014), NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthq. Spectra* **30**, 3, 1057-1085, DOI: 10.1193/070113EQS184M.

Brune, J.N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res*. **75**, 26, 4997-5009, DOI: 10.1029/JB075i026p04997.

Campbell, K.W., and Y. Bozorgnia (2014), NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra, *Earthq. Spectra* **30**, 3, 1087-1115, DOI: 10.1193/062913EQS175M.

Chiou, B.S.-J., and R.R. Youngs (2014), Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **30**, 3, 1117-1153, DOI: 10.1193/072813EQS219M.

Hwang, H., and J.-R. Huo (1997), Attenuation relations of ground motion for rock and soil sites in eastern United States, *Soil Dyn. Earthq. Eng*. **16**, 6, 363-372, DOI: 10.1016/S0267-7261(97)00016-X.

IBC (2009), *2009 International Building Code*, International Code Council, Inc., Country Club Hills, USA.

Idriss, I.M., and J.I. Sun (1992), SHAKE91: A computer program for conducting equivalent linear seismic response analyses of horizontally layered soil deposits. User's guide, Center for Geotechnical Modeling, Department of Civil and Environmental Engineering, Univ. of California, Davis, USA.

Iman, R.L., and W.J. Conover (1980), Small sample sensitivity analysis techniques for computer models, with an application to risk assessment, *Commun. Stat. Theor. Meth*. **9**, 17, 1749-1842, DOI: 10.1080/03610928008827996.

IS:1893 (2002), Indian standard, Criteria for earthquake resistant design of structures: Part 1, 5th rev., Bureau of Indian Standards, New Delhi, India.

Joyner, W.B., and D.M. Boore (1981), Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake, *Bull. Seismol. Soc. Am*. **71**, 6, 2011-2038.

Kayal, J.R. (2008), *Microearthquake Seismology and Seismotectonics of South Asia*, Springer, Dordrecht, Capital Publ. Co., New Delhi.

Mittal, H., A. Kumar, and R. Ramhmachhuani (2012), Indian national strong motion instrumentation network and site characterization of its stations, *Int. J. Geosci*. **3**, 6A, 1151-1167, DOI: 10.4236/ijg.2012.326117.

Mohanty, W.K., R. Prakash, G. Suresh, A.K. Shukla, M.Y. Walling, and J.P. Srivastava (2009), Estimation of coda wave attenuation for the national capital region, Delhi, India using local earthquakes, *Pure Appl. Geophys*. **166**, 3, 429-449, DOI: 10.1007/s00024-009-0448-7.

Motazedian, D., and G.M. Atkinson (2005), Stochastic finite-fault modeling based on a dynamic corner frequency, *Bull. Seismol. Soc. Am*. **95**, 3, 995-1010, DOI: 10.1785/0120030207.

Muthuganeisan, P., and S.T.G. Raghukanth (2016), Site-specific probabilistic seismic hazard map of Himachal Pradesh, India. Part II. Hazard estimation, *Acta Geophys*., DOI: 10.1515/acgeo-2016-0011.

NDMA (2011), Development of probabilistic seismic hazard map of India, Technical Report, Working Committee of Experts (WCE), National Disaster Management Authority (NDMA), New Delhi, India.

Park, C.B., R.D. Miller, and J. Xia (1999), Multichannel analysis of surface waves, *Geophysics* **64**, 3, 800-808, DOI: 10.1190/1.1444590.

Saragoni, G.R., and G.C. Hart (1973), Simulation of artificial earthquakes, *Earthq. Eng. Struct. Dyn*. **2**, 3, 249-267, DOI: 10.1002/eqe.4290020305.

Singh, S.K., R.S. Dattatrayam, N.M. Shapiro, P. Mandal, J.F. Pacheco, and R.K. Midha (1999), Crustal and upper mantle structure of Peninsular India and source parameters of the 21 May 1997, Jabalpur earthquake (Mw = 5.8): Results from a new regional broadband network, *Bull. Seismol. Soc. Am*. **89**, 6, 1631-1641.

Singh, S.K., D. García, J.F. Pacheco, R. Valenzuela, B.K. Bansal, and R.S. Dattatrayam (2004), Q of the Indian shield, *Bull. Seismol. Soc. Am*. **94**, 4, 1564-1570, DOI: 10.1785/012003214.

Toro, G.R., N.A. Abrahamson, and J.F. Schneider (1997), Models of strong ground motions from earthquakes in central and eastern North America: best estimates and uncertainties, *Seismol. Res. Lett*. **68**, 1, 41-57, DOI: 10.1785/gssrl.68.1.41.

Vucetic, M., and R. Dobry (1991), Effect of soil plasticity on cyclic response, *J. Geotech. Eng. ASCE* **117**, 1, 89-107, DOI: 10.1061/(ASCE)0733-9410 (1991)117:1(89).

Assessment
of the Relative Largest Earthquake Hazard Level
in the NW Himalaya and its Adjacent Region

Theodoros M. TSAPANOS[1], R.B.S. YADAV[2], Efthalia M. OLASOGLOU[1],
and Mayshree SINGH[3]

[1]Aristotle University of Thessaloniki, School of Geology, Geophysical Lab,
Thessaloniki, Hellas; e-mails: tsapanos@geo.auth.gr (corresponding author),
lia_olasoglou@hotmail.com
[2]Department of Geophysics, Kurukshetra University, Kurukshetra, India;
e-mail: rbsykuk@gmail.com
[3]School of Engineering, University of KwaZulu-Natal, Durban, South Africa;
e-mail: singhm5@ukzn.ac.za

A b s t r a c t

In the present study, the level of the largest earthquake hazard is assessed in 28 seismic zones of the NW Himalaya and its vicinity, which is a highly seismically active region of the world. Gumbel's third asymptotic distribution (hereafter as GIII) is adopted for the evaluation of the largest earthquake magnitudes in these seismic zones. Instead of taking in account any type of M_{max}, in the present study we consider the ω value which is the largest earthquake magnitude that a region can experience according to the GIII statistics. A function of the form $\Theta(\omega, \mathrm{RP}_{6.0})$ is providing in this way a relatively largest earthquake hazard scale defined by the letter K (K index). The return periods for the ω values (earthquake magnitudes) 6 or larger ($\mathrm{RP}_{6.0}$) are also calculated. According to this index, the investigated seismic zones are classified into five groups and it is shown that seismic zones 3 (Quetta of Pakistan), 11 (Hindukush), 15 (northern Pamirs), and 23 (Kangra, Himachal Pradesh of India) correspond to a "very high" K index which is 6.

Key words: K index, GIII, ω value, return period, NW Himalaya.

1. INTRODUCTION

Earthquake hazard assessment of a region is the ultimate goal for geoscientists, which is a useful tool for the preparation of earthquake risk mitigation policies. Earthquake hazard can be defined as the probability of occurrence of future earthquakes of a specific size during a given time interval in a seismic region of the globe. Numerous qualitative as well as quantitative methods have been formulated and applied by different researchers over the years to assess the earthquake hazard in several seismic regions of the world.

The first (GI) and the third (GIII) asymptotic distributions of extreme values (Gumbel 1958) have proven a useful tool in estimating earthquake hazard in different seismic regions, globally distributed. Both Gumbel's distributions have the advantage that they do not require analysis of the whole data set used. These procedures require predetermined fix time intervals from which the largest earthquake magnitudes are selected. These arbitrary time intervals are usually based on the seismicity rate of the region under investigation. The selected time intervals are, in some cases, a good tool to exclude automatically the fore- and after-shocks which is a necessary precondition in such kind of analysis (Yegulalp and Kuo 1974, Schenkova and Karnik 1976). On the other hand, problems concerning the earthquake hazard parameters obtained through this methodology have been discussed by Knopoff and Kagan (1977). However, Båth (1973, 1975, 1983) suggested that the dependency of the method on the largest magnitudes is the principal advantage of the technique, given that the magnitudes of the largest events are more accurately determined than those of small shocks.

Particularly, the GIII distribution has another advantage of the inclusion of an upper-bound earthquake magnitude called "ω value". This upper-bound magnitude is consistent with the concept of the finite maximum stresses, which may have been accumulated in rocks during a time interval of successive earthquakes and then subsequently released when seismic events occurred. This parameter is a unique characteristic for a region and no maximum earthquake magnitudes (*e.g.*, maximum observed) in a region can exceed this upper bound magnitude. Thus, for the computation of occurrence or expectation of earthquakes with extreme magnitudes using probabilistic models, the GIII asymptotic distribution allows a proper and natural physical interpretation. The plots of frequency-magnitude distribution usually reveal a curvature shape. This curvature has been observed especially when the distribution moves near the maximum earthquakes (Page 1968, Utsu 1971, Burton 1977, Bloom and Erdmann 1980). The parameters of the GIII distribution allow for any detectable curvature to the upper bound magnitude. Both techniques of Gumbel's first and third distribution have been applied for earthquake hazard assessment by a number of researchers in different re-

gions of the world (*e.g.*, Yegulalp and Kuo (1974) for Pacific Ocean; Makropoulos (1978) for Greece; Burton (1979) for Europe to India; Makropoulos and Burton (1985) for Pacific rim; Tsapanos and Burton (1991) for the whole world; Tsapanos (1997) for circum Pacific belt; Shanker *et al*. (2007) for Hindukush-Pamir Himalaya region; Bayrak *et al*. (2008) for Turkey; Yadav *et al*. (2012a, 2013a) for NW Himalaya; Tsapanos *et al*. (2014) for Turkey; among others).

The present paper confines itself to the evaluation of the "largest earthquake hazard level" (*K* index) for 28 predefined seismic zones in the NW Himalaya and the adjoining areas. The ω value is considered as the maximum (largest) earthquake magnitude for each one of these seismic zones and the return periods ($RP_{6.0}$) for ω values ≥ 6.0 are also assessed. The classification of the seismic zones in different groups, in terms of relative earthquake hazard level, provides an image of quantitative largest earthquake hazard in the NW Himalaya and adjacent regions.

2. SEISMOTECTONIC SETUP AND PAST SEISMIC HAZARD IN THE REGION

The study region is one of the active seismic regions in the Indian subcontinent which is bounded by latitude 25°-40°N and longitude 65°-85°E (Fig. 1). The investigated region is located at the western syntaxis of the Himalayan part of Alpide belt and its neighboring regions, which includes India, Pakistan, Afghanistan, Hindukush, Pamirs, Mangolia, and Tien-Shan. This region is one of the most seismically active continent-continent (Indian-Eurasian plates) collision type active plate margin regions of the world. Structurally, this region is controlled by large rigid lithospheric blocks of the Indian plate from the south, the Afghan block in the southwest, the Turan plate in the west, and the Tarim block in the northeast (Koulakov and Sobolev 2006, Yadav 2009, Yadav *et al*. 2012b). The region exhibits intensive folding and thrusting, which occurred in the Cenozoic and Mesozoic era (Gansser 1964). The trend of folding and faulting in the Himalayan region is of NW-SE to EW direction. The Main Himalayan Frontal Thrust (HFT) and the Hazara, Sulaiman, and Kirthar ranges highlight the collisional deformations of the Indian and Eurasian plates (Seeber and Armbruster 1981, Thingbaijam *et al*. 2009, Yadav *et al*. 2012b).

The seismic hazard level of the region under investigation is high in terms of occurrences of larger size earthquakes during previous centuries and exceedance of peak ground acceleration (PGA) estimated through probabilistic seismic hazard assessment. Two great earthquakes (1902, Caucasus (M_S 8.6) and 1905, Kangra, India (M_S 8.6)) and several moderate to large earthquakes (Figs. 1 and 2) have occurred in the region during previous century (Gutenberg and Richter 1954, Yadav *et al*. 2012b). Several researchers
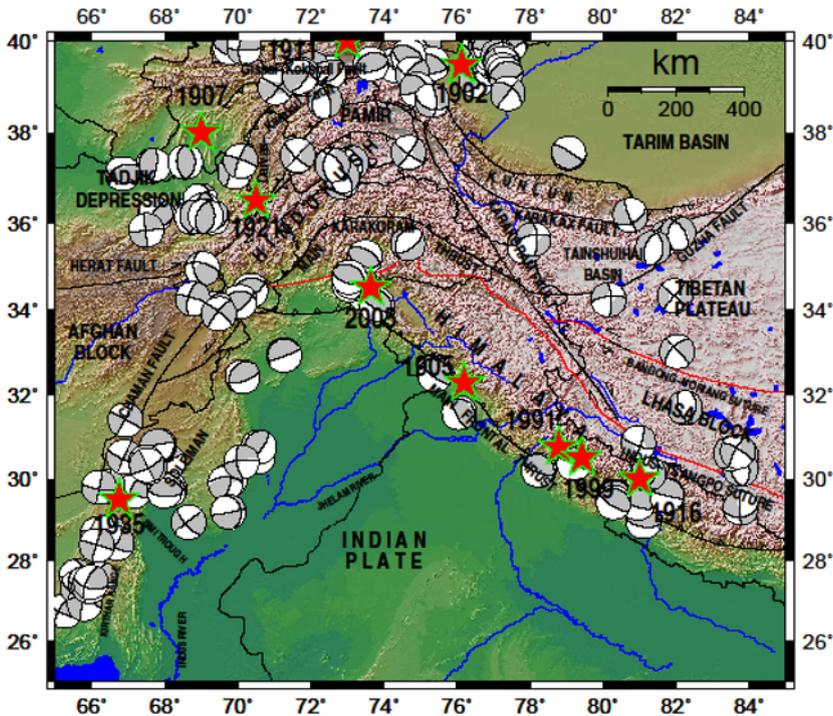
Fig. 1. Tectonic map of the NW Himalaya and adjoining regions, showing the fault and fold systems (after Koulakov and Sobolev 2006, Yadav 2009, Yadav *et al*. 2010). Focal mechanism solutions of shallow earthquakes with $M_W \geq 5.5$ obtained from Harvard GCMT catalogue during the period 1976-2010 are also shown with beach balls in the map, revealing the style of faulting in different parts of the region. Some large and damaging earthquakes are shown with stars along with year of occurrence.

(*e.g.*, Chatelain *et al*. 1980, Burtman and Molnar 1993, Fan *et al*. 1994, Lyubushin *et al*. 2010, Arora *et al*. 2012) suggested that two converging seismic regimes – northward subduction of the Indian plate beneath the Hindukush and southward subduction of the Eurasian plate below the Pamir – control the seismicity of the Hindukush–Pamir thrust zone. The western part of the Himalayan arc, consisting of the Kashmir ranges, turns to the south near Nanga Parbat to form the Hazara syntaxis (Meltzer *et al*. 2001), where the 2005 Kashmir earthquake (M_W 7.6) was nucleated. The Chaman fault zone has been associated with two major earthquakes: 1931, Mach (M_S 7.4), and 1945, Quetta (M_W 7.7). The Arabian plate is apparently being subducted northward, forming the subduction zone overlying the E-W trending Makran ranges (Quittmeyer and Jacob 1979). The seismogenic surge of the 1945
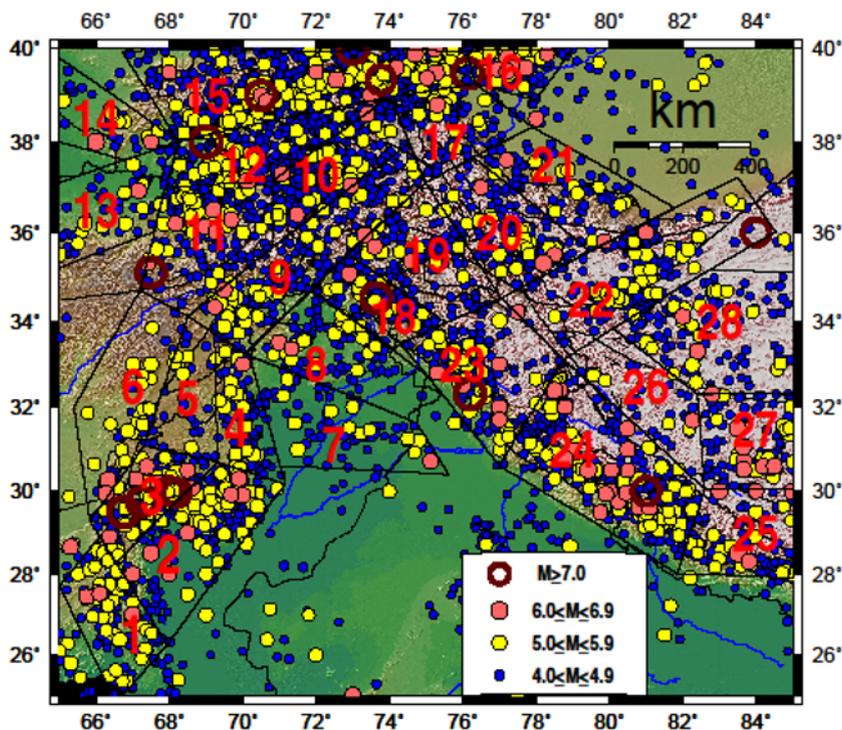
Fig. 2. Characterization of 28 possible seismic zones in the NW Himalaya and adjoining regions on the basis of seismicity, tectonics, and focal mechanism of earthquakes (after Yadav *et al.* 2012b). The epicentral distribution of independent earthquakes (main shocks) of $M_W \geq 4.0$ that occurred during the period 1900-2010 are also shown in the figure along with source zones that reveal seismic activity of each zone.

Makran (M_W 8.1) earthquake triggered a tsunami in Pakistan and Indian regions. The Himalayan Frontal Thrust belt of the Indian region has experienced three moderate to large earthquakes of M_W 6.8 and 6.5 in 1991, and 1999 at Uttarkashi and Chamoli region, respectively, and most devastating earthquake of the Kashmir Himalaya in 2005 of M_W 7.6. The study region is located in all seismic zones, V, IV, III and II on the seismic zoning map of India (Bureau of Indian Standards (BIS 2002)), with an assigned Peak Ground Acceleration (PGA) values of 0.4, 0.3, 0.2, and 0.1 g, respectively (Nath and Thingbaijam 2012).

3.   EARTHQUAKE CATALOGUE USED

The earthquake hazard assessment of a region depends upon the quality of the prepared earthquake catalogue, *i.e.*, compilation, homogenization of

magnitude scale, removal of dependent events (foreshocks and aftershocks), and finally completeness analysis with respect to magnitude and time. In the present study, the earthquake catalogue is adopted from Yadav *et al.* (2012b) who compiled a homogeneous and complete catalogue for the region under investigation using various historical and instrumental earthquakes catalogues. The initially compiled catalogue from different sources contains several earthquake magnitude scales, which was homogenized for moment magnitude (M_W) using different empirical relations established among body-wave magnitude (m_b), surface-wave magnitude (M_S), local magnitude (M_L), and moment magnitude (M_W). As we aforementioned, foreshocks and aftershocks are removed using a spatial and temporal windowing method developed by Uhrhammer (1986). In this technique, a scan for the entire catalogue has been performed within a defined spatial and temporal window for each given earthquake. All events (foreshocks and aftershocks) with epicenters falling within the defined two windows are removed after considering them as dependent events. The completeness of earthquakes catalogue is performed with respect to magnitude and time. The magnitude of completeness (M_C), also called threshold or cut-off magnitude, is defined as the lowest magnitude at which 100% of events in space-time volume are detected. The method of Entire Magnitude Range (EMR) (Woessner and Wiemer 2005) is used to estimate the M_C. The completeness with respect to time is also calculated using a simple graphical technique known as "visual cumulative method" given by Mulargia and Tinti (1985). A graph is constructed between the time and cumulative number of events for a particular magnitude range. Then, the completeness interval will be the number of years from the beginning of the period to the last year of occurrence in the catalogue. A full review of constructions and content of this catalogue can be obtained from Yadav *et al.* (2012b).

In the study of Yadav *et al.* (2012b) only the instrumental part of this catalogue during the period 1900-2010 with $M_W \geq 4.0$ was taken for the computation of earthquake hazard and the epicentral distribution of earthquakes is shown in Fig. 2. It is observed that most of the earthquakes of this catalogue are concentrated along the Hindukush–Pamir Himalaya region, the Himalayan frontal thrust belt and the Sulaiman–Kirthar ranges of Pakistan. The five largest earthquakes in this catalogue are: 1907 Hindukush earthquake of M_W 7.9; 1902 Caucasus earthquake of M_W 7.8 (M_S 8.6); 1905 Kangra, India earthquake of M_W 7.8 (M_S 8.6); 1935 Quetta, Pakistan earthquake of M_W 7.7; and 1911 Sarez, Pamir earthquake of M_W 7.7. The catalogue consists of 32 earthquakes having magnitude greater than M_W 7.0 and 73 earthquakes of $M_W \geq 6.5$, which suggest that the examined region is highly active.

4. THE METHODOLOGY APPLIED

The Gumbel's methodology, applied in the present study, is not a new one. But, the use of ω values of Gumbel's third asymptotic distribution (GIII) as the maximum magnitude is considered as a new approach for the scope of the present study. Since the GIII method is widely known to the seismological community, we shall give a very brief description below.

If M is the largest earthquake magnitude observed in successive equal time periods of a given area, then the probability that M is an extreme value magnitude is given by the following probability distribution function:

$$P(M) = \exp\left[-\left(\frac{\omega - M}{\omega - u}\right)^k\right] , \tag{1}$$

where ω is the upper bound to M, u is the characteristic value, and k is the shape parameter, with $P(u) = 1/e$ and $P(\omega) = 1$. If the extreme equal time span is one year, the return period RP (in years) for an earthquake magnitude M is given as:

$$T(M) = \frac{1}{\left[1 - P(M)\right]} , \tag{2}$$

where $[1 - P(M)]$ is the annual probability that an earthquake magnitude will be exceeded. If we have extreme intervals of N-year duration, then the corresponding distribution of $P_N(M)$ can be related with the one-year extreme $P_1(M)$ by the following relation:

$$P_1(M) = \sqrt[N]{P_N(M)} . \tag{3}$$

The GIII distribution gives an appropriate and natural physical interpretation with the inclusion of an upper bound magnitude (ω) that allows for the calculation of the occurrence of extreme magnitude earthquakes using a probabilistic model. The frequency-magnitude distribution often shows curvature, especially when the largest magnitudes are approached (Page 1968, Utsu 1971, Bloom and Erdmann 1980). The parameters of the GIII distribution allow both for any detectable curvature in addition to the upper bound magnitude. For curve fitting, Eq. 1 is transposed into the following form:

$$M = \omega - (\omega - u)\left[-\ln(P(M))\right]^\lambda , \tag{4}$$

where $\lambda = 1/k$ and plotting M as ordinate and $[-\ln(P(M))]^\lambda$ as abscissa draws a straight line with ω as intercept and $-\omega(\omega - u)$ as gradient.

Earlier, Yadav *et al.* (2012a) computed the ω values in a regional scale for the 28 seismic zones in the Hindukush–Pamir Himalaya and surrounding area. For this purpose they considered data for the period 1900-2010. In the present study, an effort is made to evaluate new ω values with additional data (2011-2013), keeping the lower cut of magnitude with $M_W \geq 4.0$, but the

obtained results are almost the same as those obtained by Yadav *et al.* (2012a). The average difference between the ω values estimated by Yadav *et al.* (2012a) and the ω values obtained by this new attempt is approximately –0.06, which is negligible. Therefore, we decided to use as the maximum magnitude (ω values) the results obtained in the present study for the estimation of the largest earthquake hazard level. A description of these values is listed in Table 1. In this table, it is obvious that ω values vary from 6.14 to

Table 1

The results of parameter ω (present study and Yadav *et al.* 2012a), M_{max}^{obs}, the return periods for $RP_{5.0}$ and $RP_{6.0}$ and the relative largest earthquake hazard level (K index) for 26 out of the 28 examined seismic zones in the NW Himalaya and adjacent regions

Zones	ω new (present study)	ω old (Yadav *et al.* 2012a)	Difference (ω new – ω old)	M_{max}^{obs}	$RP_{5.0}$ (in yrs)	$RP_{6.0}$ (in yrs)	K index
1	6.41	6.51	–0.10	6.3	3	62	2
2	6.99	6.90	0.09	6.7	3	24	3
3	7.73	8.44	–0.71	7.7	5	11	6
4	6.16	6.10	0.06	6.0	6	170	2
5	–	–	–	5.8	–	–	–
6	6.98	6.76	0.22	6.7	6	33	3
7	6.25	6.29	–0.04	6.2	8	387	2
8	6.46	6.58	–0.12	6.3	8	115	2
9	7.88	7.21	0.67	7.0	4	22	4
10	7.70	7.75	–0.05	7.4	2	4	5
11	8.04	7.67	0.37	7.5	2	3	6
12	8.1	8.26	–0.16	7.9	4	27	5
13	6.14	6.26	–0.12	6.1	8	302	2
14	7.22	7.00	0.22	6.9	12	61	3
15	7.92	8.15	–0.23	7.7	2	6	6
16	7.82	7.88	–0.06	7.6	4	18	5
17	6.63	6.54	0.09	6.3	3	245	2
18	7.63	8.58	–0.95	7.6	4	28	5
19	6.3	6.59	–0.29	6.3	5	89	2
20	7.05	7.13	–0.08	7.0	3	17	5
21	–	–	–	6.0	–	–	–
22	7.29	7.49	–0.20	7.1	2	14	5
23	8.12	8.03	0.09	7.8	4	18	6
24	7.71	7.86	–0.15	7.2	2	8	5
25	7.12	7.09	0.03	6.9	5	36	4
26	6.72	6.36	0.36	6.3	8	219	2
27	6.79	7.08	–0.29	6.7	4	18	5
28	6.47	6.62	–0.15	6.4	4	42	3

8.12, while the maximum observed magnitudes, M_{max}^{obs}, are between 5.8 and 7.9. The largest observed earthquake ($M_{max}^{obs} = 7.9$) occurred in zone 12 during the year 1907.

5.   RESULTS AND DISCUSSION

In the present study, the authors emphasize their efforts to estimate the relative largest earthquake hazard level (K index) in 28 seismic zones of the NW Himalaya and adjoining regions using a homogenous and complete earthquake catalogue covering the time period 1900-2013. A similar approach has been introduced by Papadopoulos and Kijko (1991) for the seismic zones in the Greece and Tsapanos (2001) for the seismic regions around the circum Pacific belt. Both of the above approaches considered the maximum magnitude as the maximum regional magnitude, M_{max}^{reg}, proposed by Kijko and Sellevoll (1989, 1992). In both of the after mentioned studies (Papadopoulos and Kijko 1991, Tsapanos 2001) the relative seismic hazard level was defined by the letter K. We keep this symbol to our study as well.

As mentioned above, the parameter ω derived from the GIII method is used in the present work for the estimation of largest earthquake hazard level in the considered region. A first inspection of Table 1 shows that, generally, the estimated ω values and M_{max}^{obs} values do not differ significantly. This is due to the fact that the estimation of the ω parameter in most of the analyzed seismic zones is based on relatively long duration of earthquake data, which is comparable with the seismic cycle of the strongest earthquake in respective zones. The largest ω values are evaluated for the seismic zones 11, 12, and 23 with corresponding results of 8.04, 8.10, and 8.12, respectively. These seismic zones are characterized by intense seismicity during the investigated time period. We have also listed the return periods of earthquakes with $M = 5.0$ ($RP_{5.0}$) and $M = 6.0$ ($RP_{6.0}$) in Table 1. It is observed that the return periods of earthquake $M = 6.0$ make sense since this magnitude is considered to be a hazardous shock and almost all of the 28 seismic zones have experienced such earthquakes (exception observed only in zone 5 with $M_{max}^{obs} = 5.8$). Therefore, the return periods $RP_{6.0}$ are reliable for the further assessment of the relative earthquake hazard level. Only two of the studied seismic zones (5 and 21) show abnormal $RP_{6.0}$. We interpret this non-normality on the basis of the fact that the distribution of these zones shows a linear tendency rather than a curvature one (Fig. 3). These figures demonstrate both linear and curvature distribution (as provided by GIII) in zone 5 and 21, respectively. The χ^2 test is applied for both zones in order to examine statistically which is the dominant distribution. On the basis of χ^2 test, we
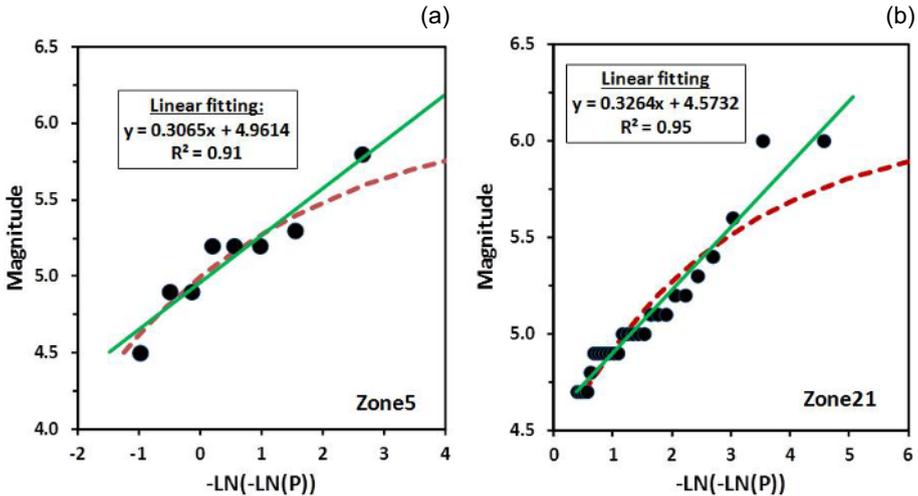
Fig. 3. Gumbel-III distribution (dash red line) and linear least square fitting (solid green line) with observed data (solid circles) for zone 5 (a) and zone 21 (b) showing that observed data set is fitted better by linear line than GIII asymptotic curve in these zones (modified after Yadav *et al*. 2012a).

observed very surprising results which show that both distributions can be accepted. For zone 5, χ^2 test for linear distribution is obtained as 0.999, while for the curvature (GIII) it is equal to 0.993. Similarly, for zone 21, χ^2 test for linear distribution is observed as 1.000 and it is 0.999 for curvature distribution. The estimated regression coefficient (R^2) for the best fit of the linear distribution is obtained as 0.91 for zone 5, while 0.95 is observed for zone 21, showing percentage linear fitting of observed data. Another evidence for linear distribution comes out from a first look on Fig. 3. The inspection of these figures shows that the curvature distribution does not fit well to the upper bound magnitude, as it was expected from the Gumbel's theory, but relents lower than the maximum observed magnitude in both zones ($M_{\max}^{\text{obs}} = 5.8$ for zone 5 and $M_{\max}^{\text{obs}} = 6.0$ for zone 21). All of the above observations make us able to conclude that the linear distribution is more proper for zones 5 and 21 than the curvature. Therefore, we excluded these two seismic zones (5 and 21) from our study.

In order to assess the relative earthquake hazard level in the considered region, we classified the 26 seismic zones under investigation in groups based on their difference in largest earthquake hazard level (K index). For this purpose, we follow a similar approach considered by Papadopoulos and Kijko (1991) and Tsapanos (2001). Therefore, we equally took into account the importance of ω values and RP$_{6.0}$. We considered that the K index is a
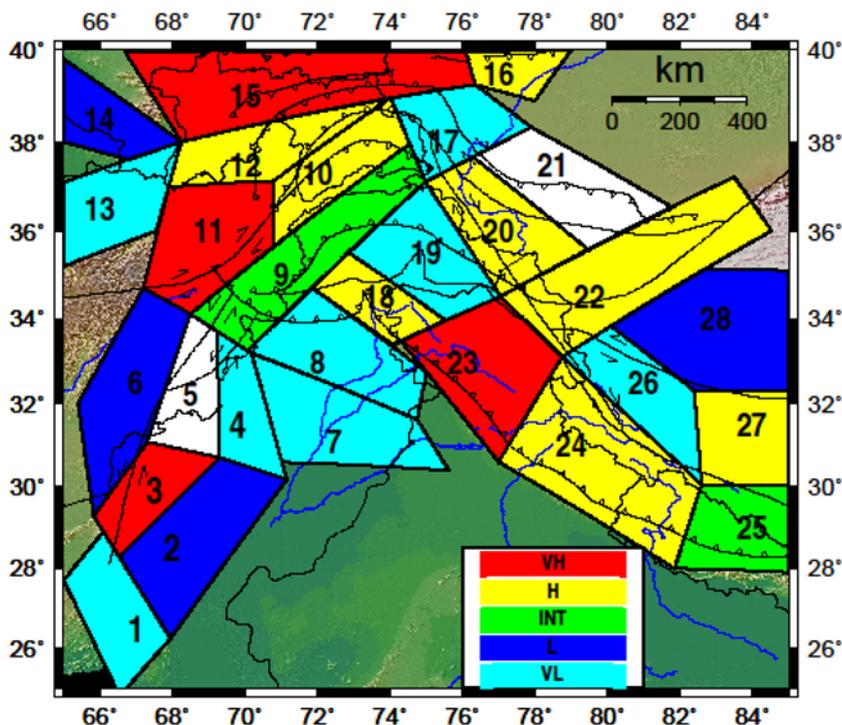
Fig. 4. Spatial distribution of relative largest earthquake hazard level (*K* index) in 26 out of the 28 seismic zones of the examined region depicting the hazard level in different zones estimated by largest earthquake magnitude expressed by ω values and return periods of *M*6.0 earthquake.

function of ω values and $RP_{6.0}$ and denoted as the term $\Theta(\omega, RP_{6.0})$. This parameter (Θ) increases with ω values and decreases with the values of return periods $RP_{6.0}$. Then, we constructed the following groups: $\omega \le 6.9$, $7.0 \le \omega \le 7.9$, and $\omega \ge 8.0$, and defined them as 2, 4, and 6, respectively. We follow the same procedure for $RP_{6.0}$ (in years) and constructed three groups of this parameter as $RP_{6.0} \le 20$, $21 \le RP_{6.0} \le 50$, and $RP_{6.0} \ge 51$ for the corresponding numbers of 6, 4 and 2, respectively. The arithmetic mean of these two parameters, *i.e.*, $K = 1/2 [\Theta(\omega) + \Theta(RP_{6.0})]$ expresses the adopted *K* index of a seismic zone. The *K* index has values equal to 2, 3, 4, 5, and 6 and characterizes five groups of *K*, which are: very low = 2, low = 3, intermediate = 4, high = 5, and very high = 6. Table 1 (last column) demonstrates the *K* index for each of the studied zones. Figure 4 illustrates the spatial distribution of *K* index in the 26 seismic zones (5 and 21 are excluded) of the NW Himalaya and adjacent area, which were defined by Yadav *et al*. (2012b).

A detail inspection of Fig. 4 shows that four seismic zones (3 − Quetta of Pakistan region, 11 − Hindukush, 15 − northern Pamirs, and 23 − Kangra, Himachal Pradesh of India) are corresponding with the "very high" K index of 6. These seismic zones are characterized by intense seismic activity in the past and Yadav *et al.* (2012a, b) also evaluated the largest earthquake magnitudes, of more than 8.0, in these zones either by GIII or Kijko–Sellevoll methods. Next "high" K index of 5 is observed in seismic zones 10, 12, 16, 18, 20, 22, 24, and 27 in tectonic regimes of the Hindukush seismic Belt, Caucasus, Kunlun fault, Kashmir, Uttarkashi–Chamoli region, and part of Tibet. These regions are seismically active and have experienced several moderate to large earthquakes in their history. Only two seismic zones (9 − southern Hindukush belt, and 25 − Nepal Himalaya) show intermediate K index having a value of 4 which corresponds with moderate size earthquakes in the past. The low K index of value 3 is observed in four seismic zones: 2, 6, 14, and 28, while very low K index of value 2 is observed in seismic zones: 1, 4, 7, 8, 13, 17, 19, and 26. These regions are associated with the low seismic activity regions of the Pakistan, Afghanistan, Karakoram fault, and parts of the Tibetan Plateau. It is interesting to note that about one half of the studied seismic zones are associated with low and very low K index, while about the other half of the studied seismic zones are corresponding to high and very high K index. Yadav *et al.* (2010, 2012a, b; 2013a, b; 2015) also investigated earthquake hazard using different statistical techniques in the same seismic zones of the studied region and observed similar level of hazard in these seismic zones, revealing the seismic characteristics of the region. A good correspondence is also observed with spatial maps prepared by Khattri *et al.* (1984), Bhatia *et al.* (1999), Lyubushin and Parvez (2010), and Nath and Thingbaijam (2012) for peak ground accelerations (PGA) in the examined region.

6. CONCLUSIONS

The parameter ω deduced from GIII method in the present study is a characteristic parameter for any seismic zone. According to our knowledge, this quantity has never been evaluated before for such kind of study in the examined region. Hitherto, only the maximum regional magnitude (M_{max}) has been applied in such kind of studies (Papadopoulos and Kijko 1991, Tsapanos 2001).

For the purpose of this study, we used the relative largest earthquake hazard level (K index) which can be computed as a function of the form of $\Theta(\omega, RP_{6.0})$. The arithmetic mean of two parameters (ω and $RP_{6.0}$) is used as a criterion to classify the seismic zones of the NW Himalaya and surrounding regions defined by Yadav *et al.* (2012b) in five groups. The considered values of K index are 2, 3, 4, 5, and 6, by which the seismic zones are ranked

as "very high", "high", "intermediate", "low", and "very low" level of K index. From the spatial map (Fig. 4) of these indexes, it is revealed that four of the seismic zones (3 − Quetta of Pakistan region, 11 − Hindukush, 15 − northern Pamirs, and 23 − Kangra, Himachal Pradesh of India) have very high K index of 6, while eight of them show high K of 5. It is observed that seismic zone 12, in which the largest observed earthquake (M_{max}^{obs} = 7.9) occurred, shows high K index and not very high as it was expected. On the other hand, zone 23 shows very high K index while it has experienced a great earthquake of M_W = 7.8 (M_S 8.6) in 1905. The spatial distribution of K index among these seismic zones is very useful for scientific purposes along with the designation of priority seismic zones for earthquake resistant design.

R e f e r e n c e s

Arora, B.R., V.K. Gahalaut, and N. Kumar (2012), Structural control on along-strike variation in the seismicity of the northwest Himalaya, *J. Asian Earth Sci.* **57**, 15-24, DOI: 10.1016/j.jseaes.2012.06.001.

Båth, M. (1973), *Introduction to Seismology*, Springer, Basel, 395 pp.

Båth, M. (1975), Seismicity of Tanzania region, *Tectonophysics* **27**, 4, 353-379, DOI: 10.1016/0040-1951(75)90004-9.

Båth, M. (1983), Earthquake frequency and energy in Greece, *Tectonophysics* **95**, 3-4, 233-252, DOI: 10.1016/0040-1951(83)90070-7.

Bayrak, Y., S. Öztürk, H. Çinar, G.Ch. Koravos, and T.M. Tsapanos (2008), Regional variation of the ω-upper bound magnitude of GIII distribution in and around Turkey: tectonic implications for earthquake hazards, *Pure Appl. Geophys.* **165**, 7, 1367-1390, DOI: 10.1007/s00024-008-0359-z.

Bhatia, S.C., M.R. Kumar, and H.K. Gupta (1999), A probabilistic seismic hazard map of India and adjoining regions, *Ann. Geophys.* **42**, 6, 1153-1164, DOI: 10.4401/ag-3777.

BIS (2002), Indian standard criteria for earthquake resistant design of structures. Part 1 – General provisions and buildings, IS 1893-2002, Bureau of Indian Standards, New Delhi, India.

Bloom, E.D., and R.C. Erdmann (1980), The observation of a universal shape regularity in earthquake frequency-magnitude distributions, *Bull. Seismol. Soc. Am.* **70**, 1, 349-362.

Burtman, V.S., and P. Molnar (1993), Geological and geophysical evidence of deep subduction of continental crust beneath the Pamir, *Geol. Soc. Am., Sp. Paper* **281**, 1-76, DOI: 10.1130/SPE281-p1.

Burton, P.W. (1977), The application of extreme value statistics to seismic hazard assessment in the European area. **In:** *Proc. Symp. Analysis of Seismicity and on Seismic Risk, 17-22 October 1977, Liblice, Czech Republic*, Vol. 1722, Academia, Prague, 323-334.

Burton, P.W. (1979), Seismic risk in southern Europe through to India examined using Gumbel's third distribution of extreme values, *Geophys. J. Int.* **59**, 2, 249-280, DOI: 10.1111/j.1365-246X.1979.tb06766.x.

Chatelain, J.L., S.W. Roecker, D. Hatzfeld, and P. Molnar (1980), Microearthquake seismicity and fault plane solutions in the Hindu Kush Region and their tectonic implications, *J. Geophys. Res.* **85**, B3, 1365-1387, DOI: 10.1029/JB085iB03p01365.

Fan, G., J.F. Ni, and T.C. Wallace (1994), Active tectonics of the Pamirs and Karakorum, *J. Geophys. Res.* **99**, B4, 7131-7160, DOI: 10.1029/93JB02970.

Gansser, A. (1964), *Geology of the Himalayas*, Interscience Publ., London, 289 pp.

Gumbel, E.J. (1958), *Statistics of Extremes*, Columbia University Press, New York, 375 pp.

Gutenberg, B., and C.F. Richter (1954), *Seismicity of the Earth and Associated Phenomena*, Princeton University Press, Princeton, 310 pp.

Khattri, K.N., A.M. Rogers, D.M. Perkins, and S.T. Algermissen (1984), A seismic hazard map of India and adjacent areas, *Tectonophysics* **108**, 1-2, 93-134, DOI: 10.1016/0040-1951(84)90156-2.

Kijko, A., and M.A. Sellevoll (1989), Estimation of earthquake hazard parameters from incomplete data files. Part I: Utilization of extreme and complete catalogs with different threshold magnitudes, *Bull. Seismol. Soc. Am.* **79**, 3, 645-654.

Kijko, A., and M.A. Sellevoll (1992), Estimation of earthquake hazard parameters from incomplete data files. Part II: Incorporation of magnitude heterogeneity, *Bull. Seismol. Soc. Am.* **82**, 1, 120-134.

Knopoff, L., and Y. Kagan (1977), Analysis of the theory of extremes as applied to earthquake problems, *J. Geophys. Res.* **82**, 36, 5647-5657, DOI: 10.1029/JB082i036p05647.

Koulakov, I., and S.V. Sobolev (2006), A tomographic image of Indian lithosphere break-off beneath the Pamir–Hindukush region, *Geophys. J. Int.* **164**, 2, 425-440, DOI: 10.1111/j.1365-246X.2005.02841.x.

Lyubushin, A.A., and I.A. Parvez (2010), Map of seismic hazard of India using Bayesian approach, *Nat. Hazards* **55**, 2, 543-556, DOI: 10.1007/s11069-010-9546-1.

Lyubushin, A.A., B.R. Arora, and N. Kumar (2010), Investigation of seismicity in western Himalaya, *Russ. J. Geophys. Res.* **11**, 1, 27-34.

Makropoulos, K.C. (1978), The statistics of large earthquake magnitude and an evaluation of Greek seismicity, Ph.D. Thesis, University of Edinburgh, Edinburgh, UK, 193 pp.

Makropoulos, K.C., and P.W. Burton (1985), Seismic hazard in Greece: I. Magnitude recurrence, *Tectonophysics* **117**, 3-4, 205-257, DOI: 10.1016/0040-1951(85)90273-2.

Meltzer, A., G. Sarker, B. Beaudoin, L. Seeber, and J. Armbruster (2001), Seismic characterization of an active metamorphic massif, Nanga Parbat, Pakistan Himalaya, *Geology* **29**, 7, 651-654, DOI: 10.1130/0091-7613(2001)029 <0651:SCOAAM>2.0.CO;2.

Mulargia, F., and S. Tinti (1985), Seismic sample areas defined from incomplete catalogues: An application to the Italian territory, *Phys. Earth Planet. In.* **40**, 4, 273-300, DOI: 10.1016/0031-9201(85)90038-X.

Nath, S.K., and K.K.S. Thingbaijam (2012), Probabilistic seismic hazard assessment of India*, Seismol. Res. Lett.* **83**, 1, 135-149, DOI: 10.1785/gssrl.83.1.135.

Page, R. (1968), Aftershocks and microaftershocks of the great Alaska earthquake of 1964, *Bull. Seismol. Soc. Am.* **58**, 3, 1131-1168.

Papadopoulos, G.A., and A. Kijko (1991), Maximum likelihood estimation of earthquake hazard parameters in the Aegean area from mixed data, *Tectonophysics* **185**, 3-4, 277-294, DOI: 10.1016/0040-1951(91)90449-3.

Quittmeyer, R.C., and K.H. Jacob (1979), Historical and modern seismicity of Pakistan, Afghanistan, northwestern India, and southeastern Iran*, Bull. Seismol. Soc. Am.* **69**, 3, 773-823.

Schenkova, Z., and V. Karnik (1976), Application of the largest values theory to Balkan earthquakes. **In:** *Proc. Sem. Seismic Zoning Maps, UNESCO, 27 October – 4 November 1975, Skopje, Yugoslavia*, Vol. 1, 193-205.

Seeber, L., and J.G. Armbruster (1981), Great detachment earthquakes along the Himalayan arc and long-term forecasting. **In:** D.W. Simpson and P.G. Richards (eds.), *Earthquake Prediction – An International Review*, Maurice Ewing Series, Vol. 4, Am. Geophys. Union, Washington, 259-277, DOI: 10.1029/ME004p0259.

Shanker, D., R.B.S. Yadav, and H.N. Singh (2007), On the seismic risk in the Hindukush–Pamir–Himalaya and their vicinity, *Curr. Sci.* **92**, 11, 1625-1630.

Thingbaijam, K.K.S., P. Chingtham, and S.K. Nath (2009), Seismicity in the North-West Frontier Province at the Indian-Eurasian plate convergence, *Seismol. Res. Lett.* **80**, 4, 599-608, DOI: 10.1785/gssrl.80.4.599.

Tsapanos, T.M. (1997), Regional variation of the ω values in the circum-Pacific belt*, Pure Appl Geophys.* **150**, 1, 113-120, DOI: 10.1007/s000240050066.

Tsapanos, T.M. (2001), Evaluation of the seismic hazard parameters for selected regions of the world: the maximum regional magnitude, *Ann. Geophys.* **44**, 1, 69-79, DOI: 10.4401/ag-3615.

Tsapanos, T.M., and P.W. Burton (1991), Seismic hazard evaluation for specific seismic regions of the world, *Tectonophysics* **194**, 1-2, 153-169, DOI: 10.1016/0040-1951(91)90278-Z.

Tsapanos, T.M., Y. Bayrak, H. Cinar, G.Ch. Koravos, E. Bayrak, E.E. Kalogirou, A.V. Tsapanou, and G.E. Vougiouka (2014), Analysis of largest earthquakes in Turkey and its vicinity by application of the Gumbel III distribution, *Acta Geophys*. **62**, 1, 59-82, DOI: 10.2478/s11600-013-0155-8.

Uhrhammer, R.A. (1986), Characteristics of northern and central California seismicity, *Earthq. Notes* **57**, 1, 21 (abstract).

Utsu, T. (1971), Aftershocks and earthquake statistics (III): Analysis of the distribution of earthquakes in magnitude, time and space with special consideration to clustering characteristics of earthquake occurrence (1), *J. Fac. Sci. Hokkaido Univ., ser. 7*, **3**, 5, 379-441.

Wessel, P., and W.H.F. Smith (1995), New version of the generic mapping tools, *EOS Trans. Am. Geophys*. *Union* **76**, 33, 329, DOI: 10.1029/95EO00198.

Woessner, J., and S. Wiemer (2005), Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* **95**, 2, 684-698, DOI: 10.1785/0120040007.

Yadav, R.B.S. (2009), Seismotectonic modeling of NW Himalaya: A perspective on future seismic hazard, Ph.D. Thesis, Department of Earthquake Engineering, IIT Roorkee, India, 198 pp.

Yadav, R.B.S., D. Shanker, S. Chopra, and A.P. Singh (2010), An application of regional time and magnitude predictable model for long-term earthquake prediction in the vicinity of October 8, 2005 Kashmir Himalaya earthquake, *Nat. Hazards* **54**, 3, 985-1014, DOI: 10.1007/s11069-010-9519-4.

Yadav, R.B.S., Y. Bayrak, J.N. Tripathi, S. Chopra, and E. Bayrak (2012a), Regional variation of the ω-upper bound magnitude of GIII distribution in Hindukush–Pamir Himalaya and the adjacent regions: A perspective on earthquake hazard, *Tectonophysics* **544-545**, 1-12, DOI: 10.1016/j.tecto.2012.03.015.

Yadav, R.B.S., Y. Bayrak, J.N. Tripathi, S. Chopra, A.P. Singh, and E. Bayrak (2012b), A probabilistic assessment of earthquake hazard parameters in NW Himalaya and the adjoining regions, *Pure Appl. Geophys*. **169**, 9, 1619-1639, DOI: 10.1007/s00024-011-0434-8.

Yadav, R.B.S., T.M. Tsapanos, G.Ch. Koravos, Y. Bayrak, and K.D. Devlioti (2013a), Spatial mapping of earthquake hazard parameters in the Hindukush–Pamir Himalaya and adjacent regions: Implication for future seismic hazard, *J. Asian Earth Sci*. **70-71**, 115-124, DOI: 10.1016/j.jseaes.2013.03.007.

Yadav, R.B.S., T.M. Tsapanos, Y. Bayrak, and G.Ch. Koravos (2013b), Probabilistic appraisal of earthquake hazard parameters deduced from a Bayesian approach in the northwest frontier of the Himalayas, *Pure Appl. Geophys*. **170**, 3, 283-297, DOI: 10.1007/s00024-012-0488-2.

Yadav, R.B.S., G.Ch. Koravos, T.M. Tsapanos, and G.E. Vougiouka (2015), A probabilistic estimate of most perceptible earthquake magnitudes in the NW Himalaya and adjoining regions, *Pure Appl. Geophys.* **172**, 2, 197-212, DOI: 10.1007/s00024-014-0864-1.

Yegulalp, T.M., and J.T. Kuo (1974), Statistical prediction of the occurrence of maximum magnitude earthquakes, *Bull. Seismol. Soc. Am.* **64**, 2, 393-414.

Gravity Data from the Teboursouk Area ("Diapirs Zone", Northern Tunisia): Characterization of Deep Structures and Updated Tectonic Pattern

Fatma HACHANI[1], Hadhemi BALTI[2], Ali KADRI[2], and Mohamed GASMI[2]

[1]Water Researches and Technologies Centre (CERTE), Technopark of Borj Cedria, Soliman, Tunisia; e-mail: hachani_fatma@yahoo.fr

[2]Department of Earth Sciences, Faculty of Sciences of Bizerte, University of Carthage, Tunisia

A b s t r a c t

Located between eastern segments of the Atlas and Tell-Rif orogenic belts, the "Dome zone" of northern Tunisia is characterized by the juxtaposition of various structures that mainly controlled the long geodynamic history of this part of the south-Tethyan Margin. To better understand the organization and deep extension of these structures, gravity data from the Teboursouk key area are proposed. These data include the plotting of Bouguer anomaly map and related parameters such as vertical and horizontal gradients, upward continuation and Euler solution. Compared to geological and structural maps available, they allow the identification of new deep structures and greater precision regarding the characteristics and organization of known ones; consequently, an updated structural pattern is proposed.

Key words: Northern Tunisia, Teboursouk, gravity data, deep structure, diapir, faults.

1.   INTRODUCTION

Northern Tunisia is an extension of the Tell-Rif eastern segments and the north-Atlasic domain (Fig. 1A). It is considered as one of the important sectors where the complex geological history of the Maghrebian south-Tethyan Margin can be reconstructed. In fact, at least since the Triassic period, this area has undergone many tectonic episodes that alternated through time. Edi-
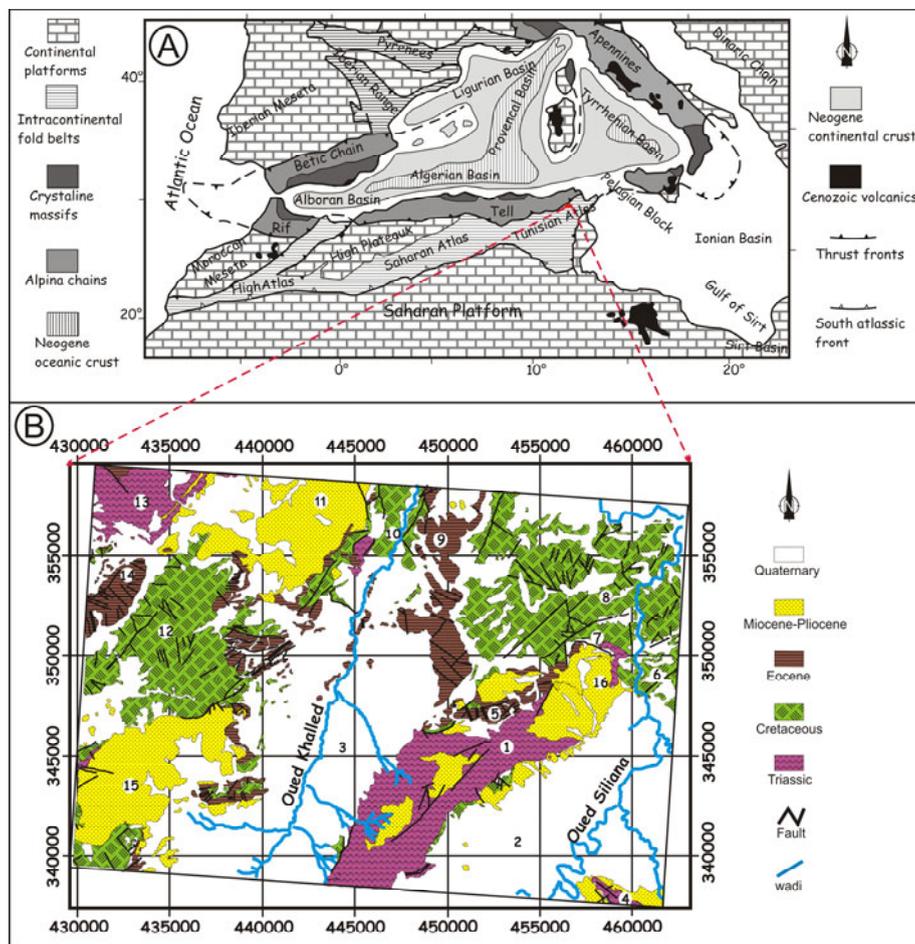
Fig. 1: (A) Tectonic pattern of the western Mediterranean domain (modified from Doglioni *et al*. 1999), (B) geological map of the study area (extracted from geological map of Teboursouk, 1:50 000). 1 –Jebel Ech Cheid, 2 – plain of El Aroussa, 3 – plain of Khalled, 4 – jebel Strassif, 5 –jebel Chetlou, 6 – Koudiat el Mahjbia, 7 – Knana, 8 – Tabet Ech Cherif, 9 – Ain Tounga Khalled, 10 – Ain Jemala, 11 – Oued Hermoucha, 12 – Oued Arkou, 13 –Thibar, 14 – Jebel Goraa, 15 – Ain Hamra-Dougga, 16 – Draa el Kantra.

fied structures were repeatedly affected by successive phases leading to the present-day structure. A set of various structures at different scales can record signatures of the main steps of this geodynamic evolution. The Teboursouk area belongs to the "Diapirs zone" of northern Tunisia which is adjacent to the northernmost Atlasic domain. Since the 1920s, particular attention has been paid to this key sector by many geologists interested in thematic investigations aiming at the reconstruction of geodynamic evolution of the Tunisian Tethys Margin and its controlling factors (Biely *et al.* 1974, Ben Ayed 1994, Chihi 1995, Gueguen *et al.* 1998, Guiraud *et al.* 2005, Rigane and Gourmelen 2011). The stratigraphical succession disturbance, the tectonic style, and the onset mode of the characteristic Triassic material are still a matter of debate; the drowning or burial of many structures makes it difficult to follow the continuation of structural units. In the gravimetric study, the subject of this paper, we propose to follow the distribution of density contrasts due to depressions and faults aiming to increase our understanding of the deep geological structure using unpublished gravity data (Hachani 2006). Interpretation of the Bouguer anomaly map and related derivative compilations serve as a base for deep structures' characterization; an updated tectonic pattern is proposed for the area replaced in its broader geodynamic context.

2. GEOLOGICAL AND PALAEOGEOGRAPHICAL CONTEXT

In northeastern Tunisia, the Teboursouk area (672 km²) is located one hundred kilometres to the NW of Tunis (Fig. 1B), between Lambert coordinate lines: $X = 430$ km and $X = 462$ km; $Y = 338$ km and $Y = 359$ km).

As a part of the "Diapirs zone", it records many stratigraphic and structural signatures of the geodynamic evolution of the eastern Maghreb Tethyan margin. By Middle to Upper Triassic, the northern Tunisia passive margin registered the first echoes of a distensive phase, better expressed after the fragmentation of the Liassic initial carbonate platform (Solignac 1927, Bolze 1950, Rouvier 1985, Ben Ayed 1986). Consequent subsidence reached its maximum during Middle Jurassic and continued until Lower Cretaceous, all in a Tethyan rifting context related to Africa–Eurasia relative movement. As a result, a mosaic of high blocks and adjacent depressions, separated by reactivated Hercynian faults, characterizes the main palaeogeographical features of the area. A northward more subsiding area, the "Tunisian Trough", developed in a tilted-block-structural pattern allowing progressive ascentional movement of Triassic material. This material has pierced its sedimentary cover since the Albian times (Chitta 1979, Perthuisot 1979, Rouvier 1985, Dali 1997) and underlined NE-SW-trending major faults.

Table 1

Lithological synthesis of Teboursouk

Age	Stages	Formations	Lithostratigraphic descriptions	Densities [g/cm^3]
Quaternary (40 to 50 m)			limestone, clay, silt et conglomerates	
Neogene	Pliocene (20 m)		conglomerates	$1.7<d<2.38$
	Miocene (1300 m)	Kechabta-Oued el Khedim	sandstone, clay, rare conglomerates	
		Oued Mellah	marl and gypsum	
		Hakima	sandstone, breccia, conglomerates, marls	
Paleogene	Oligocene (75 m)	Fortuna	sandstone	
	Upper Eocene (35 m)	Souar	marls, nummulitic limestones	$d > 2.4$
	Ypresien-lower Lut (150 m)	Metlaoui	limestones and et nummulitic limestones	
	Paleocene (300 m)	el Haria	marls	$1.7<d<1.95$
Upper Cretaceous	Campanian-Maastrichtian (365 m)	Abiod	limestones, marls et marly calcareous	$d \approx 1.9$
	Santonian (500 m)	el Kef	marl and clay with rare intercalations of limestone and marly limestone	
	Coniacian (50 m)			
	Turonian (130 m)	Bahloul	black limestones finely bedded	$d \approx 2.4$
	Cenomanian (> 250 m)	Fahdene	marl and platy limestone	
Lower Cretaceous	Albian (500 m)			$2.38<d<2.41$
	Aptian (500 m)	M'Cherga	marl, marly calcareous limestone, sandstone	
	Barremian (≈200 m)		marl, marly calcareous	
Trias			gypsum, carbonates	$1.75<d<2.6$

In relation to Africa–Europe collision, northern Tunisia underwent several compressional tectonic events. A first episode, recently attributed to the Middle Eocene (Perthuisot 1974, Adil 1993), is still a matter of discussion; however, a consensus is reached on the occurrence of three major phases dated as Middle Miocene, Late Miocene, and earliest Pliocene. The most important one, referred to as the "Atlasic phase", was responsible for important folding and reactivated diapiric movement of Triassic deposits also interpreted as salt glaciers by some authors (Vila 1995). These NE-SW trending Triassic chaotic material as well as a set of curved axis folds mainly characterize the present "Diapirs zone" of northern Tunisia. In this area, anticlines are narrow and asymmetrical with faulted south-eastern flanks; synclines are wide and infilled by Neogene deposits. Four major fault types can be identified considering their general trends: N-S, NE-SW, SE-NW, and E-W (Perthuisot and Jauzein 1972, Perthuisot 1978, 1979; Hammami 1999, Kadri and Ben Haj Ali 1999, El Ouardi 2002, Chikhaoui 2002, Balti *et al.* 2014). The stratigraphic data (Table 1) includes various successions of Triassic-to-Quaternary ages with the lack of Jurassic series.

3.   GRAVITY DATA

Gravity data were gathered by the Val d'Or Sagax company for the Tunisian geological survey (ONM 2000). They concern 648 regularly distributed stations covering the map section of the Teboursouk area. The approximately mean density is one observation per square kilometer. A Lambert North projection using the Clark ellipsoid 1880 and Carthage datum was adopted. Free-air and Bouguer gravity corrections were performed using sea level datum and a reduction density of 2.4 g/cm^3. This value results from the comparison of several methods: direct measurement of densities over several samples, and using the indirect Nettleton's method.

3.1 Bouguer anomaly map

The Bouguer gravity anomaly map ($d = 2.4$ g/cm^3) of the study area shows a variation of 34 mGals. Three anomaly groups bounded by strong gravity gradients can be identified (Fig. 2) considering their trends.

❑ **NW-SE:** these anomalies consist of $\mathbf{B_1}^+$ (Tabet Ech Cherif syncline), $\mathbf{B_1}$ (along Khalled-Ain el Hamra-Dougga), $\mathbf{B_3}^-$ (Teboursouk), and $\mathbf{B_7}^-$ (El Aroussa plain);

❑ **NE-SW:** following this direction, the anomalies $\mathbf{B_2}^+$, $\mathbf{B_3}^+$, $\mathbf{B_4}^-$, and $\mathbf{B_5}^+$ are of weak gradients and coincide with the Triassic outcrops of Ain Jamela, Jebel Ech Cheid, Ain Tounga-Khalled, and Thibar, respectively. $\mathbf{B_8}^-$ characterizes the Ain el Hamra-Dougga syncline, $\mathbf{B_5}^-$ and $\mathbf{B_6}^-$ correspond to the plains of Khalled and El Aroussa;
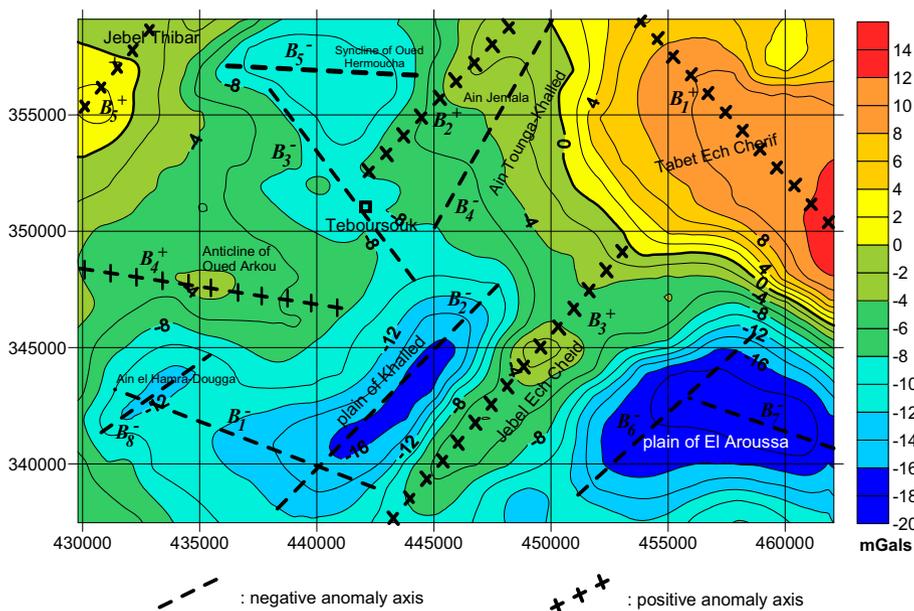
Fig. 2. Bouguer anomaly map of Teboursouk ($d = 2.4$).

- ☐ **E-W: B_4^+** and **B_5^-** overprint the Oued Arkou anticline and the Oued Hermoucha syncline, respectively.

3.2 Methods and results of interpretation of gravity anomalies

Various techniques, such as vertical and horizontal derivatives, Euler deconvolution, and upward continuation, were used to interpret the gravity map. The programs used for the gravity data interpretation in this study are the Geosoft and keating programs (Keating 1998).

Vertical derivative $\left(\dfrac{\delta g}{\delta z} \right)$

Vertical derivatives of anomalies provide a powerful technique for enhancing anomalies of smaller and near surface geological features (Evjen 1936, Baranov 1953, Marson and Klingele 1993). This derivative allows to identify the structures in the subsurface from the Bouguer anomaly. Indeed, the transformation reduces the regional element (low-frequency anomalies) which may distort or even hide the relationships near surface geological features (high-frequency anomalies) and the shape of the anomalies. Positive zones indicate rock densities exceeding the average and *vice versa*. Consequently a vision on sub-outcropping lithologies is possible and distinguished raised zones are characterized by positive anomalies where heavier platform
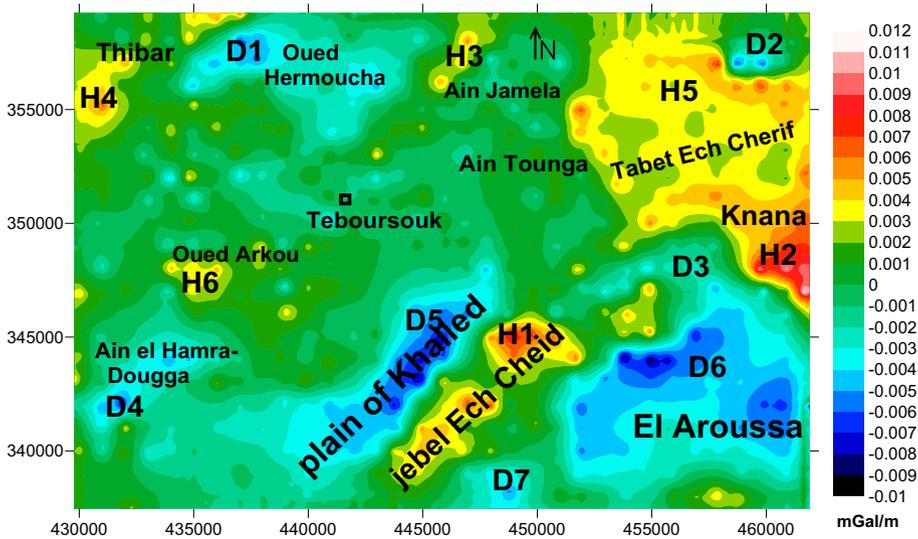
Fig. 3. First vertical derivative of the Bouguer anomaly.

is either closer to the surface or showing a light and thin cover. On the other hand, depression zones are characterized by negative anomalies indicating a light and thick cover.

Figure 3 shows that, in our case study, the raised zones are related to the Triassic outcrops (H_1 – J. Ech Cheid; H_2 – Knana; H_3 – Ain Jamela; H_4 – Thibar), to the Cretaceous series of Tabet Ech Cherif (H_5) and to the Neogene outcrops of the Oued Arkou anticline (H_6).

These raised zones strongly contrast with depressive ones filled in by light and thick series which are, from north to south: D_1 (Pliocene deposits of Oued Hermoucha), D_2 (NE the study area), D_3 (between Jebel Ech Cheid and Knana), D_4 (Ain el Hamra-Dougga), D_5 (the plain of Khalled), D_6 (two distincts subdepressions in El Aroussa), and D_7 (southern J. Ech Cheid).

The association of horizontal gradient $\left(\dfrac{\delta g}{\delta(x,y)}\right)$ and the upward continuation

The association of horizontal gradient and the upward continuation is used for the identification of the deep faults' limits and their dips (Cordell and Grauch 1985, Blakely and Simpson 1986, Blakely 1996, Phillips 1998, Archibald *et al*. 1999, Everaerts and Mansy 2001, Khattach *et al*. 2004, Najine *et al*. 2006, Chennouf *et al.* 2007, Abderbi and Khattach 2010). The Bouguer anomalies were prolonged to heights up to 4000 m with a step of 250 m. For each level, the maxima of the horizontal gradient were localized.
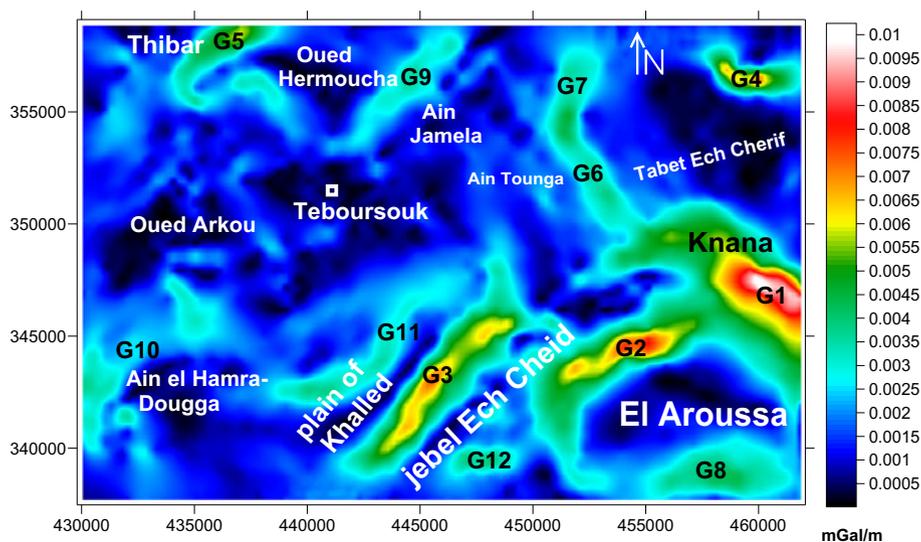
Fig. 4. Horizontal gradient map.

If the structures are vertical, all maxima are overprinting. On the other hand, the shifting of maxima for the upward extension indicates the direction of the dip.

Zones with strong gradients as shown by Fig. 4 are:

❑ **G1** (0.01 mGal/m): trending NW-SE, this zone marks a rooted and sub-outcropping fault (loose iso-anomalous contours) between Cretaceous series (north to the rift: high zone) and those of El Aroussa basin: low zones. This limit would correspond to the NE edge of El Aroussa;

❑ **G2** (0.008 mGal/m) and **G3** (0.007 mGal/m) on both sides of Jebel Ech Cheid are of a ENE-WSW and NE-SW trend;

❑ **G4** (0.007 mGal/m) trending E-W, overprints Oued Siliana and indicates a vertical deep fault separating Tabet Ech Cherif from a basin located to the NE of the study area. This fault underlines the contact between Triassic outcrops and Cretaceous series of Tabet Ech Cherif.

Other zones with weaker horizontal gradients (< 0.006 mGal/m) are also shown in Fig. 4. These consist of: **G5** (south-eastern Thibar, trending NE-SW); **G6** and **G7** (Tabet Ech Cherif; NW-SE and NNE-WSW, respectively); **G8** (south to El Aroussa, NNE-SSW); **G12** (Ech Cheid, E-W); **G11** (plain of Khalled, flexure), **G9** (North Ain Jamela, NE-SW), and **G10** (NW flank and the periclinal termination of Ain el Hamra-Dougga syncline).

The superposition of the maxima of horizontal gradients (Fig. 5) shows that some faults are sub-vertical: **F1**, **F2**, **F3**, and **F4** delimiting the Jebel
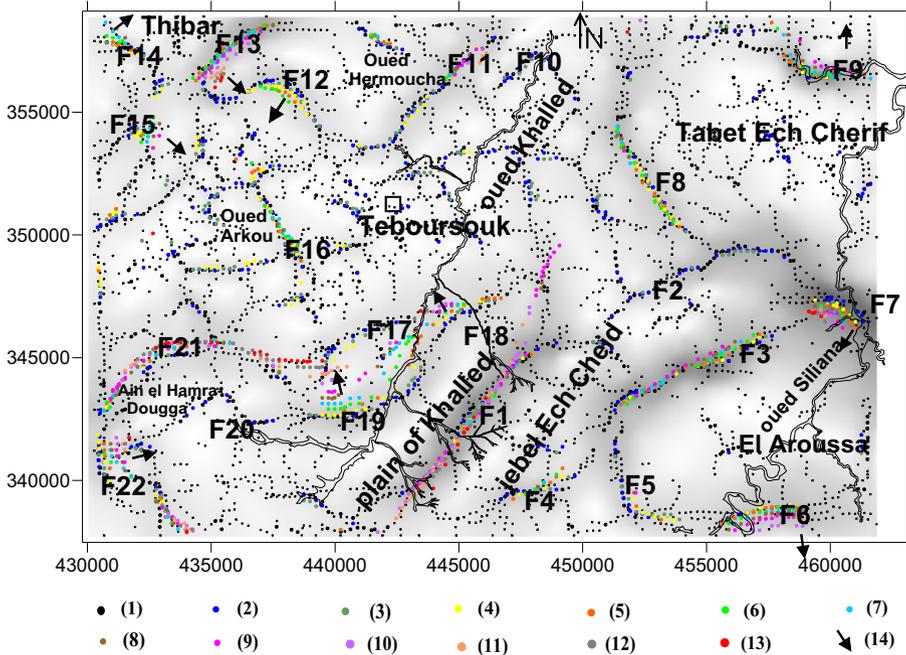
Fig. 5. Superposition of horizontal gradient maxima obtained from the Bouguer anomaly and from upward continuation map at different altitudes: (2) 250 m, (3) 500 m, (4) 750 m, (5) 1000 m, (6) 1250 m, (7) 1500 m, (8) 1750 m, (9) 2000 m, (10) 2500 m, (11) 3000 m, (12) 3500 m, (13) 4000 m, and (14) dip direction.

Ech Cheid; **F5** and **F17** affecting the plains of El Aroussa and Khalled, respectively; **F16** close to the Oued Arkou anticline; **F8** crossing the Tabet Ech Cherif syncline, and **F11** in Oued Hermoucha. The other faults show various dipping trends; from east to west they are: **F9** (N-dipping, located to the north-eastern extremity of the sector); **F7** (SW-dipping, NE J. Ech Cheid nearby Ragoubet el Mahjbia); **F6** (S-dipping, north to J. Strassif); **F18** and **F19** (NW-dipping, plain of Khalled), **F12** (SW-dipping, Oued Hermoucha); **F13** (SE-dipping, SW J. Thibar); **F14** (NE-dipping, Jebel Thibar), **F15** (SE-dipping, NW J. Goraa), and **F22** (east-dipping, Ain el Hamra-Dougga syncline).

The digitization and statistic processing of the gravimetric contacts (Fig. 6: Map 1 and panel A) reveal four fault trending groups: N-S, E-W, N30 to N70, and N120 to N150. This is in line with the geological map structures where the N40, E-W, N140, and N-S trends predominate (Fig. 6: Map 2 and panel B). The outline length of structures is nevertheless different.
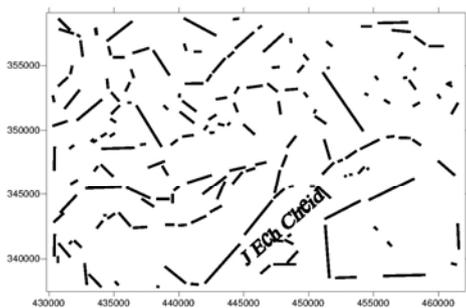
Map1 : Gravimetric lineaments map and **(A)** its
statistical analysis.
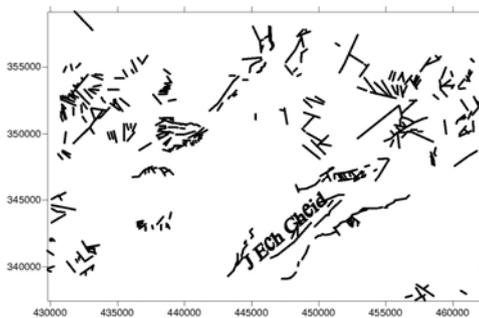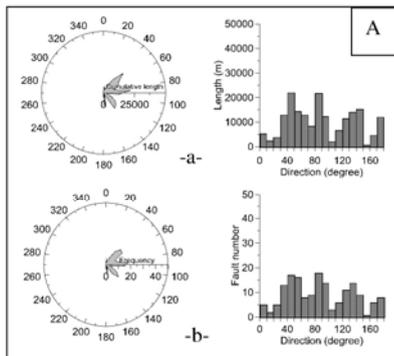
Map2 : Observed lineament map in geology and
(B) its statistical analysis.
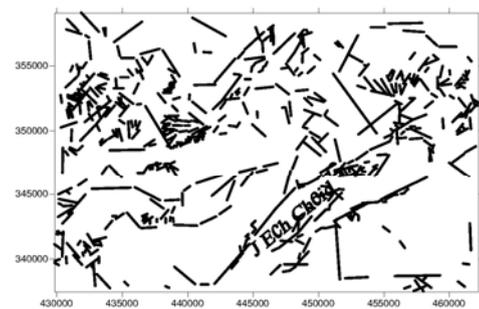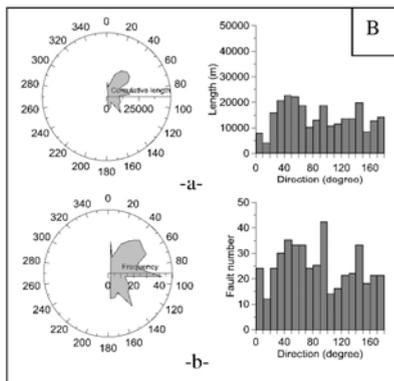
Map3 : Deduced structural map and **(C)** its
statistical analysis
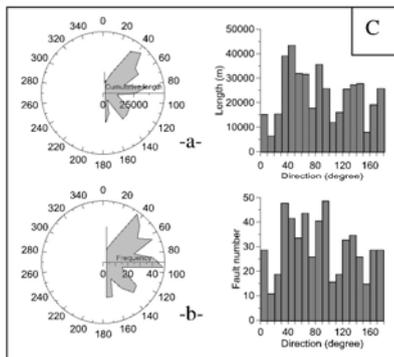
Fig. 6. Statistical analysis of accident directions obtained on the various maps:
(a) rose and histogram of cumulated lengths in meter, (b) rose and histogram of fre-
quencies.

As far as frequencies are concerned (Fig. 6: panels A and B), the identified trends are ordered from high to low as follows: E-W, N40, N135, and N-S. The final structural map obtained shows a predominance of the trends E-W, N30 to N60, N130 and N-S (in number) and of the N40, E-W, N140, and N-S trends (for the length outlines) (Fig. 6: Map 3 and panel C).

Among the identified structures, some fault outlines obtained from the processed gravimetric data and those plotted on the geological map of Teboursouk (Fig. 6: Map 1) are identical (*e.g.*, NE-SW-trending ones of southern J. Chehid Triassic outcrops) or complementary (Fig. 6: Maps 2 and 3).

The Euler deconvolution

The Euler deconvolution allows the horizontal location and the depth estimation of contacts (Thompson 1982, Reid *et al.* 1990, Keating 1998, Asfirane-Haddadj and Galdeano 2000, Mushayandebvu *et al.* 2004, Khattach *et al.* 2004, Chennouf *et al.* 2007, Saibi *et al.* 2012). It depends essentially on grid spacing, size of the deconvolution windows, and structural index (Barbosa *et al.* 2002, Stavrev and Reid 2007, Melo *et al.* 2013). Based on the calculation of horizontal gradient $\left(\dfrac{\delta g}{\delta(x,y)}\right)$, and the vertical gradient $\left(\dfrac{\delta g}{\delta z}\right)$, the Euler equation is written as follows:

$$\left(x - x_0\right)\frac{\delta g}{\delta x} + \left(y - y_0\right)\frac{\delta g}{\delta y} + \left(z - z_0\right)\frac{\delta g}{\delta z} = -NT\left(x, y\right) ,$$
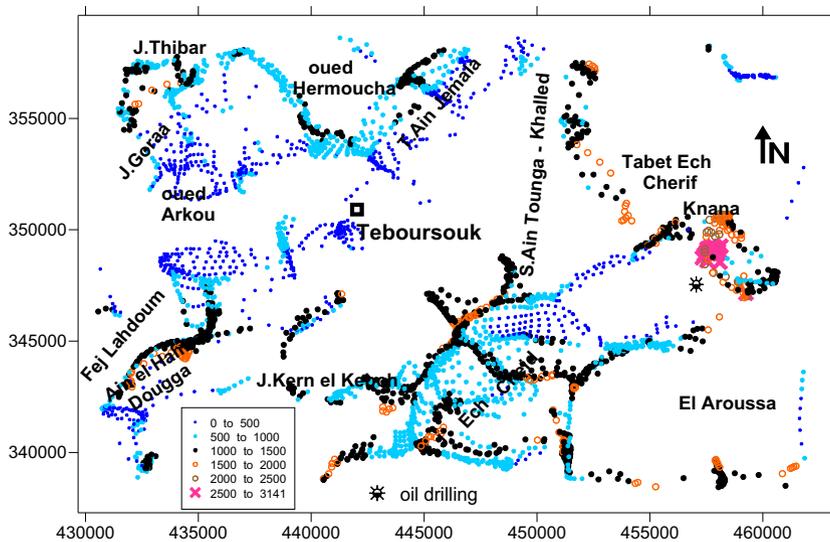
Fig. 7. Euler solution map (structural index of 0.25, 10 × 10 windows, maximum relative error of 15%).

where (x_0, y_0, z_0) are the coordinates of a gravimetric source, g is the intensity of the measured field to (x, y, z), N is the structural index and refers to the source geometry.

For good horizontal localization and source alignment we chose a structural index (SI) of 0.25.

In our case study (Fig. 7), the Euler deconvolution highlights the presence of some faults reaching 3141 m depth. The dominating faults in this area are ranging between 500 and 1500 m. The Euler solution highlights also that the Khalled basin is limited by a deep faults raging from 1500 to 2000 m.

However, several faults previously identified are absent in the Euler solution.

This is may be due to the fact that densities do not contrast enough with depths; the interference of close source responses is also a potential cause.

4. DISCUSSION

As a synthesis, the results obtained from the Euler solution, correlated to the geological data, allow to propose a new structural map for the Teboursouk area (Fig. 8a) which shows:

❑ NE-SW trending faults of important depths around J. Ech Cheid (**F1**, **F2**, and **F3**), Fej Lahdoum (**F21**), and Thibar (**F13**). Their depth varies from 500 to 1500 m, sometimes reaching 2000 m. These faults are controlling the upward movement of Triassic material. Elsewhere, the NE-SW trending faults are less deep. For the Teboursouk major fault (**F10**), the gravity data show an insignificant depth in the study area; however, the seismic reflection data have shown an important depth exceeding 1500 m (TOTAL 1978, Gasmi 2002).

❑ E-W trending major faults are less numerous; however, they have an important role in the structure of the area, such as from north to south: (i) in the north-eastern the study area, the fault **F9** is north-dipping with a depth exceeding 1000 m. It coincides with Oued Siliana and may explain its sudden change in flow direction. The relationship of the fault **F9** with the Triassic outcrop on this site is more than probable; (ii) on the eastern edge of J. Ech Cheid, along the Koudiat Tellilet-Koudiat Guella alignment, the 1500 m deep fault **F23** could be responsible for the arched aspect of Jebel Ech Cheid. The Lakhouat NS-trending fault, well individualized to the south of the study area, may have also played the common role. To the south, from Koudiat Mejemba to Koudiat Bazina Serira, the 500 to 2000 m deep fault **F4**, would be the extension of the fault **F5** of northern J. Strassif; (iii) both **F19** and **F17 faults**, with depths up to 1500 m, affect the plain of Khalled and the syncline of Ain el

Hamra-Dougga, respectively, and are interpreted as responsible for the torsion of this structure.

❑ **N-S**-trending faults are either hidden by Quaternary deposits of El Aroussa (depth = 500 to 2000 m (**F5**)); or delimited to the east and to the west, the Eocene series of Ain Tounga (**F7** and **F8**, depth = 1500 m).

❑ **NW-SE**-trending faults are the deepest and show high dipping angles. The vertical **F9** affects Tabet Ech Cherif and is considered to be at the origin of the important gradient (Fig. 3). The SW-dipping fault **F7** corresponds to the north-eastern limit of El Aroussa graben. **F24** constitutes the northern extension of this graben, bordering to the south of the Triassic outcrop of Knana. In this latter locality, major measured depths are from 2000 to 3141 m. Indeed, oil exploration works revealed important Tertiary series (1800 m) overlying the 1200 m thick Aptian–Maastrichtian successions. Eventually, in the Oued Hermoucha syncline, the faults **F12** and **F25** have depths of about 1500 m.
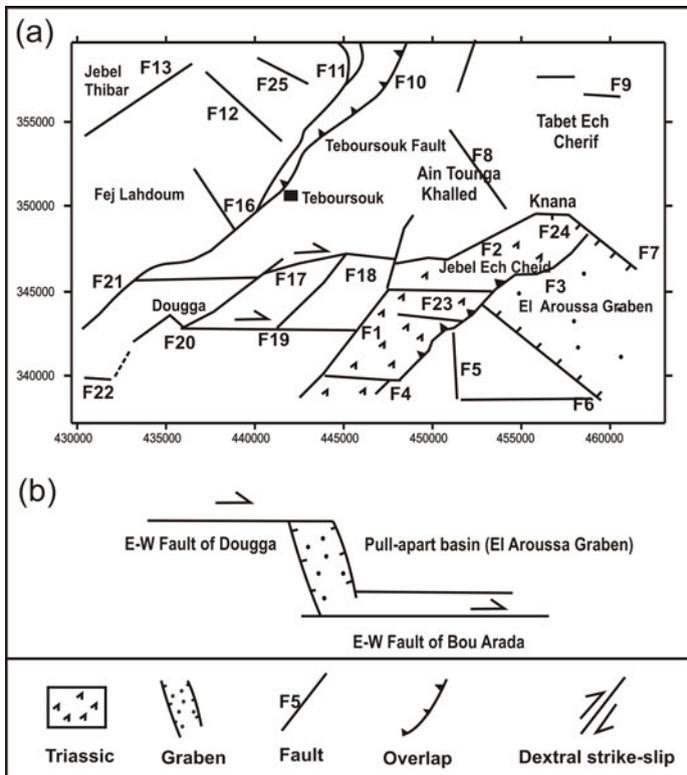
Fig. 8: (a) Synthetic structural map of Teboursouk area, (b) synthetic structure of el Aroussa graben.

The principal directions described above can be integrated into the local geodynamic context controlled mainly by the NW-SE Atlasic phase of upper Miocene (Tortonian). This compression leads to the individualization of NE-SW reverse faults and to the reactivation of old normal faults during the Tertiary. During the Aptian, the upward movement of the diapiric structures was activated. As shown in the gravity data, faults bordering the Triassic outcrops can reach 2000 m. This pleads in favor of a deep origin of the Triassic outcrop and defends its diapiric character.

In the south-eastern region, two subparallel normal faults NW-SE limit the graben of El Aroussa. The opening of this graben as a pull apart structure is guided by the E-W dextral strike-slip fault of Dougga and Bou Arada (Fig. 8b). The strike-slip fault of Dougga is linked to Fej Lahdoum one in the east. The N-S sinister strike-slip fault of Fej Lahdoum is the cause of the local derivation of the diapiric structure of Ech Cheid. This tectonics style is typical in the Tunisian Atlas (Jauzein 1967, Ben Ayed 1975, Chihi 1995).

In the western Mediterranean context, the occurrence of graben in northern and central Tunisia is related to the geodynamic evolution of the Maghreb eastern part (Algeria and Tunisia) and the Pelagian Block expansion, from Miocene to actuel. This expansion integrates as part of the general geodynamic evolution of the central and western Mediterranean by the subduction-collision systems of African and Eurasian plates (Chihi 1995).

5. CONCLUSION

The qualitative and quantitative analyses of the gravity data have a great interest in the structural studies. Thus, these techniques allowed specifying the structure and highlighting the fault system affecting the Teboursouk area. First of all we deduct the diapiric character of the J. Ech Cheid Triassic and probably the Thibar one, mainly explained by the major deep faults delimiting it (2000 m). In this study, a deep E-W fault network (1500 m), going from Fej Lahdoum (no. 18, 19) to J. Ech Cheid (no. 2, 23, 4, and 5) was also highlighted. During the upper Miocene, where the Atlas compressive phase is oriented NW-SE, and the old Quaternary (post villafranchienne), the Dougga E-W faults are dextral strike-slip and affected the architecture of Teboursouk area. Thereby, the megafold axes in Ain Hamra-Dougga, Kern el Kebch, and the J. Ech Cheid anticline are curved and vary from NE-SW to E-W direction in the neighbourhood of the E-W faults.

In Teboursouk area, the NW-SE faults are usually buried. This fault network highlighted in this study by the gravity data, affects the substratum and locally appear at the surface when it is reactivated. The fault **F10** (Fig. 8a) representing the NE edge of El Aroussa graben seems to be active in the Miocene and delimits a subsiding zone where the Tertiary series reaches

1800 m (TOTAL 1978). The El Aroussa graben is an extensional relay between the E-W Dougga fault and the E-W Bou Arada fault located in the SE of the study area.

Finally, the architecture of this study area has been controlled by the reactivation of the pre-existing deep faults. Thereby, several tectonic structures have been implemented. Gravity data processing is useful in delineating these structures (rootedness and amplitudes) and in mapping, with greater precision (depth and dip), lineations, intrusive and faults within basement rocks that are covered by sedimentary strata. These basement features tend to localize zones of weakness and, therefore, are important to the siting of man-made structure (*i.e.*, dams, water drilling, boreholes).

References

Abderbi, J., and D. Khattach (2010), Structural setting of the High Plateaus; insights from gravity data analysis (Eastern Morocco), *Bull. Inst. Sci., Rabat, Sci. Terre* **32**, 19-30 (in French).

Adil, S. (1993), Salt Triassic dynamics in northern Tunisia: Pull-apart basins form at stepover, magmatism and mining implication, Ph.D. Thesis, University of Tunis el Manar, Tunisia, 249 pp.

Archibald, N., P. Gow, and F. Boschetti (1999), Multiscale edge analysis of potential field data, *Explor. Geophys.* **30**, 1/2, 38-44, DOI: 10.1071/EG999038.

Asfirane-Haddadj, F., and A. Galdeano (2000), The use of Euler deconvolution and analytic signal for location of magnetic sources, *Bull. Soc. Géol. France* **171**, 1, 71-81 (in French).

Balti, H., F. Hachani, and M. Gasmi (2014), Hydrogeological potentiality assessment of Teboursouk Basin, Northwest Tunisia using electrical resistivity sounding and well logging data, *Arab J. Geosci.* **7**, 7, 2095-2914, DOI: 10.1007/s12517-013-1001-7.

Baranov, V. (1953), Vertical gradient calculation of the gravity field or magnetic field measured at the surface, *Geophys. Prospect.* **1**, 3, 171-191, DOI: 10.1111/j.1365-2478.1953.tb01139.x (in French).

Barbosa, V.C.F., J.B.C. Silva, and W.E. Medeiros (2002), Practical applications of uniqueness theorems in gravimetry: Part II – Pragmatic incorporation of concrete geologic information, *Geophysics* **67**, 3, 795-800, DOI: 10.1190/1.1484523.

Ben Ayed, N. (1975), Geological study of Siliana and Sers synclines (central Tunisian Atlas), Ph.D. Thesis, Pierre and Marie Curie University, Paris VI, France, 82 pp.(in French).

Ben Ayed, N. (1986), *Tectonic Evolution of the Front Alpine Chain from the Mesozoic to Current*, Annales des Mines et de la Géologie de Tunisie, Vol. 32, 286 pp. (in French).

Ben Ayed, N. (1994), The convergence of E-W and N-S convergents overthrusting strike slip fault of northern Tunisia: geometry and attempt to reconstruction of deformation conditions. **In:** *Proc. 4th Tunisian Petroleum Exploration Conference, May 1994, Tunis, Tunisia*, 25-37 (in French).

Biely, A., P.F. Burollet, and T. Lajmi (1974), Geodynamic studies of Tunisia and neighboring areas of the Mediterranean, *Notes Serv. Géol. Tunisie* **41**, 23-38 (in French).

Blakely, R.J. (1996), *Potential Theory in Gravity and Magnetic Applications*, Cambridge University Press, Cambridge, 464 pp.

Blakely, R.J., and R.W. Simpson (1986), Approximating edges of source bodies from magnetic or gravity anomalies, *Geophysics* **51**, 7, 1494-1498, DOI: 10.1190/1.1442197.

Bolze, J. (1950), Triasic diapirs and orogenic phases in the Teboursouk mountains (Northern Tunisia), *C. R. Acad. Sci. Paris* **231**, 8, 480-482 (in French).

Chennouf, T., D. Khattach, A. Milhi, P. Andrieux, and P. Keating (2007), Main structural trends in northeastern Morocco: The contribution of gravimetry, *C. R. Geoscience* **339**, 6, 383-395, DOI: 10.1016/j.crte.2007.03.006 (in French and abridged English version).

Chihi, L. (1995), Neogene-Quaternary grabens of Tunisia in the Pelagian Sea: their signification in the central mediterranean geodynamic setting, Ph.D. Thesis, University of Tunis II, Tunisia, 324 pp. (in French).

Chikhaoui, M. (2002), "Diapir zone" in Tunisia: structure and geodynamic evolution of the Meso-Cenozoic series and the Triassic salt body geometry, Ph.D. Thesis, University of Tunis, Tunisia, 318 pp. (in French).

Chitta, N. (1979), Micropaleontological study of the cretaceous (Aptian to Cenomanian) in Teboursouk region (Northern Tunisia), Ph.D. Thesis, Pierre and Marie Curie University, Paris VI, France, 192 pp. (in French).

Cordell, L., and V.J.S. Grauch (1985), Mapping basement magnetization zones from aeromagnetic data in the San Juan Basin, New Mexico. **In:** W.J. Hinze (ed.), *The Utility of Regional Gravity and Magnetic Anomaly Maps*, Society of Exploration Geophysicists, 181-197.

Dali, T. (1997), Geological study of Gafour region, Ph.D. Thesis, Pierre and Marie Curie University, Paris VI, France (in French).

Doglioni, C., M. Fernandez, E. Gueguen, and F. Sabat (1999), On the interference between the early Apennines–Maghrebides backarc extension and the Alps–Betics orogen in the Neogene geodynamics of the Western Mediterranean, *Bol. Soc. Geol. Ital.* **118**, 75-89.

El Ouardi, H. (2002), Origin of lateral variations of the Ypresian deposits in the dome zone in northern Tunisia, *C. R. Geoscience* **334**, 2, 141-146, DOI: 10.1016/S1631-0713(02)01704-2.

Everaerts, M., and J.L. Mansy (2001), Gravimetry contribution in the Boulonnais and Artois areas (France), *Bull. Soc. Géol. France* **172**, 3, 267-274 (in French).

Evjen, H.M. (1936), The place of the vertical gradient in gravitational interpretations, *Geophysics* **1**, 1, 127-136, DOI: 10.1190/1.1437067.

Gasmi, M. (2002), Geophysical contributions in recognition and managment of natural resources in Tunisia, Ph.D. Thesis, National School of Engineers of Sfax, Tunisia, 478 pp. (in French).

Gueguen, E., C. Doglioni, and M. Fernandez (1998), On the post-25 Ma geodynamic evolution of the western Mediterranean, *Tectonophysics* **298**, 1-3, 259-269, DOI: 10.1016/S0040-1951(98)00189-9.

Guiraud, R., W. Bosworth, J. Thierry, and A. Delplanque (2005), Phanerozoic geological evolution of Northern and Central Africa: An overview, *J. African Earth Sci.* **43**, 1-3, 83-143, DOI: 10.1016/j.jafrearsci.2005.07.017.

Hachani, F. (2006), Geophysical study by gravimetry of Jebel Ech Cheid and its surroundings (Northern Tunisia): structural implications, M.Sc. Thesis, Faculty of Sciences of Bizerte, Tunisia, 119 pp. (in French).

Hammami, M. (1999), Tectonic, halokinesis and mineralizations in the diapir zone (Northern Tunisia), Ph.D. Thesis, University of Tunis El Manar, Tunisia, 197 pp. (in French).

Jauzein, A. (1967), Contribution to the geological study of the confines of Tunisian dorsal (Northern Tunisia), *Ann. Mines Géol.* **22**, 475 (in French).

Kadri, A., and M. Ben Haj Ali (1999), Elements of thought on est-west and nord-sud tectonic lineaments and associated grabens in northern Tunisia, *Note Service Géol. Tunisie* **65**, 131-140 (in French).

Keating, P.B. (1998), Weighted Euler deconvolution of gravity data, *Geophysics* **63**, 5, 1595-1603, DOI: 10.1190/1.1444456.

Khattach, D., P. Keating, M. Mili, T. Chennouf, P. Andrieux, and A. Milhi (2004), Contribution of gravimetry to the study of Triffa basin structure (northestern Morocco): hydrogeological implications, *C. R. Geoscience* **336**, 16, 1427-1432, DOI: 10.1016/j.crte.2004.09.012 (in French and abridged English version).

Marson, I., and E.E. Klingele (1993), Advantages of using the vertical gradient of gravity for 3-D interpretation, *Geophysics* **58**, 11, 1588-1595, DOI: 10.1190/1.1443374.

Melo, F.F., V.C.F. Barbosa, L. Uieda, V.C. Oliveira Jr., and J.B.C. Silva (2013), Estimating the nature and the horizontal and vertical positions of 3D magnetic sources using Euler deconvolution, *Geophysics* **78**, 6, J87-J98, DOI: 10.1190/geo2012-0515.1.

Mushayandebvu, M.F., V. Lesur, A.B. Reid, and J.D. Fairhead (2004), Grid Euler deconvolution with constraints for 2D structures, *Geophysics* **69**, 2, 489-496, DOI: 10.1190/1.1707069.

Najine, A., M. Jaffal, K. El Khammari, T. Aifa, D. Khattach, M. Himi, A. Casas, S. Badrane, and H. Aqil (2006), Contribution of gravimetry to the structural study of the Tadla basin: hydrogeological implications, *C. R. Geoscience* **338**, 10, 676-682, DOI: 10.1016/j.crte.2006.04.015 (in French and abridged English version).

ONM (2000), Gravimetric campaign CG3. Teboursouk 1:50 000 map, National Office of Mines, Tunis, Tunisia.

Perthuisot, V. (1974), Middle Eocene in Teboursouk region (diapir zone, northern Tunisia); paleogeographic and structural implications, *Bull. Soc. Géol. France* **S7-16**, 3, 319-323, DOI: 10.2113/gssgfbull.S7-XVI.3.319 (in French).

Perthuisot, V. (1978), Dynamic and petrogenesis of Triassic salt extrusions in northern Tunisia, Ph.D. Thesis, Pierre and Marie Curie University, Paris VI, France, 312 pp. (in French).

Perthuisot, V. (1979), Explanatory notes for geological map of Tunisia. Sheet 33, Teboursouk, National Office of Mines, Tunisia, 52 pp. (in French).

Perthuisot, V., and A. Jauzein (1972), El Alia-Tbessa fault in Teboursouk region, *Notes Serv. Géol., Trav. Géol. Tunis.* **8**, 58-63 (in French).

Phillips, J.D. (1998), Processing and interpretation of aeromagnetic data for the Santa Cruz Basin – Patahonia Mountains area, south-central Arizona, Open-File Rep. 2002-98, U.S. Geological Survey, Reston, U.S.A.

Reid, A.B., J.M. Allsop, H. Granser, A.J. Millett, and I.W. Somerton (1990), Magnetic interpretation in three dimensions using Euler deconvolution, *Geophysics* **55**, 1, 80-91, DOI: 10.1190/1.1442774.

Rigane, A., and C. Gourmelen (2011), Tunisian Transtensive Basins in Tethyan Geodynamic Context and Their Post-Tortonian Inversion. **In:** U. Schattner (ed.), *New Frontiers in Tectonic Research – At the Midst of Plate Convergence*, InTech, Rijeka, 27-48, DOI: 10.5772/18833.

Rouvier, H. (1985), *Geology of Extreme North Tunisia: Tectonics and Paleogeographies Superposed to the Eastern Part of the Northern Maghreb Chain*, Annales des Mines et de la Géologie de Tunisie, Vol. 29, 425 pp. (in French).

Saibi, H., E. Aboud, and S. Ehara (2012), Analysis and interpretation of gravity data from the Aluto-Langano geothermal field of Ethiopia, *Acta Geophys.* **60**, 2, 318-336, DOI: 10.2478/s11600-011-0061-x.

Solignac, M. (1927), *Geological Study of Northern Tunisia*, Université Lyon and Direction Générale Travaux Publics, Tunis, 756 pp. (in French).

Stavrev, P., and A. Reid (2007), Degrees of homogeneity of potential fields and structural indices of Euler deconvolution, *Geophysics* **72**, 1, L1-L12, DOI: 10.1190/1.2400010.

Thompson, D.T. (1982), EULDPH: A new technique for making computer-assisted depth estimates from magnetic data, *Geophysics* **47**, 1, 31-37, DOI: 10.1190/1.1441278.

TOTAL (1978), Synthetic structural map of Teboursouk permit (P. De Genevraye), Report No. E7759-421, ranking No. 4454, French Oil and Gaz Co. TOTAL, Tunisia.

Vila, J.M. (1995), First terrestrial study of a large submarine salt glacier – The eastern part of the Ouenza-Ladjebel-Meridef structure (Algerian Tunisian Confines) – Proposal for an emplacement scenario and comparisons, *Bull. Soc. Géol. France* **166**, 2, 149-167 (in French).

Comparison of Factorial Kriging Analysis Method and Upward Continuation Filter to Recognize Subsurface Structures – A Case Study: Gravity Data from a Hydrocarbon Field in the Southeast Sedimentary Basins of the East Vietnam Sea

Mohammad-Reza AZAD[1], Mohammad KONESHLOO[1,2], Abolghasem KAMAKAR ROUHANI[1], and Hamid AGHAJANI[1]

[1]Shahrood University, Faculty of Mining and Petroleum Engineering and Geophysics, Shahrood, Iran

[2]Department of Chemical and Petroleum Engineering, University of Wyoming, Laramie, Wyoming, USA; e-mail: mkoneshl@uwyo.edu

A b s t r a c t

To interpret geophysical anomaly maps, it is necessary to filter out regional and sometimes noise components. Each measured value in a gravity survey consists of different components. Upward continuation (UC) is one of the most widely used filters. The shortcoming of this filter is not to consider the spatial structure of the data, and also the fact that the trial and error approach and expert's judgment are needed to adjust it. This study aims to compare the factorial kriging analysis (FKA) and UC filters for separation of local and regional anomalies in the gravity data of a hydrocarbon field in the southeast sedimentary basins of the East Vietnam Sea. As shown in this paper, FKA method permits to filter out all of the identified structures, while the UC filter does not possess this capability. Therefore, beside general and classic filtering methods, the

FKA method can be used as a strong method in filtering spatial structures and anomaly component.

Key words: anomaly separation, geophysical filtering, factorial kriging analysis (FKA), upward continuation (UC), hydrocarbon exploration.

1. INTRODUCTION

The gravity method is a very useful method in oil, gas, and mineral exploration. Despite the vast use of seismic method in oil exploration, the gravity method is always one of the initial steps in hydrocarbon exploration to recognize favourable areas. This method is usually applied before conducting next exploration steps using advanced methods such as seismic ones and drilling. Considering the fact that the density of oil-gas environments is less than their surroundings, the resulting gravity anomalies can be detected. It has also a realistic significance in discovering sedimentary basins, local structures, and areas with an oil-gas potential (Berezkin 1973, McCulloh 1980, Zeng *et al*. 2002, Tran 2004, Aghajani *et al*. 2011).

It is very important in the interpretation of gravity data for hydrocarbon exploration to recognize a sedimentary basin that can be a possible hydrocarbon area (Reynolds 1997). In addition, we know that the Bouguer anomaly maps contain three components: noise, regional, and local anomalies (Telford *et al*. 1990). Many different methods with varying methodology bases (Nettleton 1954) are used for separation of regional from residual (local) anomalies in order to distinguish areas that may be suitable for the formation of hydrocarbon reservoirs. In fact, these methods separate the local anomaly related to the oil-gas reservoir from regional anomaly.

One of the most widely known methods to extract regional anomaly is called the upward continuation (UC) method that is based on the weighted averages. The weights are calculated using upward continuation level (for details see Telford *et al*. 1990). UC can be regarded as estimated airborne survey using the ground survey. When the UC method is applied to a ground survey dataset, in fact, the data are converted as if taken from an airborne survey in an upward height or level above the ground survey. There are some problems with the method, such as the fact that the height should be known when using the method Zeng *et al*. (2007).

Factorial kriging analysis (FKA) is a variogram-based filtering technique (Matheron 1982). FKA is used to filter data with different spatial structures (Sandjivy and Galli 1984, Sandjivy 1984). Galli applied this method as a substitute to spectral analysis of magnetic data (Galli *et al*. 1984). Seguret (1993) has used FKA to extract diurnal variation from aeromagnetic data. Moreover, some experts used the FKA to reorganize geological structures on

gravity data, and also to filter seismic noisy data (Jeannée and Mari 2008, Magneron *et al*. 2009).

   This work aims to present theory of the FKA method and its application to separate gravity anomalies in a hydrocarbon field in the southeast sedimentary basins of the East Vietnam Sea. After a brief description of the study area, the variographic study of data is presented, then the FKA and UC results are demonstrated, and finally the results of the two methods are compared considering the geological features of this hydrocarbon field.

2. THEORY OF THE USED METHODS

2.1 FKA

Conventional methods (*e.g.*, UC method) have some practical limitations; firstly, a preliminary interpolation (*e.g.*, kriging) is needed to estimate the value of variable on a complete rectangular grid, when data are not on a regular grid. Secondly, their interpretation and implementation become tedious by increasing the dimension of space, and finally, by ignoring the spatial variation of the component of frequency, a local anomaly can widespread through all frequencies (Chilès and Delfiner 1999). To solve this problem, Matheron (1982) proposed an orthogonal decomposition of a regionalized variable $Z(x)$ to its components $Z_s(x)$. That means, a random function as $Z(x)$ can be assumed as an intrinsic regionalization model with mostly stationary components, in which $Z_s(x)$ is an intrinsic component (Wackernagel 1988). In this case,

$$\gamma(h) = \gamma\left(x_\alpha - x_\beta\right) = E\left\{\left[Z\left(x_\alpha\right) - Z\left(x_\beta\right)\right]^2\right\} = E\left\{\left[\sum_{i=1}^{s} Z^i\left(x_\alpha\right) - \sum_{i=1}^{s} Z^i\left(x_\beta\right)\right]^2\right\}, \quad (1)$$

where x_α and x_β represent the location of any point of α and β, $Z(x_\alpha)$ and $Z(x_\beta)$ are the measured values at these points. If the structures are assumed to be mutually uncorrelated, then this formula can be written as a summation of the components' variograms.

$$\gamma(h) = E\left\{\sum_{i=1}^{s}\left(Z^i\left(x_\alpha\right) - Z^i\left(x_\beta\right)\right)^2\right\} = \sum_{i=1}^{s} E\left(Z^i\left(x_\alpha\right) - Z^i\left(x_\beta\right)\right)^2 = \sum_{i=1}^{s}\gamma^i\left(x_\alpha - x_\beta\right). \quad (2)$$

   This kind of spatial structure is called nested variogram model, as schematically shown in Fig. 1. Matheron (1982) demonstrated the formula of this new method, named FKA, in the case of monovariable and also developed it for the case of heterotopic multivariable. The spatial structure, characterized by a covariance/variogram model, is a space domain equivalent of density spectrum in the frequency domain (Yao 2004); thus, FKA relies on the de-

(a) (b)
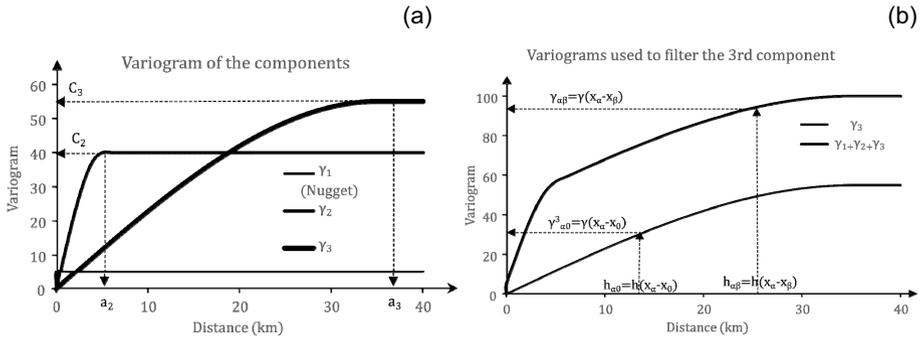
Fig. 1. Each measured gravity value of $Z(x)$ could be assumed as the summation of three different orthogonal components; as an example: noise or very short scale component as $Z^1(x)$, local as $Z^2(x)$, and regional component as $Z^3(x)$, where $Z(x) = Z^1(x) + Z^2(x) + Z^3(x)$; (a) the range and sill of variograms are shown, respectively, by a_i and C_i, the first structure is a pure nugget effect and has no range; (b) overall variogram is used in the right hand matrix of kriging to calculate variogram value between the known points and variogram of 3rd component (to be filtered) is used to complete the right hand matrix of kriging including the variogram between the known points and target.

composition of variograms, which permits to overpass the above-mentioned limitations for conventional methods for interpretation of field potential data.

Goovaerts (1997) has presented FKA in "dual form" and gave an example of the filtering properties of kriging. Wackernagel (1995) demonstrated the FKA formula to filter out the intrinsic and stationary components of a regionalized variable. An example for univariate case is the measured value in a gravity survey; the measured value could be factorized to its components which have different wavelengths. Value of intrinsic component at each point of x_0 denoted as $Z(x_0)$ can be estimated using the values of the measured variable in a given neighbourhood $Z(x_\alpha)$:

$$Z_s^*(x_0) = \sum_{\alpha=1}^{n} w_\alpha^s Z(x_\alpha) \ . \tag{3}$$

The reader is referred to Wackernagel (1988, 1995), Goovaerts (1997, page 165-166), and Chilès and Delfiner (1999, pages 342-345) for details of the mathematical framework of this method.
1. Non bias condition: The expectation of the estimation error should be nil;
2. Minimizing the variance of error;
3. Non-correlation condition between components.

$$\sigma_E^2 = \sum_{u=0}^{s-1} C^u \left(x_\alpha - x_\beta \right) - \gamma^s \left(x_\alpha - x_\beta \right) - \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} w_\alpha^s w_\beta^s \gamma^s \left(x_\alpha - x_\beta \right) + 2 \sum_{\alpha=1}^{n} w_\alpha^s \gamma^s \left(x_\alpha - x_0 \right) , \tag{4}$$

$$\sum_{\beta=1}^{n} w_\beta^s \ \gamma \left(x_\alpha - x_\beta \right) + \mu_s = \ \gamma^s \left(x_\alpha - x_0 \right) \quad \text{for} \ \ \alpha = 1, 2, ..., n \ ,$$

$$\sum_{\beta=1}^{n} w_\beta^s = 1 \ . \tag{5}$$

where w_β^s represents the weight of kriging for each point of x_β, $\gamma(x_\alpha, x_\beta)$ is the value of variogram between each two pairs of points of x_α and x_β; in the same way, $\gamma^s(x_\alpha, x_0)$ is equal to the value of variogram for the distance between x_α and x_0 (see Fig. 1b), and finally μ_s is the LaGrange multiplier.

The filtering system, expressed by Eq. 5, is similar to the system of ordinary kriging, but the variogram value of right-hand side matrix of this equation, $\gamma^s(x_\alpha, x_0)$, is calculated only using the spatial structure that must be filtered.

2.2 UC

There are many different methods concerned with separation of the regional and local or residual anomalies in geophysics, especially in gravity and magnetic geophysical techniques. The UC method is used frequently to identify regional anomalies and gravity/magnetic variations of deeper horizons (Claerbout 1986). This method is applied to filter out short-wavelength (or high-frequency) anomalies, and also, to reduce their amplitudes and decrease the noise related by near surface features (Reynolds 1997).This filter or operator is a transformation of the potential field anomaly calculated at a height greater than the observed field $A(x, y, z)$, that is, a measured or observed potential field at a location (x, y, z) in space. If there is no source in the upper half space, then (Gibert and Galdeano 1985):

$$\Delta \left[A(x, y, z) \right] = 0 \quad \text{where} \quad z \geq 0 \ . \tag{6}$$

Potential field in different points is expressed as follows (Jacobsen 1987):

$$A(x, y, z) = \frac{z}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{A(\xi, \eta, 0)}{\left[(x - \xi)^2 + (y - \eta)^2 + z^2 \right]^{3/2}} d\xi d\eta \ , \tag{7}$$

where ξ and η are the coordinates. Equation 7 can be written as a convolution (Gibert and Galdeano 1985) given by Eq. 8:

$$A(x, y, z) = A(x, y, 0) * P_z(x, y) \, , \tag{8}$$

where * denotes convolution; hence, the Fourier transformation of Eq. 8 is equal to the Fourier transform of the first term multiplied by the Fourier transform of the second terms (Everett 2013), which is

$$P_z(x, y) = \frac{z}{2\pi \left(x^2 + y^2 + z^2 \right)^{3/2}} \, , \tag{9}$$

$A(x, y, 0)$ is the upward continued potential field, *i.e.*, the potential field taken to the height 0, that is, higher than the height z related to the location of the observed or measured field $A(x, y, z)$; it should be noted that the direction of z axis downward is positive.

The UC operator is a filter that eliminates high frequency components and, in contrast, the low-frequency components will be strengthened. In particular, the UC filter tries to filter out the shorter wavelength anomalies that are regarded as near surface structures, and the subsequent process is highlighting the effect of deep subsurface structures. Although one of the most widely used filtering techniques in the processing and interpretation of potential field data is the UC method, there are some disadvantages too, the most important being the fact that it does not consider the spatial structure of the data, and also the lack of knowledge about the optimum UC height for a particular case. In this paper, a geostatistical filtering method, called FKA, is used instead of the UC method. Then each of the structures (including anomalies and noise) can be separated using the FKA theory explained in detail in Section 2.1. In the following, we will first briefly review of the study area and then will apply the methods separately, and finally, compare the results.

3. OVERVIEW OF THE STUDY AREA

3.1 Geology of the study area

The study area is located in the southeastern part of a Vietnamese basin situated in the east of Vietnam Sea (Fig. 2a). The studied field covers an area of 215×215 km. There are numerous oil-gas boreholes drilled in the basin. The prior studies have shown some oil-gas reservoirs developed in the Eocene-Oligocene and the Early Miocene formations with the marine-river source, but many of them have been concentrated in the broken rock of the Pre-Cenozoic strata. The multiple oil-gas reservoirs are developed at different depths and connected together horizontally. In the area, the thickness of sedimentary layer varies very strongly, from 1 km along the coastline to 13 km in the center of the sedimentary basin. Many studies have pointed out that the rocks fill up the grabens in the Eocene-Oligocene strata, which are in

(a) (b)
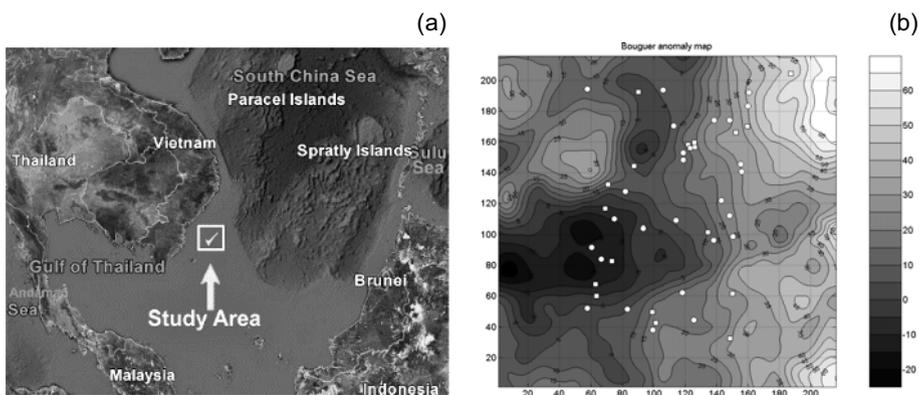
Fig. 2: (a) Location of the 215×215 km study area (Tran 2004), and (b) Bouguer anomaly map and drilled wells position map of the study area (○ – wet bore, □ – dry bore). The origin of the map coordinate system is located in the lower-left corner of the map, with positive values extending along the axes up and to the right of the origin point. The values of *X* and *Y* coordinates are in km. The contour interval is 5 mGals. Negative Bouguer anomalies are indicated by darker colors.

contact with the horst of carbonate basement rock, and create the potential oil-gas reservoirs (Que and Tran 1996, Vu 2003, Tan and Bo 2002, Tran 2004).

3.2 Gravity data

The gravity anomalies have been measured on a very fine grid with a high quality. The gravity map has been drawn with a contour interval of 2 mGals, and all the necessary computations in this regard have been performed on a grid of 1×1 km (Fig. 2b). The used data are drawn by resampling from this exhaustive data base on a grid of 30×30 km to investigate the capability of the FKA and UC methods on the larger grid. The reason for the increase of the grid size from 1×1 km to 30×30 km was the long time of the execution of the FKA computer program when the huge gravity data taken from 1×1 km grid size was used in the program. However, when the grid size increased, a considerable reduction in the run time was observed due to the reduction of the number of gravity data in the grid. Although different grid sizes were tested, the 30×30 km grid size containing much less data than 1×1 km grid size was the largest grid size for which its FKA results were comparable with the UC results. The values of the gravity anomalies vary from –30 to +90 mGals. The negative anomalies indicate sedimentary materials of the basins, and have amplitudes of about tens of mGals lying in the central area but not in the sedimentary basin. It is attributed to the accretion

of the basement structure. The main trends of the gravity anomalies are northeast-southwest. Some bands of strong gravity gradient are also observed here in northeast-southwest and northwest-southeast directions, some being hundreds kilometers long.

4. GEOSTATISTICAL FILTERING

The first step in a geostatistical filtering (FKA) procedure is the spatial structure analysis. This step permits to recognize the components and model them by a nested variogram model in which each component is represented by a simple structure.

4.1 Structural analysis of the gravity data and variogram modeling

Variographic study is performed in different directions. Main directions of anisotropy are along N30E and N120E (Figs. 3a, b). Both types of anisotropy, including zonal anisotropy and geometric anisotropy, are observed. If directional variogram shows different effect ranges in several directions, this type of anisotropy is called the geometric anisotropy. Zonal anisotropy can occur when variograms in different directions suggest a different value for the sill (Wackernagel 1988). Once experimental variograms have been calculated along main directions of anisotropy, one should fit a credible model to them. The procedure of analyzing the anisotropy and choosing the anisotropy and fitting the variogram to the model is explained in many geostatistical books (*e.g.*, Armstrong 1998, Isaaks and Srivastava 1989).The best fitted model is a nested model that includes three structures or components indicated in the right hand side of the following equation:

$$\gamma(h) = 10.\text{Nugget} + 170.\text{Cub}\left(\frac{N30E}{145}, \frac{N120E}{190}\right) + 335.\text{Cub}\left(\frac{N30E}{200}, \frac{N120E}{370}\right) . \quad (10)$$

Each coefficient shows the portion of the component from the total variance of variable, and the number in the parentheses shows the main axes of anisotropy and the value of ranges (see Fig. 3). A cubic variogram model is defined by the following equation (Chilès and Delfiner 1999):

$$\gamma(h) = \begin{cases} C_0\left[7\left(\frac{h}{a}\right)^2 - \frac{35}{4}\left(\frac{h}{a}\right)^3 + \frac{7}{2}\left(\frac{h}{a}\right)^5 - \frac{3}{4}\left(\frac{h}{a}\right)^7\right] & \text{if} \quad h < a \\ C_0 & \text{if} \quad h \ge a \end{cases} \quad (11)$$

Ellipses of anisotropy are shown in Fig. 3c. The bigger ellipse concerns the larger cubic structure (the third structure indicating a regional anomaly) and the smaller one shows the second structure.
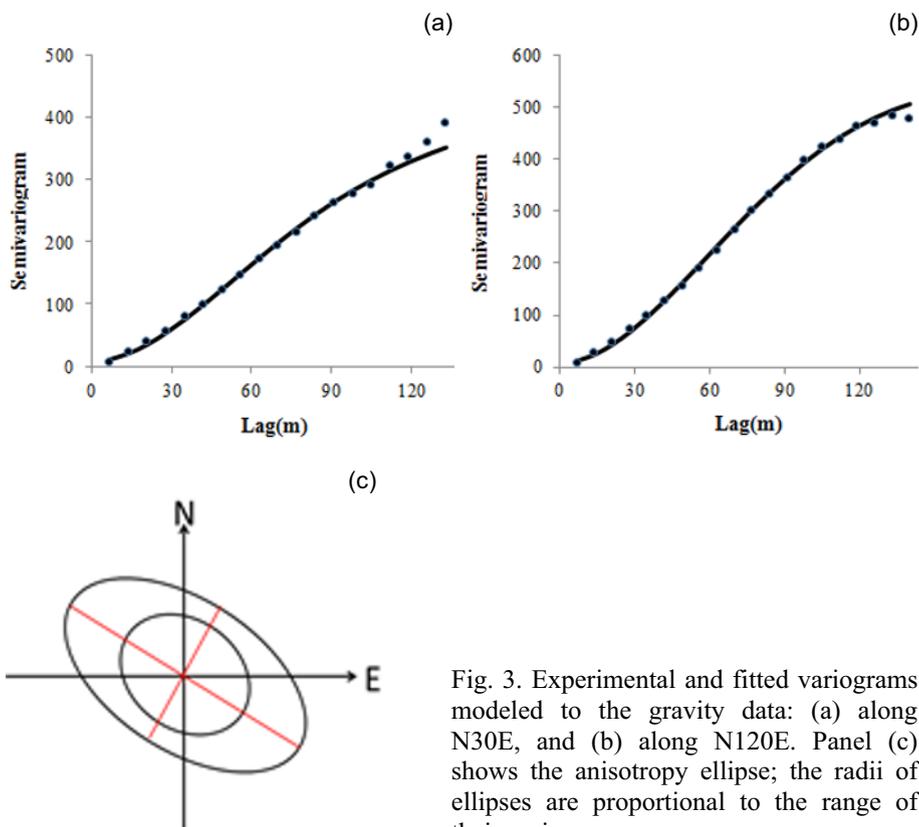
Fig. 3. Experimental and fitted variograms modeled to the gravity data: (a) along N30E, and (b) along N120E. Panel (c) shows the anisotropy ellipse; the radii of ellipses are proportional to the range of their variograms.

The portion of nugget is very small in comparison with two other structures; there are two possible interpretations for a nugget effect:

(i) The distance between gravity stations (30 km) is too large to provide detailed information about small range models, which may be associated with very high-frequency spatial variations of the gravity data due to very local geologic phenomena;

(ii) Data may be affected by a measurement error, which directly causes a discontinuity in the variogram behavior at the origin. Experimental variogram of data reveals a hierarchical structure, modeled by the mentioned nested structures (Eq. 10), each structure being characterized by its own range and direction. The short range component (145 km in N30E and 190 km in N120E) corresponds to the local changes in the Bouguer anomaly of the basin, and long-range component may be interpreted as more regional geological or sedimentation changes of the basin. Long-range components seems to be more anisotropic.

4.2 Separation of structures in the gravity data by FKA method

The purpose of this step is to filter out identified spatial components at all nodes of the grid. The 3rd structure has been filtered using Eq. 5. Figure 4a shows the filtered large-scale component results obtained as a consequence of applying FKA method on the gravity data. Then, the values of this component are subtracted from the measured gravity values to obtain the residual values to filter out the second structure. As a result, the short-scale structure component of the gravity data has been shown in Fig. 4b. The final residual structure has been assumed as the equivalent variable of the nugget effect component (Fig. 4c). The coordinates are in kilometers.

   Normally, the regional anomaly is in relation with deep structures effect and the residual anomaly points the near surface structures effect (Reynolds 1997). As can be seen in Fig. 4a, the regional component keeps a decreasing trend from east to west of the study area. An anomaly with low values is notable in the western part of the map of the regional component shown in
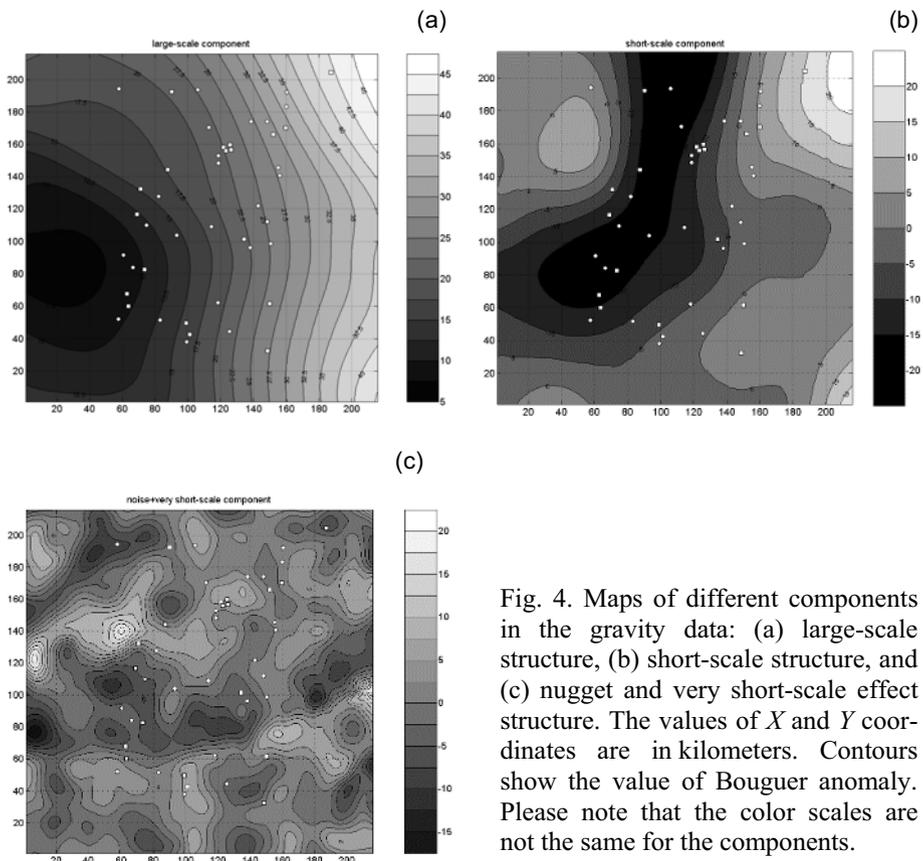
Fig. 4. Maps of different components in the gravity data: (a) large-scale structure, (b) short-scale structure, and (c) nugget and very short-scale effect structure. The values of X and Y coordinates are in kilometers. Contours show the value of Bouguer anomaly. Please note that the color scales are not the same for the components.

Fig. 4a. That may be interpreted as the effect of low-density sedimentary layers. This structure is distinct from the measured gravity data map. The short-scale component map (Fig. 4b) shows completely different features. High values in this map indicate high-density geological structures such as dense rocks or features present in the area. Moreover, a very notable negative anomaly exists that has been extended from north to the center of the map with a north-southwest trend. This anomaly can be due to the existence of low density sedimentary structures in this part of the area. This part can be regarded as a suitable zone for hydrocarbon reservoirs occurrence, as shown in the previous research by Tran (2004), and the drilled bores often reach to the oil reservoirs in the study area.

5. FILTERING THE GRAVITY DATA BY UC METHOD

Tran (2004) and Aghajani (2009) have applied normalized total gradient (NTG) method on the gravity data to detect the gravity anomalies caused by oil-gas reservoirs in the marine sedimentary basins of the study area. Tran (2004) has also pointed out that applying other processing and interpretation techniques on the gravity data is needful to check and enrich the results of NTG method. The UC filter has been applied using different heights of 15, 20, 25, 30, 35, and 40 km, and the results of upwarding for 15, 20, 25, and 30 km have been presented as the maps shown in Fig. 5.

In the interpretation of gravity data for hydrocarbon exploration, it is significant to be able to distinguish between the sedimentary basins which are good possible hydrocarbon targets, and other low-density geological structures that have no prospect for hydrocarbon. These geological structures as well as sedimentary basins produce negative gravity anomalies. In this paper, to determine the depositional areas for possible hydrocarbon occurrence, UC filter with heights of 15, 20, 25, and 30 km have been employed. As can be seen from Fig. 5, the gravity anomalies become smoother, and the anomalies embedded at intermediate depths may even disappear as the height of the UC filter increases.

Hence, we can see the main subsurface structures or anomalies in the maps obtained as a result of applying UC filter with heights of 15, 20, and 25 km; however, the effects of some of these subsurface structures are largely disappeared in the UC map with a height of 30 km. As can also be seen from all of the maps shown in Fig. 5, a lower Bouguer anomaly can be seen in the southwest and west of the study area, and an increase in the Bouguer anomaly is observed toward the east of the area. This low-value Bouguer anomaly can be interpreted as being due to low density geological structures with special geometrical shapes, such as a geo-anticline or salt domes. In contrast, high value Bouguer anomaly can be explained by an increase in the thickness or density sediments, which are located in central and eastern parts
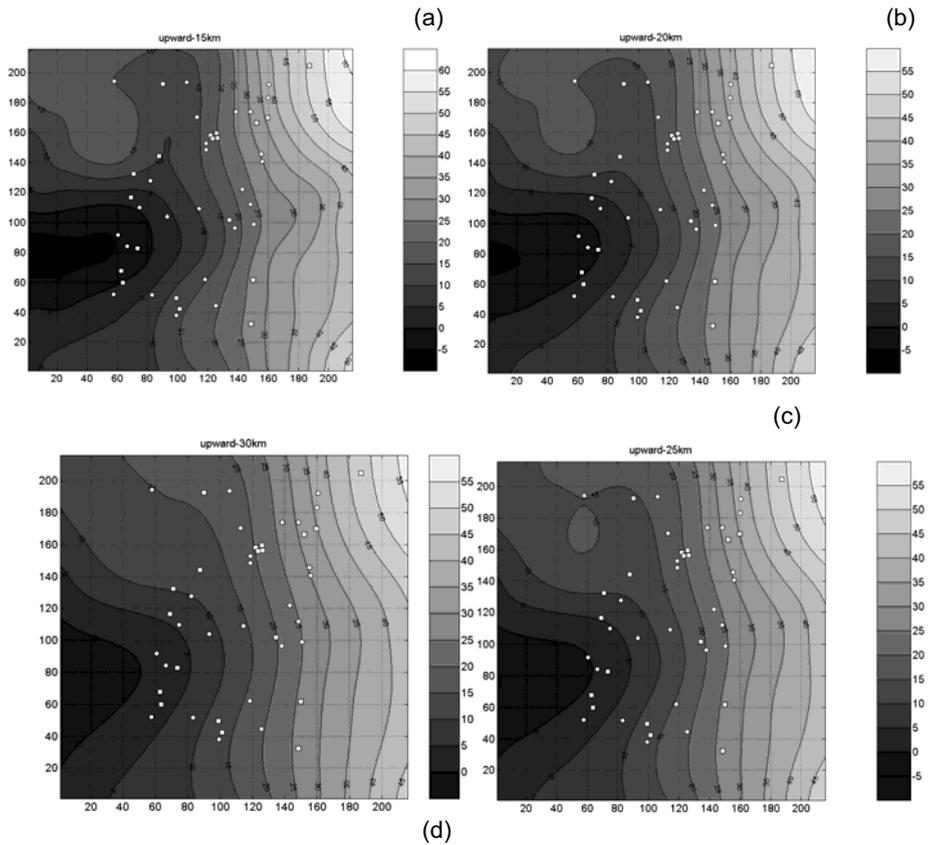
Fig. 5. The results of applying UC filter for different heights: (a) 15 km, (b) 20 km, (c) 25 km, and (d) 30 km, on the gravity data. The coordinates are in kilometers. Due to different ranges of values, the color scales are not the same.

of the study area. Selection of the height of 25 km among the other heights seems to be the best or optimum UC height to depict the regional anomaly or anomalies from the gravity data in the area. This is because the effects of the subsurface structures in the UC map obtained with a height greater than 25 km are the same as those in the UC map with the height of 25 km. Besides, Zeng *et al*. (2007) introduced a criterion or quantitative method for choosing the optimum UC height. Using this criterion or quantitative method, in which the curve of UC gravity data *versus* the UC height is drawn to obtain the optimum UC height, the height of 25 km was approximately obtained.

If we subtract the obtained regional anomaly from the Bouguer anomaly,

the residual map is obtained, in

which the local anomalies have

been   shown

(a)   (b)

(c)   (d)
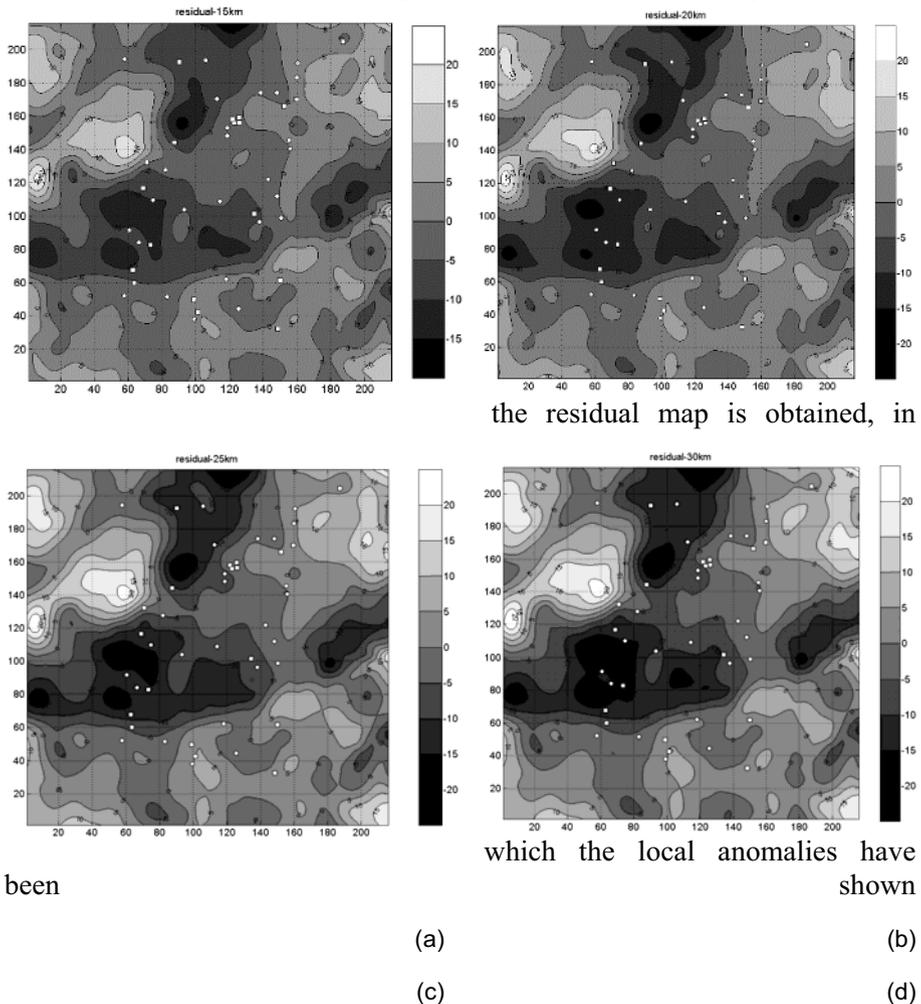
Fig. 6. Residual anomalies obtained as a result of applying UC filter for different heights: (a) 15 km, (b) 20 km, (c) 25 km, and (d) 30 km. The range of variable in panel (a) is smaler than others. The coordinates are in kilometers.

(Fig. 6). From the residual maps, shown in Fig. 5, we can identify possible location of oil-gas anomalies. Reservoirs are considered as negative residual anomaly values in the maps. Positive anomalies in 4 corners of the study area (shown in Fig. 6) can be attributed to the rocks of higher density. It is difficult to judge about the geological features of these corners of the study area as there is not enough information about the geology of the area. A re-

markable zone, identified by a negative anomaly with a north-south exten-
sion in the central part of the maps shown in Fig. 6, may be a location for the
occurrence of oil-gas reservoirs in the subsurface.

However, a correct interpretation of the maps and accurate determination
of existence of oil-gas reservoirs in the subsurface of the area require more
information, such as subsurface geological information, well logs, and drill-
ing data from the area. However, this zone, and also other negative anomaly
zones, can be considered as the preference zones for the next exploration
stages, *e.g.*, seismic surveys or drilling operation. The optimum UC height
has been selected as 25 km based on maximum deflection of the curve pre-
senting correlation between the results of two successive heights of
upwarding *versus* the height (for details see Zeng *et al*. (2007) and Guo *et al*.
(2013)). The corresponding local or residual anomaly map for the height of
25 km has been considered for comparison in the next section.

6. COMPARISON OF THE RESULTS OBTAINED FROM APPLYING FKA AND UC FILTERING METHODS

For the purpose of comparison of the two filtering methods discussed in this
paper, the obtained results of the relevant components from the two methods
have been compared together. Therefore, the results of large-scale compo-
nent of FKA have been compared with the obtained results of applying UC
with a large height. Similarly, the results of small-scale component of FKA
have been compared with the results of applying UC with a low height. Ini-
tially, the regional anomalies obtained from the two methods have been
compared together. For the UC method, the regional anomaly component
has been extracted from the results
of applying optimum UC height of
25 km to the gravity data (Fig. 7a).
In fact, the correct height of up-
ward continuation is qualitatively
justified by an expert although dis-
pute   on   which   height
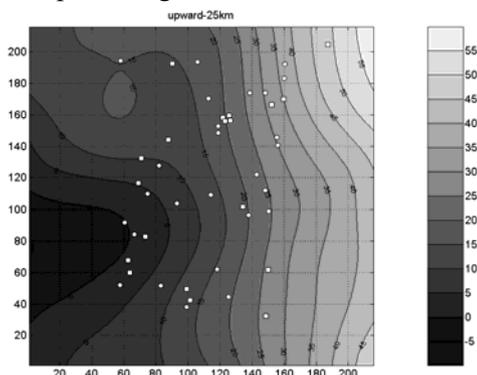
(a)

(b)
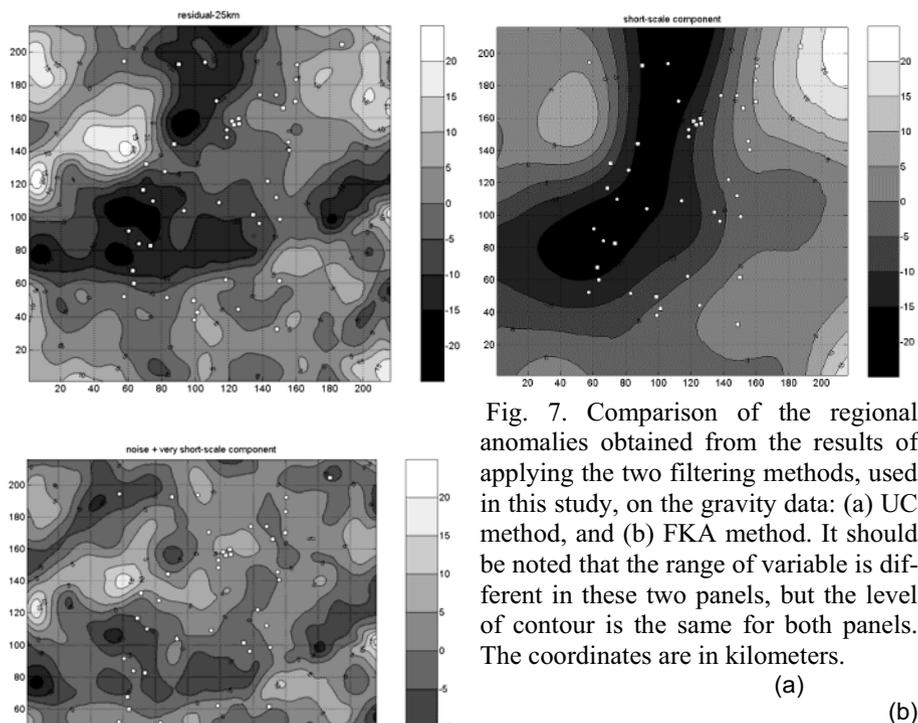
Fig. 7. Comparison of the regional anomalies obtained from the results of applying the two filtering methods, used in this study, on the gravity data: (a) UC method, and (b) FKA method. It should be noted that the range of variable is different in these two panels, but the level of contour is the same for both panels. The coordinates are in kilometers.

(a)

(b)
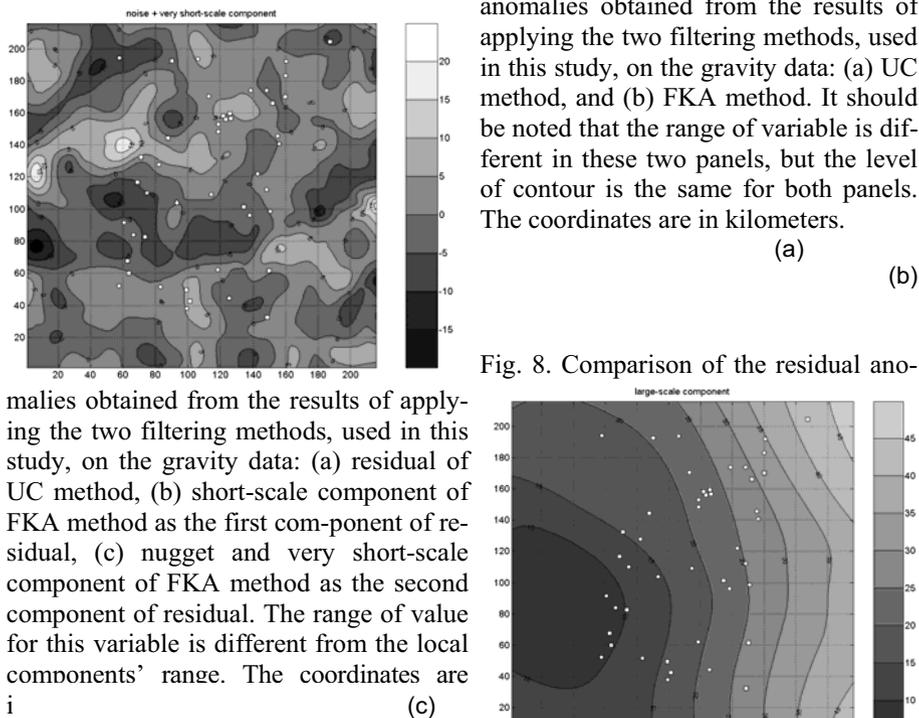
Fig. 8. Comparison of the residual ano-

malies obtained from the results of applying the two filtering methods, used in this study, on the gravity data: (a) residual of UC method, (b) short-scale component of FKA method as the first com-ponent of residual, (c) nugget and very short-scale component of FKA method as the second component of residual. The range of value for this variable is different from the local components' range. The coordinates are i

(c)

should be the optimum UC height still remains. As can be seen from Fig. 7, the two maps obtained from the results of FKA and UC methods indicate similar anomalies, trends, and quality, although slight differences between the quantities or values of the two maps are observed.

The map of short-scale component of FKA and the residual map of UC filter for a height of 25 km are shown in Fig. 8a. The negative and positive anomalies are situated in almost the same positions in both maps. However, the variable in the FKA map (Fig. 8b) is obviously more continuous than the gravity value of the residual map (Fig. 8a) obtained from the UC method.

The map of the short-scale component of FKA shows a continuous long negative anomaly extended from north to the center and then to the west of the study area. This continuous negative anomaly can be interpreted as a possible long or broad sedimentary zone with a low density compared to the densities of the surrounding rocks. Unlike the FKA map, the residual map shows various negative anomalies or zones in north, center, west, and even east of the study area, implying the existence of various local low density geological features in these parts of the area.

7. CONCLUSIONS

Separation of regional and residual anomalies from geophysical data is carried out using various methods. In this paper, the capability of FKA method to separate all the components (not just low-pass and residual as UC) interpretable as regional anomaly (large-scale component) from residual anomaly (short-scale component) in the gravity data from a hydrocarbon field in Vietnam has been demonstrated, and the obtained results have been compared with the results of applying UC filtering method on the gravity data. The UC filtering method is commonly used in gravity and magnetic geophysical methods for separation of regional from local or residual anomalies. We have also shown the superiority of FKA method over UC filter for separation of spatial structures or anomalies. The results of the FKA method depend on the decomposition of the random field model (Bouguer anomaly) into the various structures or factors. The basic difference between the UC and FKA methods is that the FKA method takes into account the spatial structure of data. The geophysical UC filter is often used in trial and error approach on the gravity and magnetic data. In every engineering calculation and modeling, the expert choices will be important and somehow applied in system. FKA is based on spatial structure, and it is possible that two experts choose two different variogram models, but these models could not be much different, as the models are adjusted to experimental variograms calculated directly from data. The FKA method is a multivariate geostatistical method that is applied for the separation of signal from noise based on the recognition of spatial structures with different scales. In addition, the FKA method permits to filter out all of identified structures while this is not applicable in the case of UC method. Furthermore, the sizes and trends of anomalies can be determined from FKA results. Therefore, in addition to general and classic filtering methods, we can also use the FKA method as a strong method in

filtering spatial structures and anomaly components, and, also, in separation of signal from noise. As the mathematical structure of FKA is similar to ordinary kriging, it can be shown that, in the same way, FKA can handle the problem of clustered data in non-uniform sampled fields, and also is generalized to *n*-dimensional space cases. The results of filtering anomalies in this research work are generally confirmed by geological findings from the study area. However, the accuracy of both FKA and UC methods in filtering anomalies in this case study should be judged when comprehensive or sufficient information becomes available from the subsurface of the area. This information can be obtained from extensive subsurface geological studies, drilling operations, well logging interpretation as well as analyses of cores obtained from numerous wells drilled in the area. Due to the existence of very local changes in the basin characteristics revealed by the closeness of wet and dry wells, a more detailed exploration could be justified in the favorable zones of this basin.

R e f e r e n c e s

Aghajani, H. (2009), Investigation of the capabilities of the normalized full gradient method for detecting the hydrocarbon potential areasusing gravity data, Ph.D. Thesis, Shahrood University of Technology, Shahrood, Iran (in Persian).

Aghajani, H., A. Moradzadeh, and H. Zeng (2011), Detection of high-potential oil and gas fields using normalized full gradient of gravity anomalies: A case study in the Tabas basin, eastern Iran, *Pure Appl. Geophys*. **168**, 10, 1851-1863, DOI: 10.1007/s00024-010-0169-y.

Armstrong, M. (1998), *Basic Linear Geostatistics*, Springer, Berlin Heidelberg, 155 pp., DOI: 10.1007/978-3-642-58727-6.

Berezkin, W.M. (1973), Application of gravity exploration to reconnaissance of oil and gas reservoir, Neda Publishing House (in Russian).

Chilès, J.-P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons, New York, 695 pp.

Claerbout, J.F. (1986), Fundamentals of geophysical data processing, *Geophys. J. Roy. Astron. Soc*. **86**, 1 217-219.

Everett, M.E. (2013), *Near-Surface Applied Geophysics*, Cambridge University Press, Cambridge.

Galli, A., F. Gerdil-Neuillet, and C. Dadou (1984), Factorial kriging analysis: A substitute to spectral analysis of magnetic data. **In:** G. Verly, M. David, A.G. Journel, and A. Marechal (eds.), *Geostatistics for Natural Resources Characterization. Part 1*, D. Reidel Publ. Co., Dordrecht, 543-557, DOI: 10.1007/978-94-009-3699-7_31.

Gibert, D., and A. Galdeano (1985), A computer program to perform transformations of gravimetric and aeromagnetic surveys, *Comput. Geosci.* **11**, 5, 553-588, DOI: 10.1016/0098-3004(85)90086-X.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 483 pp.

Guo, L., X. Meng, Z. Chen, S. Li, and Y. Zheng (2013), Preferential filtering for gravity anomaly separation, *Comput. Geosci.* **51**, 247-254, DOI: 10.1016/j.cageo.2012.09.012.

Isaaks, E.H., and R.M. Srivastava (1989), *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 561 pp.

Jacobsen, B.H. (1987), A case for upward continuation as a standard separation filter for potential-field maps, *Geophysics* **52**, 8, 1138-1148, DOI: 10.1190/1.1442378.

Jeannée, N., and J.-L. Mari (2008), Applying geostatistical filtering techniques to near-surface geophysics: two examples for refraction surveying and gravimetry. **In:** *Proc. 70th EAGE Conference & Exhibition, 9-12 June 2008, Rome, Italy*, 5 pp.

Magneron, C., M. Bourges, and N. Jeannée (2009), M-Factorial Kriging for seismic data noise attenuation. **In:** *11th Int. Congr. Brazilian Geophysical Society and EXPOGEF 2009, 24-28 August 2009, Salvador, Brazil*, 1651-1654.

Matheron, G. (1982), Pour une analyse krigeante des données régionalisées, Report N-732, Centre de Géostatistique, Fontainebleau, France (in French).

McCulloh, T.H. (1980), Mass properties of sedimentary rocks and gravimetric effects of petroleum and natural-gas reservoirs. **In:** D. Guion and C. Prieto (eds.), *Gravity Magnetic Exploration Applications*, Society of Exploration Geophysicists.

Nettleton, L.L. (1954), Regional, residuals, and structures, *Geophysics* **19**, 1, 1-22, DOI: 10.1190/1.1437966.

Que, B.C., and T.D. Tran (1996), Geodynamics and mineral potential of the offshore area of Vietnam, Final report KT-03-02, Marine Research Program, Hanoi, Vietnam.

Reynolds, J.M. (1997), *An Introduction to Applied and Environmental Geophysics*, John Wiley & Sons, 806 pp.

Sandjivy, L. (1984), The factorial kriging analysis of regionalized data: Its application to geochemical prospecting. **In:** G. Verly, M. David, A.G. Journel, and

A. Marechal (eds.), *Geostatistics for Natural Resources Characterization. Part 1*, D. Reidel Publ. Co., Dordrecht, 559-571, DOI: 10.1007/978-94-009-3699-7_32.

Sandjivy, L., and A. Galli (1984), Kriging analysis versus spectral analysis, *Inform. Geol.* **21**, 115-124.

Séguret, S.A. (1993). Analyse krigeante spatio-temporelle appliquée à des données aéromagnétiques, *Cah. Géostat.* **3**, 115-138 (in French).

Tan, T.K., and N.Q. Bo (2002), The main structural elements in Cenozoic of the continental shelf of Vietnam. **In:** *Proc. Scientific Conference "Bien Dong 2002", 16-19 September 2002, Nha Trang, Vietnam.*

Telford, W.M., L.P. Geldart, and R.E. Sheriff (1990), *Applied Geophysics*, 2nd ed., Cambridge University Press, New York.

Tran, D.T. (2004), Two and three dimensional normalized total gradient of gravity anomalies and its application for detecting the oil-gas potential areas in the southeast basins of the East Vietnam Sea. **In:** *7th SEGJ Int. Symp. "Imaging Technology", 24-26 November 2004, Sendai, Japan*, 1-6.

Vu, N.G. (2003), Structural evolution of the block 102 and 106 Red river basin-implication for hydrocarbon potential. **In:** *Proc. Conf. Vietnam Petroleum Institute: 25 Years of Development and Achievement*, 283-309.

Wackernagel, H. (1988), Geostatistical techniques for interpreting multivariate spatial information. **In:** C.F. Chung, A.G. Fabbri, and R. Sinding-Larsen (eds.), *Quantitative Analysis of Mineral and Energy Resources*, NATO ASI Series, Vol. 223, D. Reidel Publishing Co., 393-409, DOI: 10.1007/978-94-009-4029-1_24.

Wackernagel, H. (1995), *Multivariable Geostatistics: An Introduction with Applications*, Springer, Berlin Heidelberg, 387 pp.

Yao, T. (2004), Reproduction of the mean, variance, and variogram model in spectral simulation, *Math. Geol.* **36**, 4, 487-506, DOI: 10.1023/B:MATG.0000029301.22150.28.

Zeng, H., X. Meng, C. Yao, X. Li, H. Lou, Z. Guang, and Z. Li (2002), Detection of reservoirs from normalized full gradient of gravity anomalies and its application to Shengli oil field, east China, *Geophysics* **67**, 4, 1138-1147, DOI: 10.1190/1.1500375.

Zeng, H., D. Xu, and H. Tan (2007), A model study for estimating optimum upward-continuation height for gravity separation with application to a Bouguer gravity anomaly over a mineral deposit, Jilin province, northeast China, *Geophysics* **72**, 4, I45-I50, DOI: 10.1190/1.2719497.

New Methods
for Modeling Laterolog Resistivity Corrections

Jadwiga Anna JARZYNA[1], Adam CICHY[1], Dezső DRAHOS[2],
Attila GALSA[2], Maria Joanna BAŁA[1], and Andrzej OSSOWSKI[1]

[1]AGH University of Science and Technology, Faculty of Geology,
Geophysics and Environment Protection, Kraków, Poland;
e-mails: jarzyna@agh.edu.pl (corresponding author), cichy@agh.edu.pl,
bala@geol.agh.edu.pl, ossowski@geol.agh.edu.pl

[2]Eötvös Loránd University, Department of Geophysics and Space Sciences,
Budapest, Hungary; e-mails: drahos@pangea.elte.hu, gali@pangea.elte.hu

A b s t r a c t

The paper presents methods for laterolog response modeling. In Coulomb's charges method, Laplace's equation is solved for the electric field distribution in rock medium with internal boundaries between different resistivity layers. There, the boundary problem is reduced to Fredholm integral equation of the second kind. The second method uses a finite element array to model apparent resistivity from laterolog. The task is treated as DC problem and the Laplace equation is solved numerically. The presented methods were applied to borehole data covering a typical stratigraphic section of the Fore-Sudetic Monocline in southwestern Poland. Apparent resistivity was calculated using the Coulomb's charges method and alternatively modeled using a finite element method which gave similar results. Then, a series of linear corrections for borehole, shoulder bed, and filtration effects for apparent resistivity obtained by the Coulomb's charges method demonstrated the feasibility of calculating true resistivity of virgin and invaded zones. The proposed methods provide a flexible solution in modeling which can be adapted to other logs.

Key words: laterolog response, Coulomb's charges method, finite element modeling, apparent resistivity correction.

1. INTRODUCTION

The modeling of resistivity response observed from various resistivity logs employs analytical and numerical methods that come with tradeoffs. Analytical methods may require numerical solutions such as series representation of Bessel functions or numerical evaluation of integrals. Geologic structures and invasion zone features around the borehole may complicate analytical solutions. Even with recent advances in computational power, analytical methods can only be applied to simple 1D or 2D rock formation models (Drahos 1984, Anderson 2001).

Numerical solutions lend themselves more readily to modeling resistivity in complicated geologic media. They have been adapted to 2D and 3D examples of logging environments, but depend on finite difference or finite element routines that generate a large number of computationally intensive linear equations. Fast computers, if available, can easily solve these equations.

Laterolog electrode tools can be modeled using relatively simple analytical expressions and numerical solutions. Attempts have been made to numerically correct for the Groningen effect on LLD resistivity (Lovell 1993). Potential formulation based on a Poisson/Laplace equation was shown to be effective. The shape and size of electrodes present challenges in calculating laterolog resistivity responses and numerical modeling, however, and thus require certain assumptions that may affect results.

A large body of computer code has been developed to model resistivity response (Anderson 2001). Methods are based on analytical (Moran and Timmons 1957) as well as numerical solutions (Nam *et al.* 2010). Several of these methods are described here as relevant examples of efforts to improve resistivity models. Wang *et al.* (1998) designed LL3D, a 3D finite element code for computing laterolog response in arbitrary geometries. A later version of this code addressed anisotropic media. Davydycheva *et al.* (1996) developed the DC3D method, a 3D finite difference tool for computing laterolog response in a complicated geometry. This approach uses impedance boundary conditions and improves accuracy in cases of high resistivity contrasts through material averaging (Moskow *et al.* 1999).

The variety of algorithms for resistivity modeling arises not only from differences in numerical methods but also from specific approach to geological model performance from performance requirements of different geologic environments (Ribeiro and Carrasquilla 2013). The above-mentioned authors included a polynomial mathematical representation of an invaded zone having variable diameter and resistivity. A simplified method of resistivity modeling in complicated geologic structures (multiple beds, boreholes, and invaded zones) used a general recursive algorithm and local reflection and

transmission operators from a single bed boundary (Chew *et al.* 1991). In this method, computational requirements scaled linearly, rendering the method more efficient than the finite element approach to the same problem. Another approach was based on optimization of the finite element *hp* mesh where *h* was the size of the finite element and *p* was the polynomial element order of approximation (Pardo *et al.* 2005, Nam *et al.* 2010).

The above summary of analytical solutions and numerical modeling demonstrates the state of the art as well as the significance of further efforts to improve resistivity data interpretation from dual laterolog measurements. The results of two different approaches to dual laterolog resistivity modeling are presented below. These methods build on previous work and can be adapted to future refinement of resistivity analysis.

2. DUAL LATEROLOG RESISTIVITY MEASUREMENT

Dual laterolog techniques use nine electrodes operating according to galvanic conduction principles. LLD and LLS refer to respective deep and shallow operating modes of the device. These measurements are a popular and still viable version of multielectrode laterologs (Ellis and Singer 2007). Low-frequency alternating currents (AC) are used to reduce contact-impedance electrical noise and interference from electromagnetic fields. The log is typically considered a direct current (DC) measurement, however (Ellis and Singer 2007, Nam *et al.* 2010). Total current emitted by the laterolog LLD device returns to the torpedo (the lower section of the armored cable) and the N electrode (located on the bridle, an insulated cable run between torpedo and the top of the device) provides a reference voltage reading. When the torpedo and device enter a resistive unit, the N electrode voltage becomes negative (Trouiller and Dubourg 1994). For proper resistivity recording, the reference voltage should equal zero. It can become negative due to skin effects from the 35 Hz AC of the device, due to the presence of a high resistivity unit above a lower resistivity unit, or due to steel casing apparatus. Reference electrode effects have always appeared in LLD measurements and are referred to as the Delaware and Groningen effects. They cause an increase in apparent resistivity of otherwise low resistivity beds. A number of technical fixes, including additional electrodes are used to reduce these effects, but their resultant noise and uncertainties persist.

This paper considers the DLT-GA dual laterolog device made by Halliburton Co. (Halliburton 1992) and compares both LLD and LLS operating modes.

Well log analysts have encountered the problem of elevated apparent resistivity in trying to interpret a Permian unit referred to as the Main Dolomite, which appears in cores and boreholes from the Fore-Sudetic Monocline of southwestern Poland. This unit exhibits medium and low apparent resis-

tivity occurring beneath a high resistivity unit referred to as the Basal Anhydrite (Table 2). Apparent resistivity of a typical section observed in borehole data from the Fore-Sudetic Monocline was calculated using the Coulomb's charges method (Cichy and Ossowski 2015). Drahos and Galsa (2007) solved the same problem using finite element method. Results from the two different approaches are compared below. Similarities in apparent resistivity datasets from each method were interpreted as a successful correctness test.

3.   COULOMB'S CHARGES METHOD FOR MODELING RESISTIVITY

In the Coulomb's charges method, electric field distribution within rock media having internal boundaries between horizons of different resistivity is obtained by solving the Laplace equation with the following internal boundary conditions:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{r}{R}\frac{\partial V}{\partial r}\right)+\frac{\partial}{\partial z}\left(\frac{1}{R}\frac{\partial V}{\partial z}\right)=0 \ , \tag{1a}$$

$$\frac{1}{R^i}E_n^i(r,z)=\frac{1}{R^{i+1}}E_n^{i+1}(r,z) \ , \tag{1b}$$

$$E=-\nabla V \ . \tag{1c}$$

where V is the electric field potential, R^i and R^{i+1} are resistivities of homogenous layers i and $i+1$; $E_n^i(r,z)$ and $E_n^{i+1}(r,z)$ are normal components of the electrical field at boundaries separating the i and $i+1$ domains, and r and z are coordinates of the cylindrical system.

Surface electric charge density $\sigma(r,z)$ was calculated according to the formula:

$$\sigma(r,z)=2\varepsilon_0\,K(r,z)\,E_n^{\mathrm{av}}(r,z) \ , \tag{2}$$

where $K(r,z)=(R^{i+1}-R^i)/(R^{i+1}+R^i)$ is the reflectivity coefficient, $E_n^{\mathrm{av}}(r,z)=(E_n^i(r,z)+E_n^{i+1}(r,z))/2$ is an average of the electric field normal components, and ε_0 is the permittivity of free space.

The normal component of the electric field, $E_n^{\mathrm{av}}(r,z)$, is a sum of components related to primary sources, $E_n^{sc}(r,z)$, charge induced on horizontal boundaries between layers, $E_n^h(r,z)$, charge induced on cylindrical boundaries, $E_n^c(r,z)$, and charge from cylindrical electrodes of the device, $E_n^s(r,z)$,

$$E_n^{\mathrm{av}}(r,z)=E_n^{sc}(r,z)+E_n^h(r,z)+E_n^c(r,z)+E_n^s(r,z) \ . \tag{3}$$

The boundary problem for the Laplace equation with internal boundary conditions reduces to the Fredholm integral equation of the second kind (Alpin 1964, Alpin *et al.* 1985, Cichy and Ossowski 2015).
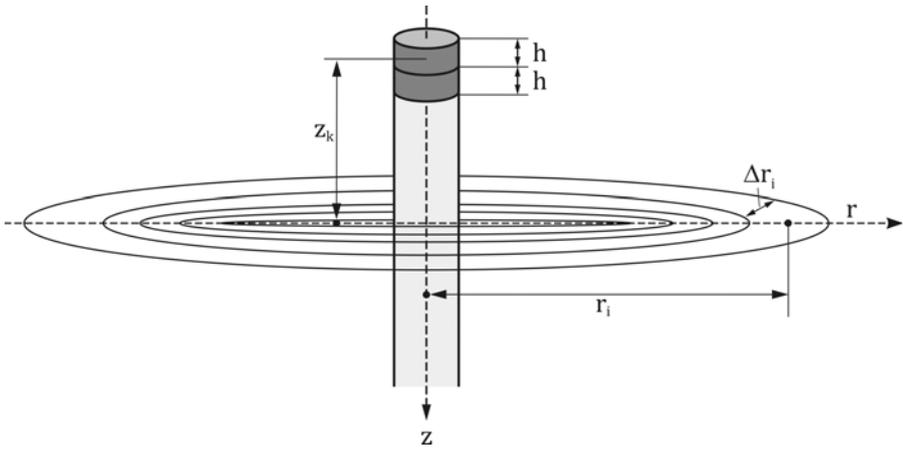
Fig. 1. Diagram showing cylindrical and horizontal boundaries for elementary cylinders and rings (after Cichy and Ossowski 2014).

In the next step, horizontal and cylindrical boundaries between layers of the rock medium that vary in terms of resistivity were discretized as elementary cylinders, rings, and horizontal layers (Fig. 1). The method then determines a set of linear equations with unknowns defined as charge density on the elementary units. The electric potential at any point in the borehole and rock formation from all current sources and charges on all elements (cylinders and horizontal surfaces) is calculated as a superposition.

The size of the model domain was determined empirically. The maximum space for calculation was determined for z values of up to 50 m and r values of up to 230 m. The constructed vertical cylinder approximated a borehole of 0.20 m diameter, with height extending both 25 m above the laterolog's central electrode A0, and 25 m below. The borehole was divided into elementary cylinders of 0.20 m diameter and 0.05 m height. Calculations addressed one thousand of such cylinders within the 50 m vertical section of the borehole. Each horizontal boundary consisted of 500 rings. The size, Δr_i, of the first ring and the next 249 ones was fixed at 0.05 m. The sizes, Δr_i, of the proceeding 250 rings increased in geometric progression by a factor of about 1.02. Iterative testing identified optimal size, height, and diameter parameters for the model. This discretization of the model allowed us to solve the system of linear equations to within 2% accuracy of calculations (Cichy and Ossowski 2015). The accuracy was determined by comparing results of the Coulomb's charges calculation with those of the analytical solution described in Alpin *et al.* (1985).

4. LATEROLOG RESPONSE

Electric current and potential were calculated as a superposition of appropriate signals from normal logs. Single normal log response was calculated in the presence of all cylindrical electrodes of the laterolog device in the borehole when only one electrode was supplied with current, I_0 (Fig. 2). Resistivity of the cylindrical electrodes was 10^{-5} ohm m. The size of cylindrical electrodes was set according to specification of the device's Technical Manual (Halliburton 1992). The device's trunk material was assumed to be an isolator. Electrical charges generated on the trunk surface were not included in the sum of electrical field components (Eq. 3). The method thus determined the influence of all electrodes and each differentiated element of the rock formation on electric field generation.

Currents listed in Fig. 2 caption satisfy Eqs. 4 and enabled proper functioning of the laterolog.

$$I_1 = I_{A1} + I_{A1'} , \qquad \text{and} \qquad I_2 = I_{A2} + I_{A2'} , \tag{4}$$

Table 1 gives current, potential and geometric conditions for the DLT-GA laterolog device.
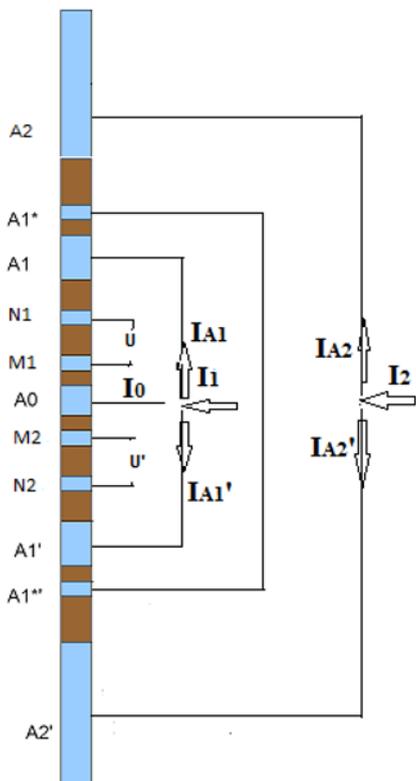
Fig. 2. Model scheme for the DLT-GA laterolog device (Halliburton 1992): A0, A1, A1′, A2, A2′ − current electrodes; A1*, A1*′ – monitoring electrodes; M1, M2, N1, N2 – measuring electrodes; U, U' – voltage between pair of electrodes M1 and N1, and between M2 and N2, respectively; I_0, I_1, I_{A1}, I_{A1}', I_2, I_{A2}, I_{A2}' − currents flowing through respective electrodes A0, A1, A1′, A2, A2′; A1, A1′ − short connected; A2, A2′ − short connected; A1*, A1*′ − short connected; M1, M2 – not connected; N1, N2 – not connected.

Conditions for the DLT-GA device according to Fig. 2

LLD	LLS
Currents I_1 and I_2 adjusted so that the measured electrode potentials satisfy the equation: $$(V_{N1} + V_{N2}) - (V_{M1} + V_{M2}) = 0$$ Current I_2 adjusted so that all potentials are equal: $$V_{A1} = V_{A2} = V_{A1'} = V_{A2'}$$ Current lines discharged from the A electrodes are closed to the return B electrode located at the surface or within the borehole (fish) so as to satisfy the equation: $$I_B = I_{A0} + I_{A1} + I_{A1'} + I_{A2} + I_{A2'}$$ Potential is measured by electrodes M1, N1, M1′, N1′: $$V^* = 1/4(V_{M1} + V_{M2} + V_{N1} + V_{N2})$$ Reference electrode N resides 80 ft beneath the torpedo.	Current I_1 adjusted so that the measured electrode potentials satisfy the equation: $$(V_{N1} + V_{N2}) - (V_{M1} + V_{M2}) = 0$$ Current $I_2 = I_{A0} + I_{A1} + I_{A1'}$ Potential is measured by electrodes M1, N1, M2, N2: $$V^* = 1/4(V_{M1} + V_{M2} + V_{N1} + V_{N2})$$ Reference electrode N resides 80 ft beneath the torpedo.

Potentials at measuring (detector) electrodes M1, M2, N1, N2 are a linear combination of potentials measured by normal devices of suitable length:

$$V_{M1} = \beta V_{A1M1}^0 + \gamma V_{A1'M1}^0 + \delta V_{A2M1}^0 + \lambda V_{A2'M1}^0 + V_{A0M1}^0 ,$$

$$V_{N1} = \beta V_{A1N1}^0 + \gamma V_{A1'N1}^0 + \delta V_{A2N1}^0 + \lambda V_{A2'N1}^0 + V_{A0N1}^0 ,$$

$$V_{M2} = \beta V_{A1M2}^0 + \gamma V_{A1'M2}^0 + \delta V_{A2M2}^0 + \lambda V_{A2'M2}^0 + V_{A0M2}^0 ,$$

(5)

$$V_{N2} = \beta V_{A1N2}^0 + \gamma V_{A1'N2}^0 + \delta V_{A2N2}^0 + \lambda V_{A2'N2}^0 + V_{A0N2}^0 ,$$

where $V_{A_iM1}^0$, $V_{A_iM2}^0$, $V_{A_iN1}^0$, $V_{A_iN2}^0$ are potentials measured by normal logs of length A_iM1, A_iM2, A_iN1, A_iN2, where i refers to each electrode in sequence ($i = 0, 1, 2$). Normal logs were supplied with I_0 current.

Multiplication factors, β, γ, δ, and λ (Eq. 6), represent ratios of currents I_{A1}, $I_{A1'}$, I_{A2}, and $I_{A2'}$ as discharged from electrodes A_1, $A_{1'}$, A_2, and $A_{2'}$ to current I_0 flowing from the central electrode A_0.

$$\beta = \frac{I_{A1}}{I_0}, \qquad \gamma = \frac{I_{A1'}}{I_0}, \qquad \delta = \frac{I_{A2}}{I_0}, \qquad \lambda = \frac{I_{A2'}}{I_0}, \qquad (6)$$

In the case of LLD mode laterolog in non-homogenous rock media, multiplication factors for currents I_1 and I_2 (Eq. 4) (Table 2) served to equalize potentials on measuring electrodes and were calculated using Eqs. 7-10:

$$\beta = \alpha g_{A1}, \quad \gamma = \alpha g_{A1'}, \quad \delta = \alpha w g_{A2}, \quad \lambda = \alpha w g_{A2'}, \qquad (7)$$

$$\alpha = \frac{I_1}{I_0} = \frac{\left(V^0_{A0N1} + V^0_{A0N2}\right) - \left(V^0_{A0M1} + V^0_{A0M2}\right)}{g_{A1}\left(V^0_{A1M1} + V^0_{A1M2}\right) + g_{A1'}\left(V^0_{A1'M1} + V^0_{A1M2}\right) + w\left[g_{A2}\left(V^0_{A2M1} + V^0_{A2M2}\right) + g_{A2'}\left(V^0_{A2'M1} + V^0_{A2'M2}\right)\right]}$$

$$\overline{-\left\{g_{A1}\left(V^0_{A1N1} + V^0_{A1N2}\right) + g_{A1'}\left(V^0_{A1'N1} + V^0_{A1'N2}\right) + w\left[g_{A2}\left(V^0_{A2N1} + V^0_{A2N2}\right) + g_{A2'}\left(V^0_{A2'N1} + V^0_{A2'N2}\right)\right]\right\}}$$
$$(8)$$

$$w = \frac{I_2}{I_1} = \frac{g_{A1}V^0_{A1}}{g_{A2}V^0_{A2}}, \qquad (9)$$

$$g_{A1} = \frac{I_{A1}}{I_1} = \frac{V^0_{A1'}}{V^0_{A1} + V^0_{A1'}}, \qquad g_{A2} = \frac{I_{A2}}{I_2} = \frac{V^0_{A2'}}{V^0_{A2} + V^0_{A2'}},$$

$$g_{A1'} = \frac{I_{A1'}}{I_1} = \frac{V^0_{A1}}{V^0_{A1} + V^0_{A1'}}, \qquad g_{A2'} = \frac{I_{A2'}}{I_2} = \frac{V^0_{A2}}{V^0_{A2} + V^0_{A2'}}. \qquad (10)$$

Relevant multiplication factors for LLS mode currents are defined by the Eqs. 11-12.

$$\beta = \alpha g_{A1}, \quad \gamma = \alpha g_{A1'}, \quad \delta = (1-\alpha)g_{A2}, \quad \lambda = (1-\alpha)g_{A2'}, \quad w = -1, \qquad (11)$$

$$\alpha = \frac{g_{A2}\left[\left(V^0_{A2M1} + V^0_{A2M2}\right) - \left(V^0_{A2N1} + V^0_{A2N2}\right)\right] + g_{A2'}\left[\left(V^0_{A2'M1} + V^0_{A2'M2}\right) - \left(V^0_{A2'N1} + V^0_{A2'N2}\right)\right] +}{g_{A1}\left(V^0_{A1M1} + V^0_{A1M2}\right) + g_{A1'}\left(V^0_{A1'M1} + V^0_{A1M2}\right) + w\left[g_{A2}\left(V^0_{A2M1} + V^0_{A2M2}\right) + g_{A2'}\left(V^0_{A2'M1} + V^0_{A2'M2}\right)\right] -}$$

$$\frac{\left(V^0_{A0N1} + V^0_{A0N2}\right) - \left(V^0_{A0M1} + V^0_{A0M2}\right)}{-\left\{g_{A1}\left(V^0_{A1N1} + V^0_{A1N2}\right) + g_{A1'}\left(V^0_{A1'N1} + V^0_{A1'N2}\right) + w\left[g_{A2}\left(V^0_{A2N1} + V^0_{A2N2}\right) + g_{A2'}\left(V^0_{A2'N1} + V^0_{A2'N2}\right)\right]\right\}}$$
$$(12)$$

Potential measured in LLD and LLS modes was calculated according to Eq. 13 while apparent resistivity was determined according to Eq. 14.

$$V^* = 1/4\left(V_{M1} + V_{M2} + V_{N1} + V_{N2}\right), \qquad (13)$$
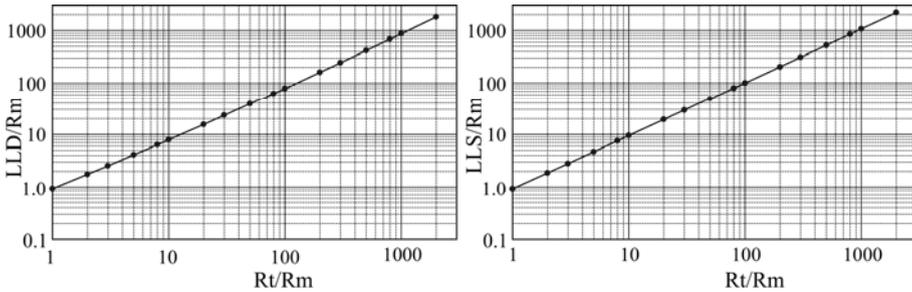
Fig. 3. Apparent resistivity *versus* true resistivity in a homogenous rock section sampled from a borehole of 0.20 m diameter with mud resistivity of $R_m = 1$ ohm m. Results on the left are from LLD mode, results on the right from LLS mode.

$$R_a = K \frac{V^*}{I_0} \; . \tag{14}$$

Geometric coefficient values were adopted from the DLT-GA device Technical Manual as $K^{LLD} = 0.81$ m and $K^{LLS} = 1.45$ m.

Test calculations were performed using a homogenous rock section sampled from a borehole of 0.2 m diameter with mud resistivity $R_m = 1$ ohm m. The tests gave rock formation resistivities of 1, 2, 3, 5, 8, 10, 20, 30, 50, 80, 100, 200, 300, 500, 800, 1000, and 2000 ohm m. Figure 3 shows the results of apparent resistivity calculations.

5.   ELTE METHOD FOR RESISTIVITY MODELING

The finite element software package, COMSOL Multiphysics v. 4.2a, was used to model the apparent resistivity measured in LLD mode for an inhomogeneous medium. The task was framed as a DC problem due to the low frequency of device measuring conditions, allowing the Laplace Eq. 1 to be solved numerically.

Figure 4 shows finite element implementation of the deep penetration function of the DLT-GA dual laterolog device. Guard electrodes (A_1, A_1^*, and A_2 in Fig. 2) above measuring electrodes were merged into an electrode array referred to as A_1, which was treated as a metal body (grey) with uniform potential. The surface of the device between A_1 and A_2 (Fig. 2) was treated as an insulating boundary condition (blue). The lower guard electrode was treated in a similar manner and referred to as A_2. Reference electrode N and the return electrode B were positioned 25 and 125 m, respectively, above the center of the device. Point and line current sources (red) served as respective sources for short (A_0 and B) and long (A_1 and A_2) current electrodes.
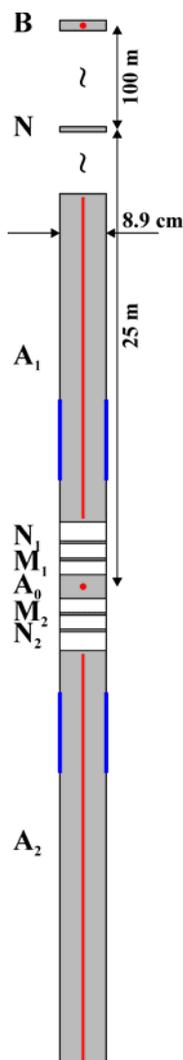
Fig. 4. Physical model (5 × horizontal enlargement) on which the deep penetration function of the DLT-GA dual laterolog is based. Grey color denotes metal apparatus assumed to have a resistivity of 10^{-5} ohm m, white is an insulator with resistivity of 10^{5} ohmm, blue indicates insulating boundary conditions, and red denotes current sources.

The superposition theorem was used to ensure the potential equivalence between measuring electrodes M_1 and N_1, as well as between M_2 and N_2 (Fig. 4). The electrode potentials caused by individual current electrodes A_0, A_1, A_2, and B were calculated as the middle, two guard and return current electrodes, respectively. The guard electrode current was determined as follows. Due to the laterolog principle, potentials on electrodes M_1 and N_1, as well as on M_2 and N_2, are set to be equal: $V_{M1} = V_{N1}$, $V_{M2} = V_{N2}$, where

$$V_{M1} = I_0 \left[\tilde{R}_{A0M1} + n_1 \tilde{R}_{A1M1} + n_2 \tilde{R}_{A2M1} - \left(1 + n_1 + n_2\right) \tilde{R}_{BM1} \right], \tag{15a}$$

$$V_{\text{N1}} = I_0 \left[\tilde{R}_{\text{A0N1}} + n_1 \tilde{R}_{\text{A1N1}} + n_2 \tilde{R}_{\text{A2N1}} - \left(1 + n_1 + n_2 \right) \tilde{R}_{\text{BN1}} \right], \tag{15b}$$

$$V_{\text{M2}} = I_0 \left[\tilde{R}_{\text{A0M2}} + n_1 \tilde{R}_{\text{A1M2}} + n_2 \tilde{R}_{\text{A2M2}} - \left(1 + n_1 + n_2 \right) \tilde{R}_{\text{BM2}} \right], \tag{15c}$$

$$V_{\text{N2}} = I_0 \left[\tilde{R}_{\text{A0N2}} + n_1 \tilde{R}_{\text{A1N2}} + n_2 \tilde{R}_{\text{A2N2}} - \left(1 + n_1 + n_2 \right) \tilde{R}_{\text{BN2}} \right], \tag{15d}$$

and the potential on the reference electrode due to the current sources is

$$V_{\text{N}} = I_0 \left[\tilde{R}_{\text{A0N}} + n_1 \tilde{R}_{\text{A1N}} + n_2 \tilde{R}_{\text{A2N}} - \left(1 + n_1 + n_2 \right) \tilde{R}_{\text{BN}} \right]. \tag{16}$$

In Eqs. 15a-d, I_0 denotes current discharged from the middle electrode (A_0), $\tilde{R}_{ij}$ is the transfer resistance between the current electrode i and the measuring electrode j (*i.e.*, the potential on electrode j as a consequence of the current flowing out from electrode i), n_1 and n_2 are factors representing the guard electrode currents relative to the middle electrode current ($n_1 = I_1/I_0$, $n_2 = I_2/I_0$, where I_1 and I_2 are the currents flowing from guard electrodes A_1 and A_2, respectively). From Eqs. 15a-d, the two unknown multipliers can be obtained for an arbitrary heterogeneous medium. Multipliers calculated by the methods above ensure a focused current field for A_0 and thus maximal penetration depth for the device.

A two-dimensional axial geometry was used to solve the problem. The finite element model was extensively validated by refinement of the mesh resolution, size of the model domain, and device performance at layer boundaries.

Numerical tests served to minimize the effects of mesh resolution on electrode potential. Global mesh size adjustments did not necessarily provide more exact results. Local adjustments around areas of the device that experience high current density can, however, significantly affect the accuracy of calculation (Drahos and Galsa 2007). Figure 5 shows local mesh refinements around the metal electrodes and the relative potential variation on electrodes as a function of the number of finite elements. The last three points in Fig. 5b are local manual refinements. The final mesh contained 366 250 finite elements with a maximum size of 1 mm along the surface of the device and 2 mm within the metal electrodes. The potentials on short current electrodes A_0 and B varied by almost 3% due to variation in mesh unit size. The geologic medium was assumed to be homogeneous with a resistivity of 1 ohm m. Only the middle and the return electrodes discharged current, with $I_0 = 1$ A and $I_B = -1$ A.

The geologic medium surrounding the device was modeled along the vertical symmetry axis of a cylinder of radius R and height H. The outer boundary of the cylinder was grounded. A smaller cylinder surrounding the device was introduced into the model to provide finer mesh around the

a

b
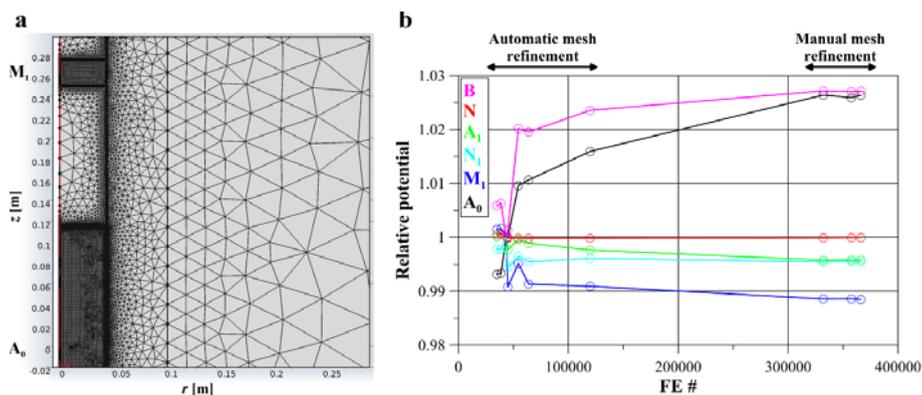
Fig. 5: (a) Finite element mesh surrounding the metal electrodes and within the borehole; (b) Effects of the number of finite elements on electrode potential in a homogeneous medium of $R_t = R_m = 1$ ohm m. Relative potential is compared to the normal mesh in COMSOL software.

a

b
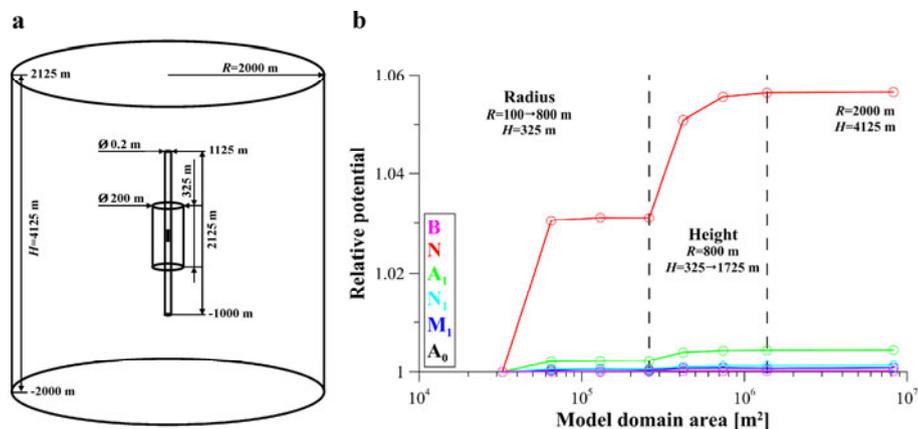
Fig. 6. Sketch of the model geometry (a) and relative potential variation of the electrodes plotted against the area of the model domain (b).

laterolog. Figure 6a shows the geometry of the model. The influence of different R and H values on electrode potential demonstrated effects of the model domain size. Figure 6b shows the corresponding variation in electrode potentials as a function of the model domain volume in three steps: (i) the radius of the domain was increased from 100 to 800 m, (ii) the height of the domain was increased vertically from 325 to 1725 m, and (iii) given no effect on potentials, the domain was enlarged by a factor of more than two, where $R = 2000$ m and $H = 4125$ m. The model assumed a borehole of 8″ diameter (20.3 cm) and 2125 m length centered around the electrode A_0

(Fig. 6b). The numerical model of this geometry gave 746 020 finite elements. Within elements, electric potentials were approximated by a quadratic function, giving 1.5 million total degrees of freedom to be solved. For a given depth, solutions required approximately 1 min CPU time and less than 4 GB memory on an Intel Server.

In practice, the apparent resistivity is calculated from the potential difference between the electrode M_1 and the reference electrode

$$R_a^{\text{LLD}} = K^{\text{LLD}} \frac{V_{\text{M1}} - V_{\text{N}}}{I_0} \,, \tag{17}$$

where the geometric factor of the device with a finite extension, K^{LLD}, is unknown. The geometry of the device (long and adjacent electrodes) precludes accurate pointwise approximation. The electric potential for electrodes M_1 and N in a homogeneous geologic medium with true resistivity of $R_t = R_m = 0.1$, 1, 10, and 100 ohm m at $I_0 = 1$ A was calculated to determine the geometric factor,

$$K^{\text{LLD}} = R_t \frac{I_0}{V_{\text{M1}} - V_{\text{N}}} \cong 0.90693 \,\text{m} \,. \tag{18}$$

This calculation yielded a geometric coefficient, $K^{\text{LLD}} = 0.90693$ m, that recovered the true resistivity values to within a standard deviation of less than 0.3%.

Numerical modeling was also used to investigate the effects of borehole mud on apparent resistivity. The diameter of the borehole and the mud resistivity was fixed ($d = 0.20$ m and $R_m = 1$ ohm m), while the true resistivity varied from $R_t = 0.1$ to $R_t = 1000$ ohm m. Figure 7 shows potential field and current line structure for three different cases. At low R_t, elevated potential was localized near current electrodes. As R_t increased, areas of elevated
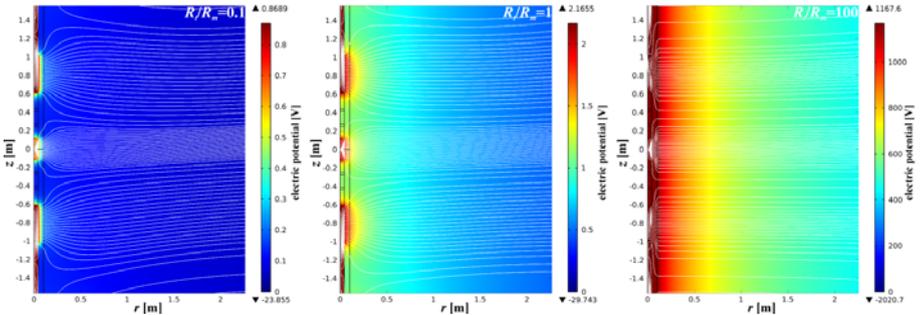
Fig. 7. The electric potential field (colored) and current line distribution (white lines) around middle and guard electrodes at different ratios for true resistivity and mud resistivity ($R_t/R_m = 0.1$; 1; 100).
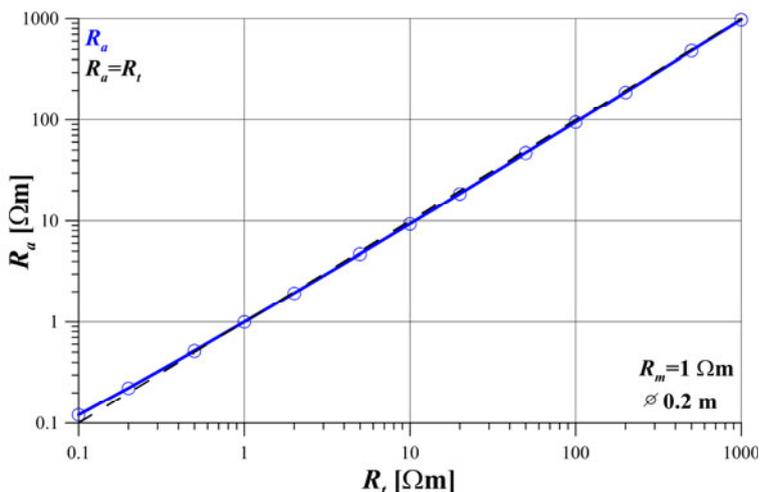
Fig. 8. Relation between the apparent and true resistivity in the presence of a borehole mud having $R_m = 1$ ohm m; borehole diameter $d = 0.20$ m.

potential became more elliptical, encircling the whole device. Current lines become more perpendicular to the borehole due to focusing which ensures the deep penetration of the LLD device. Figure 8 demonstrates that the device focused in LLD mode was basically insensitive to mud in the borehole. The apparent resistivity approximates true resistivity, with small deviation observed only at very low R_t values.

Further testing demonstrated device performance at horizontal layer boundaries. Figure 9a displays the inhomogeneous current density and current line distribution when the true resistivity of the upper layer is larger by 2 orders of magnitude and the boundary is at the middle of A_0. The modeled log (Fig. 9b) confirms that the apparent resistivity approximates true resistivity with some distance from the boundary (in the presence of mud, R_a is lower than R_t). Global overshoot occurs just above the boundary in the absence of a borehole while a local overshoot occurs in the case of a mud filled borehole. The Groningen effect is observable as an apparent resistivity increase beneath the boundary when the reference potential electrode, N, is positioned at the resistive layer (the distance between A_0 and N is 25 m). It is worth noting that the two curves (blue and red) converge in the lower layer where $R_m = R_t = 1$ ohm m.

A final set of numerical tests investigated the effects of a thin, high resistivity layer on R_a (Fig. 10). The layer was assigned a thickness of 12 m (about half of the distance separating A_0 and N electrodes) and a resistivity exceeding the shoulder bed resistivity by two orders of magnitude. Apparent
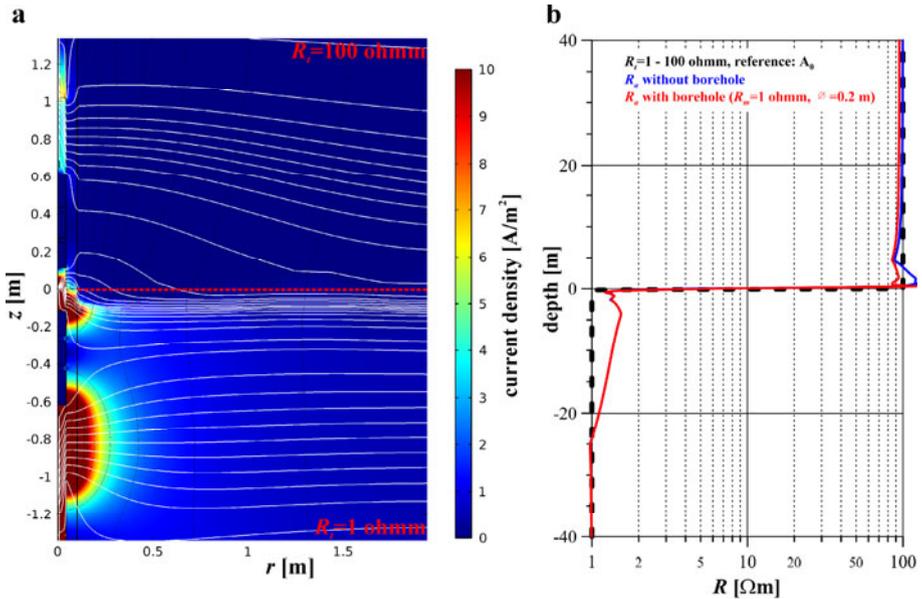
Fig. 9: (a) Current density and current line distribution in the vicinity of the electrodes when A_0 is opposite to the layer boundary separating the layers with high (up) and low (down) resistivity; (b) Relation of the apparent and true resistivity logs across the layer boundary. Red line denotes the apparent resistivity log with borehole, blue line without borehole, and black dashed line gives the true resistivity.

resistivity maxima appear above and below the bed. The stratigraphically lower maximum is of greater magnitude due to the Groningen effect. Within the bedded layer, R_a is lower than R_t due to effect of the thin layer. The apparent maxima in the presence of low resistivity mud are localized at the boundaries.

The apparent resistivity curve calculated using the Coulomb's charges method, assuming a 0.2 m diameter borehole and 1 ohm m mud resistivity, was compared with results calculated by the ELTE method (Fig. 10). The curves are visibly similar but theoretical resistivity and obtained apparent resistivity also differ in several important ways. Generally, ELTE results obtained with borehole are closer to the theoretical resistivity model. Differences can be attributed to the different physical approaches used by each method, errors related to approximations and calculation methods and model uncertainties. Differences between theoretical and calculated values become distinctly visible at points where there are large shifts in resistivity (Fig. 10). Calibration curves calculated by the Coulomb's charges method cause diverging results for apparent and true resistivity (Cichy and Ossowski
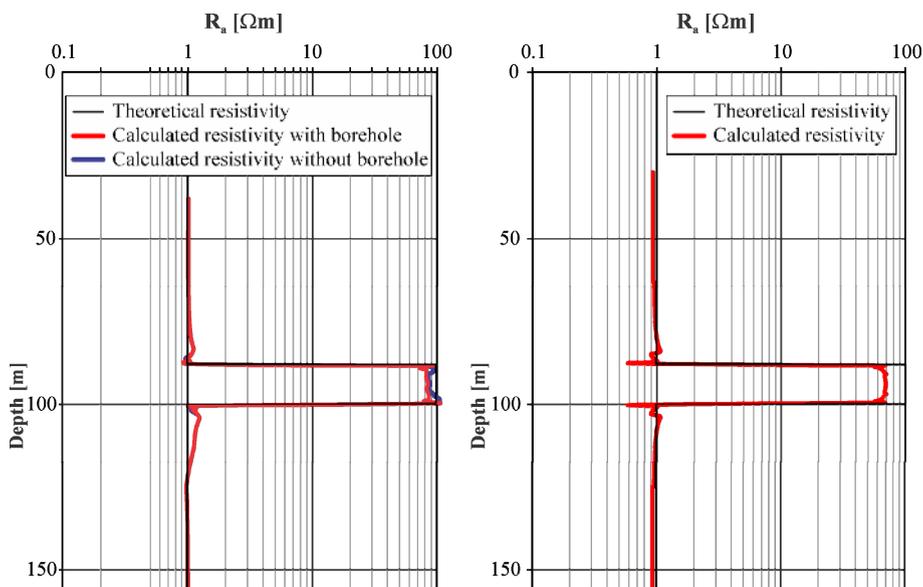
Fig. 10. Comparison of apparent and true resistivity logs across a thin, resistive layer. The left panel shows ELTE model results for apparent resistivity with the borehole in red and without the borehole in blue, and the true resistivity of the strata shown by the black line. The right panel shows Coulomb's charges method results for the apparent resistivity log with borehole in red and the true resistivity of the strata in black.

2015) (Fig. 3). The lower K^{LLD} value of 0.81 m (relative to $K^{LLD} = 0.90693$ m from ELTE method) gave lower apparent resistivity for Coulomb's charges method results. For a bed having a true resistivity of 100 ohm m, the Coulomb's charges method gave 78.53 ohm m. The apparent resistivity curve (Fig. 10 right) is perfectly symmetric and did not exhibit Groningen effects because of how the methods summed potential generated by normal device components.

6. COMPARISON OF METHODS USING GEOLOGIC DATA

Table 2 describes a stratigraphic section typical of the Fore-Sudetic Monocline in southwestern Poland, constructed from geological data and well logs.

The profile was used to compare results obtained from the Coulomb's charges and ELTE methods. A key unit in the profile was the Main Dolomite (cyclothem Z2, Stassfurt), which is a porous, hydrocarbon- or water-saturated carbonate overlain (in succession) by the Basal Anhydrite, Older Salt, and Main Anhydrite (Table 2). The thickness of the Main Dolomite car-

Table 2

Geological profile typical of the Fore-Sudetic Monocline, southwestern Poland

Strata I.D.	Top [m]	Bottom [m]	Thickness [m]	Lithology	Age and/or stratigraphic name	Resistivity [ohm m]
1	0	978	978		Quaternary, Tertiary, Jurassic	
2	978	1966	988	claystone, sandstone, mudstone	Upper Triassic	
3	1966	2237	271	limestone, dolomitic limestone, marly claystone, dolomitic claystone	Middle Triassic, Muschelkalk	150
4	2237	2349	112	claystone, limestone, dolomitic limestone, marl	Upper Bunter Sandstone, Red	100
5	2349	2576	227	claystone, mudstone, sandstone	Middle Bunter Sandstone	60-80[#]
6	2576	2905	329	claystone, lime-claystone, dolomitic claystone, sandstone, mudstone	Lower Bunter Sandstone	100
7	2905	2932	27	claystone	Transient Claystones	30
8	2932	2980	48	halite	The Youngest Salt	2000
9	2980	2994	14	anhydrite and clay	Pegmatite Anhydrite (1.5 m), Red Clay	20
10	2994	3051	57	halite	Younger Salt	2000
11	3051	3059	8*	anhydrite, gray clay, anhydrite	Main Anhydrite, Gray, Salt Clay, Covering Anhydrite	2000
12	3059	3297	238	halite	Older Salt	2000
13	3297	3303	6	anhydrite	Basal Anhydrite	2000
14	3303	3307	4*	dolomite (up to 60 m thick)	Main Dolomite, Z2, Stassfurt	40-150[#]
15	3307	3349	42	anhydrite	Upper Anhydrite	2000
16	3349	3370	21	halite	The Oldest Salt	2000
17	3370	3518	148	anhydrite	Lower Anhydrite	2000
18	3518	3519.9	1*	limestone (up to 80 m thick)	Zechstein Limestone	40-150[#]
19	3519	3520	1	shale	Copper Shale	5
20	3520	3652	132	sandstone	Rotliegend	10-30

*[)] thickness of unit may differ in different wells,
#[)] range of values observed for the unit.
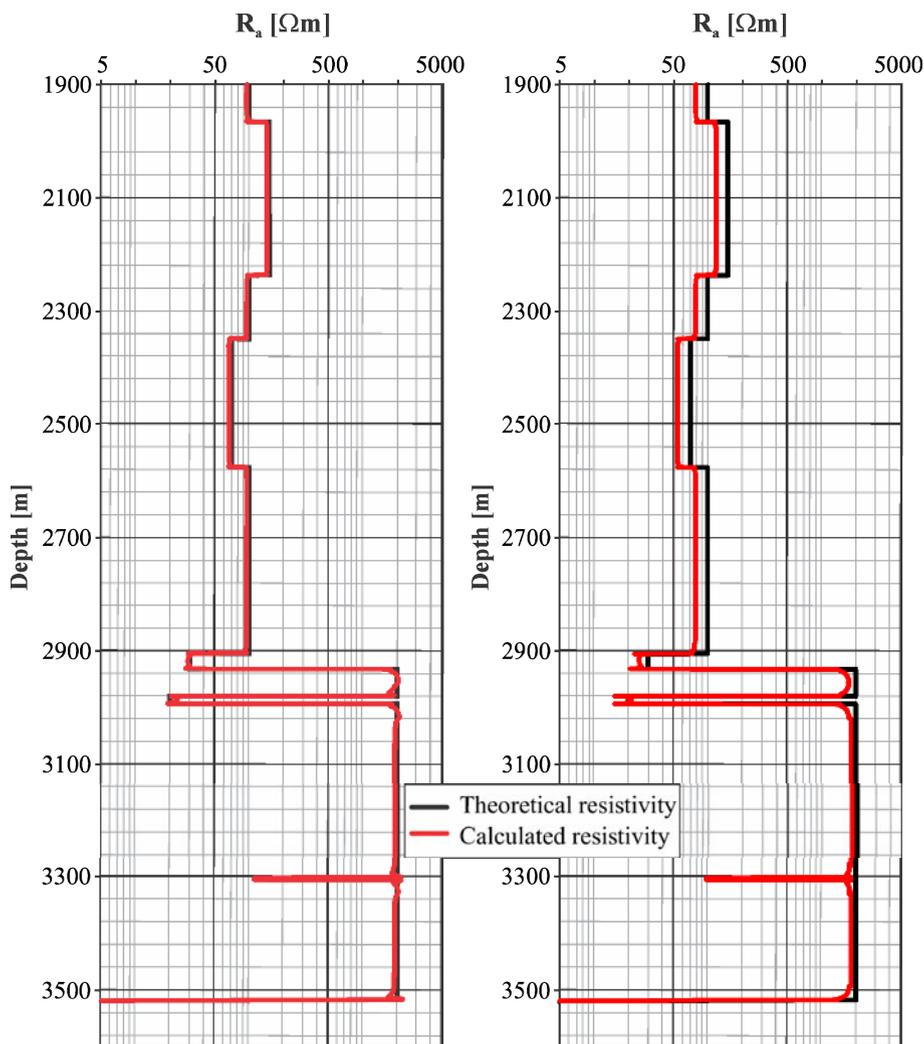
Fig. 11. Calculations and modeling results for apparent resistivity of a constructed geologic profile. Theoretical resistivity model is shown in black and calculated resistivity is shown in red. The left panel shows ELTE model results and the right panel shows Coulomb's charges results.

bonate varied, as did the thick, overlying accumulation of the Basal Anhydrite, Older Salt, and Main Anhydrite. This succession of strata was key to our analysis due to the challenges it posed for accurate resistivity recording.

The calculated and modeled results are visibly similar (Fig. 11). A lower assumed K^{LLD} caused lower apparent resistivity results for the Coulomb's

charges method. A bed having a true resistivity of 2000 ohm m gave a calculated apparent resistivity of 1811 ohm m.

7. FURTHER APPLICATION OF MODELING RESULTS

Comparison of the apparent resistivity curves demonstrated that both methods provide similar results and may be used as a numerical tool for determining and interpreting apparent dual laterolog resistivity. Calculated and theoretical resistivity values were obtained by the Coulomb's charges method using the assumed resistivity values R_t, R_{xo}, R_m and geometrical parameters (D/d) shown in Table 3.

Table 3

Input values for theoretical resistivity calculations

D/d	R_{xo}/R_m	R_t/R_m	D/d	R_{xo}/R_m	R_t/R_m	R_{xo}/R_m	R_t/R_m
1	2	3	4	5	6	7	8
1.5	1	1	12	10	10	250	250
2	1.25	1.25	14	15	15	300	300
2.5	1.5	1.5	16	20	20	400	400
3	1.75	1.75	24	30	30	500	500
4	2	2	32	40	40	1000	1000
5	2.5	2.5		50	50	2000	2000
6	3	3		75	75	3000	3000
7	4	4		100	100	5000	5000
8	5	5		150	150	7500	7500
10	7.5	7.5		200	200	10000	10000

Explanations: R_t – true resistivity of unit, R_m −mud resistivity, R_{xo} – invaded zone resistivity, d and D – respective radii for borehole and invaded zone; columns 7 and 8 represent continuations of columns 5 and 6, respectively.

The results, along with analysis of discrepancies and uncertainty, suggest a processing methodology for dual laterolog apparent resistivity (Fig. 12). The method determines the true resistivity of the virgin and invaded zones together with invaded zone diameter, D. The proposed methodology implements corrections in a manner similar to methods used by Halliburton and Schlumberger (www.halliburton.com; www.slb.com). It consists of two linear corrections in sequence, (i) for borehole influence, and (ii) influence of the interpreted bed thickness and shoulder bed resistivity. A third correction, (iii) removes filtration effects according to tornado charts. Corrections may be implemented in an automatic or semiautomatic fashion (in the case of thickness and shoulder bed resistivity).
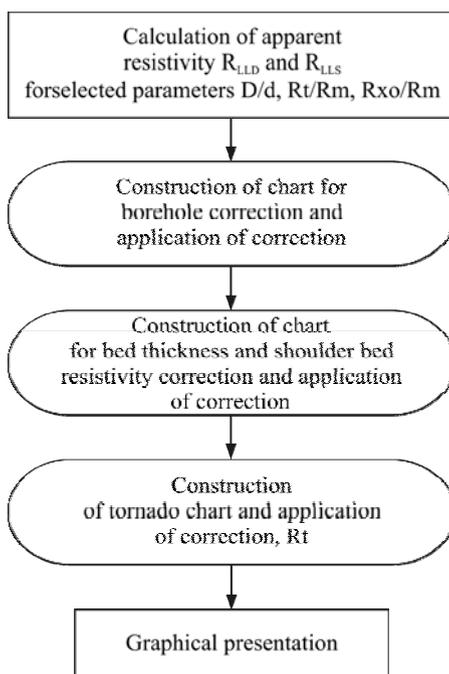
Fig. 12. Flow chart of apparent resistivity correction procedure. Dual laterolog apparent resistivity is recorded and interpreted to determine the true resistivity of the virgin zone.

A calculation based on the Coulomb's charges method was used to compute apparent resistivity for various resistivity conditions and geometric models (Table 3). The data were also used to construct nomograms for comparison with established reference charts for apparent resistivity borehole corrections used by Halliburton (Fig. 13). Values in Fig. 13 show that the correction procedures described here give the same results for borehole corrected resistivity as those given by Halliburton reference charts.

Charts for bed thickness and shoulder bed resistivity corrections were also constructed using the dataset in Table 3 (Fig. 14). Similar modeling of efficient corrections for layer thickness and shoulder bed resistivity was performed for well logging in horizontal wells (Shaogui *et al.* 2009).

A third type of correction for filtration effects in the reservoir rock (*i.e.*, tornado charts) was also applied using the dataset in Table 3. Equation 19 was used to select suitable data and construct tornado charts:

$$R_t / R_{\text{LLD}} = f \left[\log \left(\frac{R_{\text{LLD}}}{R_{\text{LLS}}}; \frac{R_{\text{LLD}}}{R_{R_{xo}}} \right) \right]. \tag{19}$$

LLD

LLS
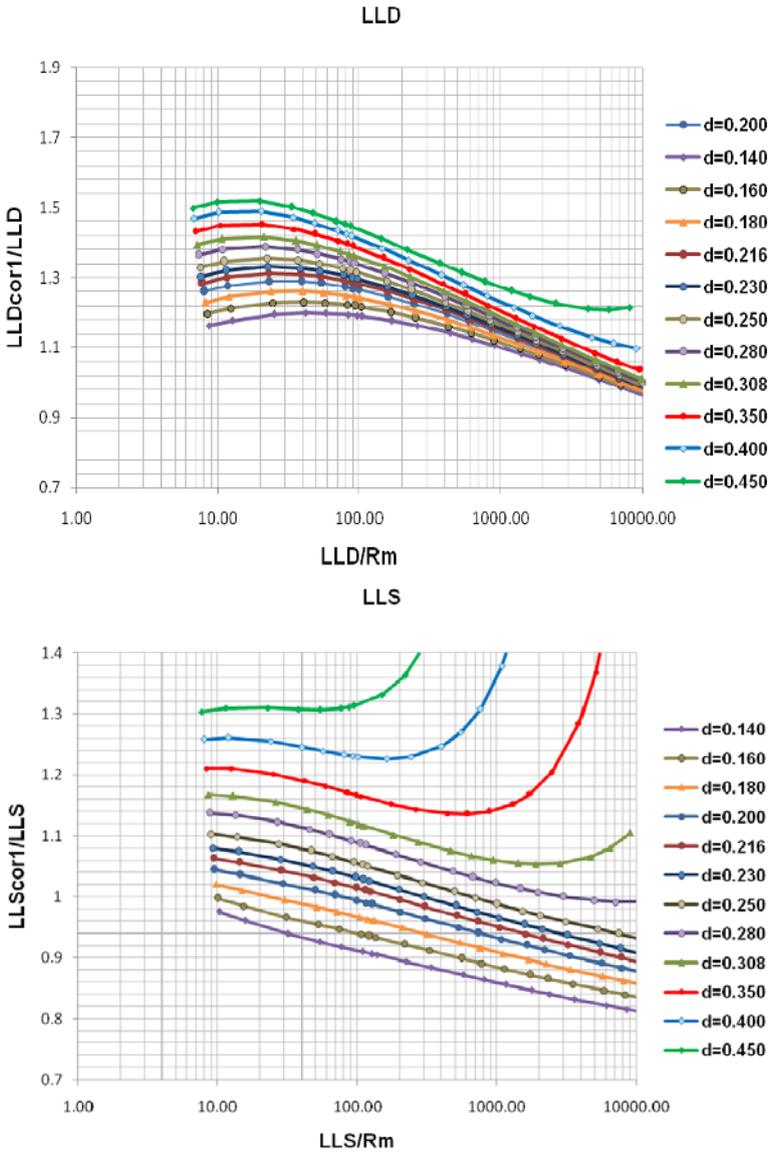
Fig. 13. Borehole correction charts for LLD (upper panel) and LLS (lower panel) re-sistivity.

Figure 15 gives an example of a tornado chart based on an invaded zone and mud resistivity ratio of $R_{xo}/R_m = 100$ ($R_m = 1$ ohm m). This apparent re-sistivity correction procedure uses two sets of curves with parameters R_t/R_m and D/d to determine true resistivity, R_t. The initial point of the tornado
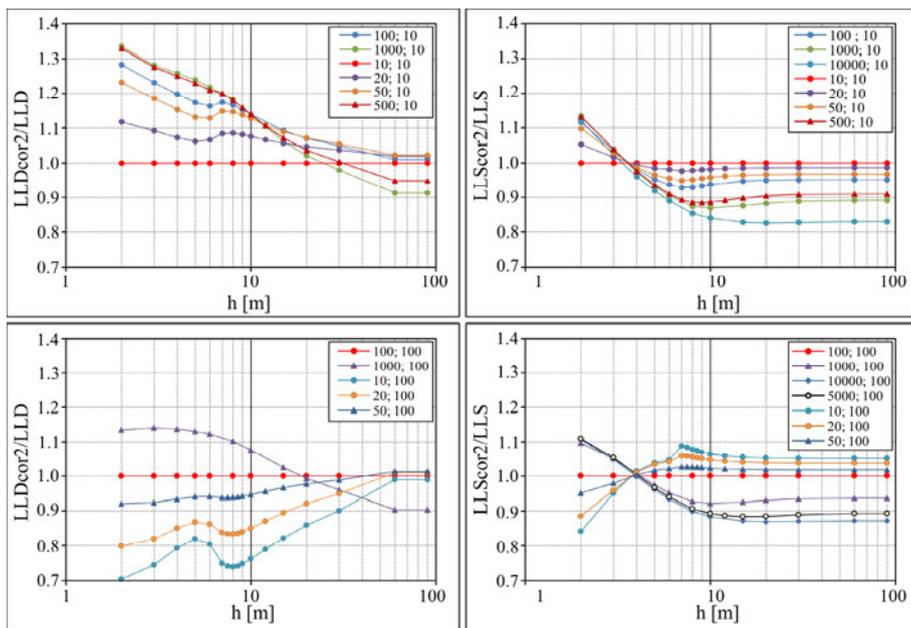
Fig. 14. Example of thickness and shoulder bed resistivity correction charts for LLD (left panel) and LLS (right panel) resistivity; the first number in legends means R_t, the second R_s.
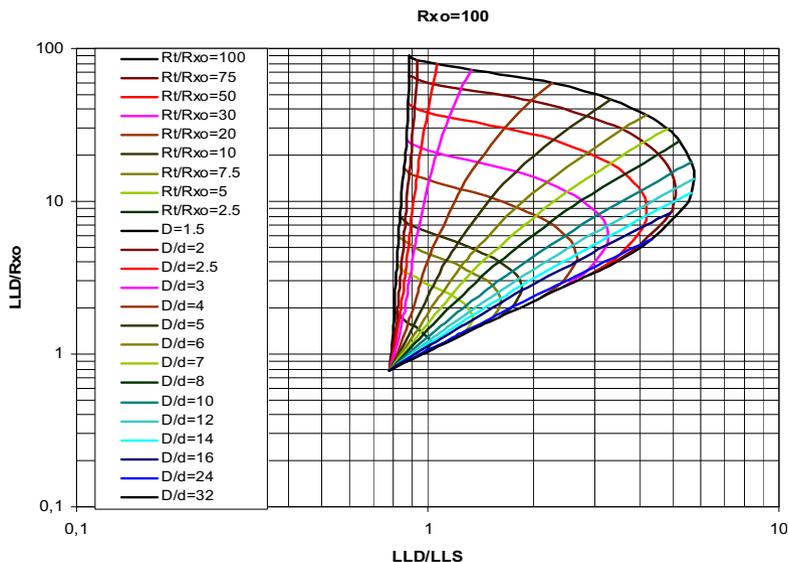
Fig. 15. Example of a tornado chart: R_t − true resistivity, R_{xo} − invaded zone resistivity, LLD and LLS refer to deep and shallow laterolog operating modes, d − borehole diameter, D − invaded zone diameter.
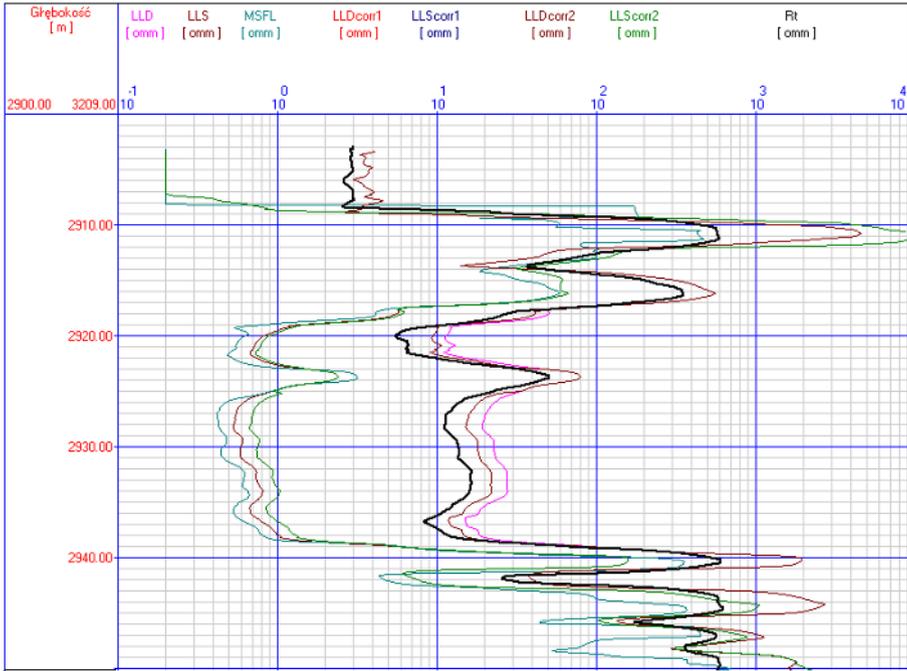
Fig. 16. Resistivity correction procedure applied to actual geological data according flow chart (Fig. 12; W5 borehole, Main Dolomite section of the Fore-Sudetic Monocline). Primary apparent resistivity is shown by LLD (pink), LLS (brown), and MSFL (forest green). Resistivity after borehole correction is shown by LLDcorr1 (red). Resistivity after layer thickness and shoulder bed correction is shown by LLDcorr2 (brown) and LLScorr2 (green). All corrections were applied to the true resistivity R_t (black) including tornado chart adjustments.

curves shifts visibly due to a discrepancy in the calculated apparent resistivity and theoretical resistivity of the model (Cichy and Ossowski 2015). Coordinates of this point are known from calculated results and are used in the final step of the correction procedure.

The apparent resitivity correction procedure gave results that, like true resistivity, were less than apparent resistivity values. Figure 16 demonstrates the influence of applied corrections. This example was based on well log data from well W5 located in the Fore-Sudetic Monocline of southwestern Poland. This part of the stratigraphic section includes the Main Dolomite (2912.0-2964.0 m), which shows significant variation in resistivity. The upper part of the Main Dolomite (beneath the Basal Anhydrite, up to 2918.5 m) is highly resistive, the middle part exhibits a relatively low resistivity, and the bottom part is highly resistive. The casing shoe is located at 2908.0 m depth. The upper part of borehole was cased with steel pipes of 7″ diameter.

Processing steps addressed apparent resistivity of LLD and LLS operating modes of the devices. The first step applied a semiautomatic correction for influence of bed thickness and shoulder bed resistivity (borehole correction performed by a service company). The invasion effect was then reduced by the tornado chart. All corrections were performed using software dedicated to the proposed solution and according to flow chart sequence (Fig. 12). The corrections generated lower resistivity values (Fig. 16).

8. CONCLUSIONS

This paper adds two methods to the body of procedures used to correct laterolog resistivity. Both methods provide comparable results using modern numerical tools to calculate apparent resistivity in complex geologic media and geometric situations. The methodology derives true resistivity from apparent resistivity corrections of DLT-GA laterolog data and provides a flexible approach that can also be adapted to specific cases.

R e f e r e n c e s

Alpin, L.M. (1964), Some theoretical method for solving direct resistivity logging problem, *Izv. AN SSSR, Geof.*, 2.

Alpin, L.M., D.S. Dajev, and A.D. Karinskij (1985), *The Theory of Alternating Fields for Exploratory Geophysics*, Niedra, Moskva.

Anderson, B.I. (2001), Modeling and inversion methods for the interpretation of resistivity logging tool response, Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands.

Chew, W.C., Z. Nie, Q.-H. Liu, and B. Anderson (1991), An efficient solution for the response of electrical well logging tools in a complex environment, *IEEE Trans. Geosci. Remote Sens.* **29**, 2, 308-313, DOI: 10.1109/36.73673.

Cichy, A., and A. Ossowski (2015), Modeling electrical field distribution in layered geological rock formations with a borehole using the Coulomb charges method, *Acta Geophys.* **63**, 5, 1244-1255, DOI: 10.2478/s11600-014-0253-2.

Davydycheva, S., V. Druskin, and T. Habashy (1996), Solution of Maxwell's equation in an arbitrary 3-D inhomogeneous anisotropic media with sharp discontinuities using a finite-difference scheme on a regular Cartesian grid with material averaging, Tech. Report EMG-002-96-15, Schlumberger–Doll Research Center, Cambridge, USA.

Drahos, D. (1984), Electrical modeling of the inhomogeneous invaded zone, *Geophysics* **49**, 10, 1580-1585, DOI: 10.1190/1.1441566.

Drahos, D., and A. Galsa (2007), Finite element modeling of penetration electric sonde, *Magy. Geofiz.* **48**, 1, 22-30 (in Hungarian).

Ellis, D.V., and J.M. Singer (2007), *Well Logging for Earth Scientists*, 2nd ed., Springer, Dordrecht.

Halliburton (1992), Technical manual of dual laterolog DLT-GA device, Halliburton Co., Houston, USA, http://www.halliburton.com.

Lovell, J.R. (1993), Finite element methods in resistivity logging, Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands.

Moran, J., and J. Timmons (1957), The mathematical theory of ring electrode laterolog response in thick-beds, Project RR-602, Schlumberger–Doll Research Center, Cambridge, USA.

Moskow, S.,V. Druskin, T. Habashy, P. Lee, and S. Davydycheva (1999), A finite difference scheme for elliptic equations with rough coefficients using a Cartesian grid nonconforming to interfaces, *SIAM J. Numer. Anal.* **36**, 2, 442-464, DOI: 10.1137/S0036142997318541.

Nam, M.J., D. Pardo, and C. Torres-Verdín (2010), Assessment of Delaware and Groningen effects on dual-laterolog measurements with a self-adaptive hp finite-element method, *Geophysics* **75**, 6, F143-F149, DOI: 10.1190/1.3496670.

Pardo, D., C. Torres-Verdin, and L. Demkowicz (2005), Self-adaptive goal-oriented hp-finite element simulations of induction and laterolog measurements in the presence of steel casing. **In:** *Proc. Fifth Annual Meeting "Joint Industry Research Consortium on Formation Evaluation", 1 August 2005, The University of Texas, Austin, USA.*

Ribeiro, M., and A. Carrasquilla (2013), New approach to calculate the mud invasion in reservoirs using well logs. **In:** *Proc. 13th Int. Congr. Brazilian Geophysical Society and EXPOGEF, 26-29 August 2013, Rio de Janeiro, Brazil.*

Shaogui, D., L. Zhuqiang, and L. Zhiqiang (2009), Response of dual laterolog and fast correction for layer thickness and shoulder bed in horizontal wells, *Petrol. Explor. Develop.* **36**, 6, 725-729, DOI: 10.1016/S1876-3804(10)60005-5.

Trouiller, J.C., and I. Dubourg (1994), A better deep laterolog compensated for Groningen and reference effects. **In:** *SPWLA 35th Annual Logging Symposium, 19-22 June 1994, Tulsa, USA*, SPWLA-1994-VV.

Wang, H.M., L.C. Shen, and G.J., Zhang (1998), Dual laterolog response in 3-D environments. **In:** *SPWLA 39th Annual Logging Symposium, 26-28 May 1998, Keystone, USA*, SPWLA-1998-X.

ELRIS2D: A MATLAB Package
for the 2D Inversion of DC Resistivity/IP Data

Irfan AKCA

Ankara University, Faculty of Engineering,
Department of Geophysical Engineering, Ankara, Turkey;
e-mail: iakca@eng.ankara.edu.tr

A b s t r a c t

ELRIS2D is an open source code written in MATLAB for the two-dimensional inversion of direct current resistivity (DCR) and time domain induced polarization (IP) data. The user interface of the program is designed for functionality and ease of use. All available settings of the program can be reached from the main window. The subsurface is discretized using a hybrid mesh generated by the combination of structured and unstructured meshes, which reduces the computational cost of the whole inversion procedure. The inversion routine is based on the smoothness constrained least squares method. In order to verify the program, responses of two test models and field data sets were inverted. The models inverted from the synthetic data sets are consistent with the original test models in both DC resistivity and IP cases. A field data set acquired in an archaeological site is also used for the verification of outcomes of the program in comparison with the excavation results.

Key words: 2D inversion, GUI, finite elements, hybrid mesh.

1. INTRODUCTION

Direct current resistivity (DCR) method is one of the most frequently used geophysical tools for the solution of near surface problems. Impressive developments occur in the multichannel resistivity measurement technologies,

thus enabling thousands of readings in a few hours. This ability leads to explore a survey site in more details, increasing the volume of the data to be processed. An exhaustive review about the recent developments in DCR imaging may be found in the paper by Loke *et al.* (2013) where some applications for a variety of survey objectives are included. Although three-dimensional (3D) data inversion and interpretation techniques are available (*e.g.*, Pidlisecky *et al.* 2007), two-dimensional interpretation of DCR and IP data is still a valuable option. Two-dimensional inversion of DCR data is an extensively-studied research area which produced several commercial and academic inversion codes and software (*i.e.*, Res2DInv 2014, EarthImager 2009, DC2DInvRes 2014). These packages have their own originalities with plenty of customization options and routines for the inversion and visualization of DCR/IP data. In addition to the above-mentioned packages, there are several open source codes (*i.e.*, Karaoulis *et al.* 2013, Pidlisecky and Knight 2008) developed for different purposes.

This paper presents a new open source two-dimensional inversion program named ELRIS2D. The novelty of the program arose from the discretization routine involving the hybrid use of structured and unstructured meshes and the design of the user interface. ELRIS2D provides a simple and convenient user interface designed for the ease of use to invert and visualize DCR/IP data which is written in MATLAB. All available options can be reached from the main window by clicking on the corresponding user interface element. This approach in the design of user interface speeds up the process of multiple data sets. A hybrid mesh application for the finite-element method reduces the computational requirements and consumed CPU time. ELRIS2D uses the smoothness constrained least squares method (Loke and Barker 1996, Tsourlos *et al.* 1998) for the definition of objective function to be minimized. The parameter correction vector is calculated by solving the linear system of equations via the MATLAB built-in function "mldivide". A variety of techniques have been suggested for the incorporation of a smoothing operator to stabilize the inversion process and extract structural information from the geolectrical models (see Zhou *et al.* 2014, Akca and Başokur 2010, Başokur and Akca 2011). A five point Laplacian operator is used as the smoothness constraint in ELRIS2D for simplicity. The user interface provides some options and tools to enhance the view of the data and model sections. Outputs of the program may be exported as a snapshot, pdf or text files. ELRIS2D was tested with the responses of synthetic models and a field data set measured at the ancient city of Pisidian Antioch, in western Turkey. Test results show that the new open source code ELRIS2D is a powerful tool for the detection of subsurface targets and additional functionalities can easily be added by other researchers.

2. DESIGN AND IMPLEMANTATION

2.1 Mesh generation

Two-dimensional finite element meshes are usually constructed using triangular elements that provide more flexibility for the representation of known interfaces among subsurface units. Most commonly used finite element meshes are classified as "structured" and "unstructured" according to the placement and shape of the elements. Figure 1a, b illustrates structured and unstructured triangular mesh examples used in 2D modeling and inversion of DCR data. Structured mesh generation is based on dividing the rectangular (or tetragonal) elements into two or four triangles (Fig. 1b), whilst unstructured meshes are generated following the Delaunay triangulation rules constrained by the triangle quality (Fig. 1a). Shewchuk (1997) and Si (2008) gave comprehensive description of unstructured mesh design in two and three dimensions, and Rücker (2011) reviews the usage of unstructured meshes in DCR modeling and inversion. Among many others, the main advantage of using an unstructured mesh is reducing the size of the mesh and thus governing matrices used in forward calculations. Additionally, it is a flexible tool to outline irregular interfaces precisely while it permits the use of greater sized triangles in parts of the model mesh where the data sensitivity is low. The unstructured mesh given in Fig. 1a consists of 363 nodes which is nearly half the number of nodes of the structured mesh designed for the discretization of the same model. The number of triangles forming the structured mesh is nearly three times greater than that of the unstructured mesh.

Besides the abovementioned advantages of the unstructured meshes, some problems arise when they are used to parameterize an unknown subsurface model. For example, the arrangement of the triangles is somewhat irregular, which may cause flower or star shaped bodies to appear in the inverted model sections. This case gives rise to difficulties in the interpretation of the resistivity sections. For a better understanding of model and finite element mesh construction, it will be helpful to remember the following basic aspects regarding the discretization of subsurface.

Discretization of the subsurface is essential to the fulfillment of two requirements:

(1) A conceptual geophysical model is required to present the subsurface distribution of a physical parameter with certain geometry;

(2) Finite elements solution requires a bounded and a fine discretized domain.

The fulfillment of the two requirements listed above with totally unstructured or structured meshes causes the unnecessary growth of the mesh size (Fig. 1a, b). Therefore, some researchers distinguish parameter and calculation meshes. For example, Günther (2004) and Rücker (2011) suggest a triple
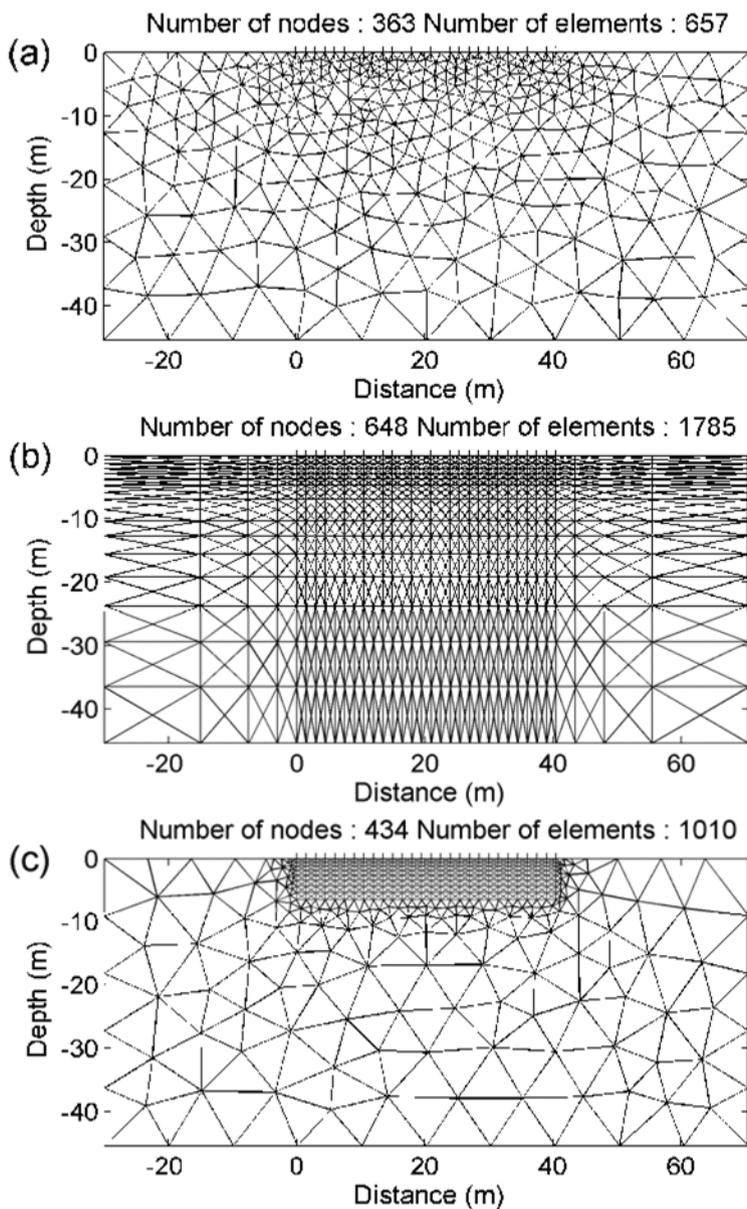
Fig. 1: (a) Unstructured, (b) structured, and (c) hybrid finite element meshes.

grid technique where three different unstructured meshes were used to "combine a resolution-dependent model parameterization with accurate forward calculations using finite elements".

In the present paper, a combination of regular and unstructured finite element meshes is used to employ the advantages of both types of meshes. The model mesh is divided into two subdomains. The first subdomain bounds the model region where measured data are sufficiently sensitive to solve the model parameters with high fidelity. The model region is bounded by the first and last electrode positions in *x*-direction and with maximum depth of investigation in *z*-direction. The maximum depth of penetration is calculated according to Edwards (1977). This portion of the mesh is constructed by dividing regularly arranged rectangles into triangular elements and each rectangular cell is treated as a model parameter. The remaining part of the mesh is required for the numerical implementation of finite elements algorithm and designed by using triangular elements with enlarging size towards the outer boundaries of the mesh (Fig. 1c). This leads to the reduction of the mesh size compared to a structured mesh keeping the conventional arrangement of the rectangular model cells. An example of the mentioned approach may also be seen in the tutorial by Günther and Rücker (2015). The unstructured portion of the mesh is produced by the program developed by Shewchuk (1997).

2.2 Forward modeling

The following elliptic equation is solved to calculate the potential distribution $\phi(x, y, z)$ over a 2D conductivity structure (σ) due to a current source (I):

$$-\nabla\big[\sigma(x, z)\nabla\phi(x, y, z)\big] = I(x, y, z) \ . \tag{1}$$

Equation 1 governs the three-dimensional potential distribution over a model where the conductivities differ only in *x*- and *z*-directions. The numerical solution of Eq. 1 may be achieved by the use of a numerical approach such as finite difference (Mufti 1976, Dey and Morrison 1979) or finite elements (Coggon 1971, Rijo 1977, Pelton *et al.* 1978). Solution of Eq. 1 is discussed by a plenty of researchers and well understood. Therefore, the paper will focus on the subsurface discretization and practical aspects of the application of finite elements method used in the program.

Application of finite element method to Eq. 1 applying Neumann and Dirichlet boundary conditions at the boundaries of a discretized domain results in a linear system of equations expressed as:

$$\mathbf{K} \cdot \mathbf{v} = \mathbf{S} \ , \tag{2}$$

where **K** is a positive definite, symmetric band matrix, **v** is a column vector comprising the unknown potentials at each mesh node, and **S** is the source term. Equation 2 can be solved by a matrix division in MATLAB as: $v = K \backslash S$ using Gaussian elimination. The matrix division operator in MATLAB al-

lows the solution of potentials due to all possible single current sources at once. Therefore, each column of **S** includes one nonzero element corresponding to the strength of the virtual current source which is set to $2\pi A$. Other elements of the source matrix are set as 0. Although the numerical calculations are based on conductivity and potential distribution, in practice the potential difference and apparent resistivity quantities are used. Therefore, the potential differences should be converted to apparent resistivities based on the layout of electrodes. Marescot *et al.* (2006) gave a comprehensive description of approaches for the definition of apparent resistivities. The definition called "general approach" is given by:

$$\rho_a = \frac{\Delta V}{\Delta V_0} \rho_0 \ , \tag{3}$$

where ΔV_0 and ΔV are respective potential differences for a homogeneous half-space with resistivity ρ_0 and a model with arbitrary resistivity distribution. ρ_0 may be assigned as 1 ohm-m for simplicity. This yields in a normalized potential difference with the response of system to the unity. The potentials at each node are calculated by assuming each electrode acts as a current pole. Therefore, any combination of current-potential electrodes may be used to calculate the apparent resistivity values. The apparent resistivity values are demonstrated as pseudosections following the basis given by Edwards (1977).

In the IP forward modeling an additional parameter, namely chargeability (η), is introduced to describe the polarization of subsurface (Seigel 1959). Chargeability is a physical parameter that can be used to explain the microscopic IP phenomena in a macroscopic manner (Oldenburg and Li 1994). I refer to Seigel (1959), Sumner (1976), Bertin and Loeb (1976), Pelton *et al.* (1978), Fink *et al.* (1990), and Ward (1990) for the basics, case histories, and mathematical background of forward modeling and inversion of IP data.

The response of a 2D chargeability model $\eta(x, z)$ can be calculated by calling the DC resistivity forward operator F_{dc} twice for the conductivity models $\sigma(1 - \eta)$ and σ, respectively, and evaluating the following equation:

$$\eta_a = \frac{F_{dc}\left[\sigma(1-\eta)\right] - F_{dc}\left[\sigma\right]}{F_{dc}\left[\sigma(1-\eta)\right]} \ , \tag{4}$$

where η_a is the apparent chargeability.

2.3 Inversion

Inversion of DC resistivity data is a non-linear and ill-posed problem. An initial model supplied by the user is updated by an iterative process. Usually, a kind of model smoothness constraint is required to stabilize the inversion.

The model smoothness constraints link the neighboring model parameters (resistivity of adjacent rectangular cells) in order to prevent sharp changes in the individual resistivities which may result in a meaningless geoelectrical section. The inversion routine used in the program is based on the solution of the following equation (Loke 2014, Wolke and Schwetlick 1988, Farquharson and Oldenburg 1998):

$$\Delta \mathbf{m_i} = \left(\mathbf{J}^T \mathbf{W}_d^T \mathbf{W}_d \mathbf{J} + \lambda \mathbf{C} \right)^{-1} \left(\mathbf{J}^T \mathbf{W}_d^T \mathbf{W}_d \Delta \mathbf{d} - \lambda \mathbf{C} \mathbf{m}_{i-1} \right) , \tag{5}$$

where $\Delta \mathbf{m_i}$ is the model correction vector, $\mathbf{J}$ is the Jacobian matrix, λ is the regularization parameter, $\Delta \mathbf{d}$ is the data discrepancy vector, i denotes the iteration number, $\mathbf{W}_d$ is data weighting matrix, and $\mathbf{C}$ is the five-point finite difference Laplacian. The data weighting matrix is in the given form:

$$\mathbf{W}_d = \text{diag} \left(1/\boldsymbol{\varepsilon_i} \right) , \tag{6}$$

where $\boldsymbol{\varepsilon_i}$ denotes the individual standard deviation of the measurements in case of availability. Otherwise, the data weighting matrix is calculated by the following equation:

$$\mathbf{W}_d = \text{diag} \left(\frac{1}{\sqrt[4]{\boldsymbol{\rho}_{\text{app}}}} \right) . \tag{7}$$

The data weighting procedure aims to reduce the effects of extreme readings to the inversion. The smoothness operator C is formed by using the "delsq" MATLAB function. The function requires key numbers of a parameter grid representing the location of the parameter in the 2D model (see MATLAB documentation for further details). The initial value of damping factor λ is based on the standard deviation of the logarithms of the apparent resistivities and lowered to half after each iteration while it is greater than 0.01.

Following Trip *et al.* (1984), Sasaki (1994), and Spitzer (1998), the Jacobian matrix (**J**) was calculated by differentiating Eq. 2 with respect to cell conductivities, yielding:

$$\mathbf{K} \frac{\partial \mathbf{v}}{\partial \boldsymbol{\sigma_i}} = -\frac{\partial \mathbf{K}}{\partial \boldsymbol{\sigma_i}} \mathbf{v} . \tag{8}$$

In Eq. 8 only the partial derivatives of potentials with respect to cell conductivities are unknown. **K** and **v** matrices are already produced during the forward solution. The partial derivative of the stiffness matrix with respect to model parameters can be easily calculated by assigning 1 to relevant conductivities and 0 to all others. Because the approach of Marescot *et al.* (2006) is used for apparent resistivity evaluation, **K** is calculated twice. In the first

case all conductivities are assigned as 1 and result is stored as $\mathbf{K}_1$. Therefore, it is sufficient to pick the values of $\mathbf{K}_1$ relevant to the triangles inside a parameter block to calculate its derivative with respect to a certain parameter. This is easily implemented by creating an index of triangles relating them with the parameters during the mesh creation. Equation 8 then can be handled as a linear system of equations in the form of $Ax = b$ and must be solved for each model parameter. This is done as described previously for Eq. 2. In this manner, the Jacobian matrix is calculated for every iteration. An updated model is achieved by adding the calculated corrections to the model parameters. The new model is verified by comparing its response with the measured data. Data fitness is measured in the sense of RMS defined as:

$$\text{RMS} = \sqrt{\frac{\left(\mathbf{W}_d \Delta \mathbf{d}\right)^{\text{T}} \left(\mathbf{W}_d \Delta \mathbf{d}\right)}{N}} \ , \tag{9}$$

where N is the number of datum and superscript T denotes matrix transpose. The misfit function given in Eq. 9 is the cost function to be minimized by the inversion routine. The inversion process usually converges after 5 to 8 iterations.

   For the IP data inversion I followed the third method proposed by Oldenburg and Li (1994). The proposed method aims to minimize the objective function given in Eq. 9 by solving a nonlinear inverse problem where the data discrepancies are now defined as

$$\Delta d = \eta_{am} - \eta_{ac} \ ,$$

where η_{am} and η_{ac} are the measured and calculated apparent chargeability values in mV/V, respectively. The methodology used in the inversion of apparent resistivity data is used in the same form for apparent chargeability inversion. Equation 5 is solved once more to calculate the model parameter corrections where the model vector now contains chargeability values of the model cells and the data vector is formed by the observed apparent chargeability values. The sensitivity matrix in Eq. 5 is replaced with the following one:

$$J_{ij}^{\eta} = -\frac{\partial d_i}{\partial \eta_j} = -\sigma \left(1 - \eta\right) \frac{V_\sigma^i}{\left(V_\eta^i\right)^2} J_{ij} \ , \tag{10}$$

where V_σ and V_η are the calculated potentials for conductivity and chargeability models, respectively; and i and j are the data and parameter indices, respectively. J is the sensitivity matrix for the DC resistivity problem (Oldenburg and Li 1994). A known resistivity model is required for the evaluation of the response of a chargeability model. Therefore, the IP data

inversion is performed sequentially after the inversion of DC resistivity data. The recovered resistivity model, the Jacobian matrix at the last successful iteration and the calculated potentials are substituted in Eqs. 4 and 10 for necessary calculations. The same misfit function is used for the verification of the recovered chargeability models.

3.   USER INTERFACE

The user interface of ELRIS2D is designed for ease of use with all controls and options available on the same window (Fig. 2). The main window is divided into several panels. The first panel located on top left of the window is a useful and simple file explorer. As soon as the program starts, root folder (where the program is called) is scanned by a routine to find the data files with the supported file format. The data file format supported by ELRIS2D is identical to the format of well-known program RES2DInv. The program currently supports Pole-Pole, Pole-Dipole, Dipole-Dipole, Wenner, and Schlumberger arrays. The data are assumed to be measured by equally spaced electrodes. Please refer to the User Manual located in the root folder of the program for further details. Unsupported data file formats are simply
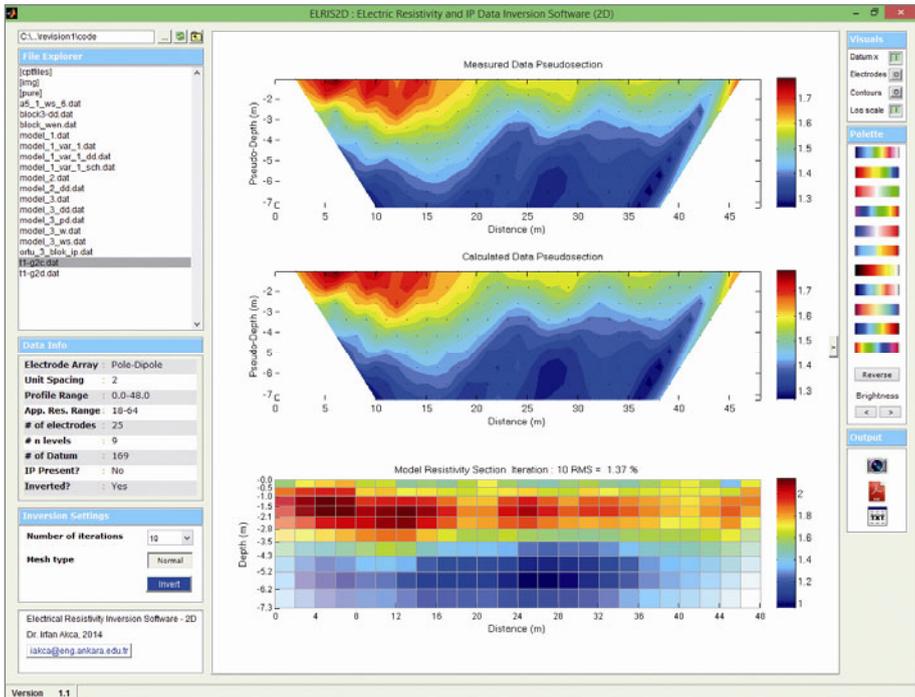
Fig. 2. The ELRIS user interface.

not listed in the File Explorer panel. The folder names in the current directory (if any) are given at the top of the list enclosed with brackets. This list box permits some interactions:

(1) double clicking on a folder name opens and scans the selected folder for supported data files,

(2) clicking on the name of a data file directly displays the content of the file as a pseudo-section,

(3) double clicking on to a data filename displays the results of previous inversion if available.

Browse and Directory Up buttons are also located at the top of the list box for changing the working directory (Fig. 2). The Refresh button may be used to update the file list in case of copying or producing new data files. Information about a selected data file is displayed in the next panel titled as "Data Info". In this context, the user can explore the content of a file just by clicking on the filename as a pseudosection supported by basic data information.

This is more practical than using a popup menu used in the order: File > Read Data > Select Folder > Select File > Ok. If the data file has already been inverted, it is indicated in the Data Info panel and the user can recall the results by just double clicking the filename.

The Inversion Settings panel includes two options: number of iterations and the mesh type. Available mesh types are "normal" and "fine". Normal mesh corresponds to a discretization where a rectangular cell is located between each adjacent electrode. Fine mesh is constructed by dividing the space between adjacent electrodes into two rectangular cells. Inversion must be restarted after changing the mesh type.

The panels on the right hand side include most common options for displaying DCR/IP data and 2D model sections. The pseudosection plotting is done following the pseudo-depth concept defined by Edwards (1977). The x-location of a datum is assumed to be at the midpoint of the electrode array for a single measurement. Following this aspect, the datum coordinate is plotted on the section with a black dot. The user can toggle to display or not to display the locations of datum. In a similar way, the user has the option of displaying electrode locations. Pseudosections are plotted as color coded sections. However, the user has the option to display the contour lines as well. Resistivities of earth materials vary in a very wide range. Therefore, logarithmic color scales are favorable for a better representation of the changes in the measured data and model resistivities. An option is present in the "Visuals" panel to toggle the logarithmic and linear color scales. Any change in the available options is applied immediately to the displayed sections.

Depending on the range of resistivity data and model resistivities, different color maps may be more suitable to obtain the best image for the interpretation. For this purpose a plenty of color schemes are made available with thumbnail previews at the palette panel. The color scales can be reversed and the darkness can be adjusted by clicking the buttons located in the palette panel.

The outcome of an inversion may be exported into other formats by three available options. A snapshot of the sections panel may be captured by snipping tool of Microsoft Windows activated by clicking a button located in output panel. Alternatively, the data and model section can be exported as a pdf file. Finally, a text file containing cell resistivities and data fitness may be generated by clicking the icons located at output panel. These options become available only when three sections are displayed in the main panel. Some other customization options are available via context menus that can be activated by right clicking on the sections. Labels, titles, length of color scales, color scale limits can be modified using several input boxes. The sections panel may be enlarged by hiding the tools located on the right hand side of the user interface by clicking the toggle button located at the right edge of the sections panel.

The current version of the code has some limitations. ELRIS supports only frequently used 5 electrode arrays listed before. Non-conventional arrays and borehole electrodes are not supported. The earth surface is assumed to be flat. The user interface of the code does not have the ability to incorporate *a priori* information to the inversion. However, the source code of the program is open which permits the potential users and contributors to eliminate the aforementioned limitations. Apart from the listed limitations of the present program, the following features may also be added with little effort:

❑ allows user to select multiple data files at a time and invert them subsequently with the user supplied settings,

❑ combining the inverted resistivity models of a survey site to display a pseudo three dimensional resistivity distribution.

4. EXAMPLES

4.1 Test model 1

ELRIS2D was tested with synthetic and field data sets. The first example is a synthetic data set calculated for the model given in Fig. 3d. The model consists of two embedded bodies and a laterally discontinued overburden. The resistivities of the background, overburden and embedded bodies are set as 100, 250, and 2500 ohm-m, respectively. The embedded bodies simulate a cross-section of walls perpendicular to the section plane. The data were simulated for dipole-dipole electrode array with 25 equally spaced electrodes.
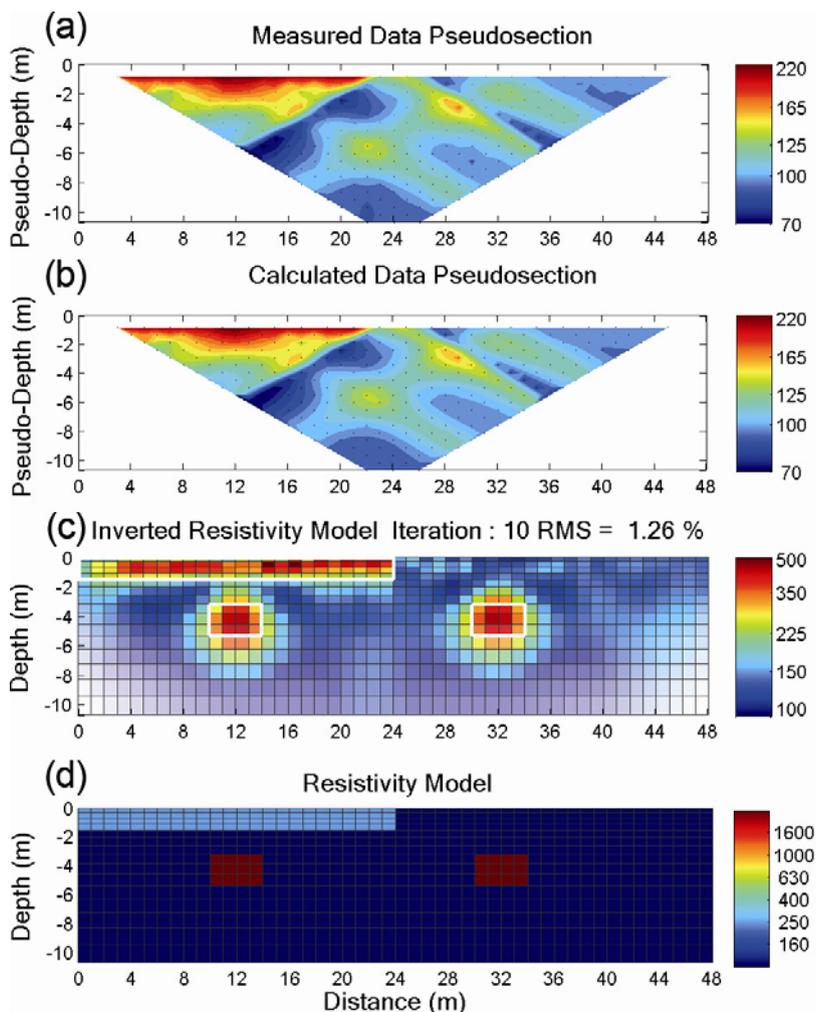
Fig. 3. Inversion results of the test model 1.

Electrode spacing was set to 2 m. Data were contaminated with Gaussian noise of 3%. Setting the response of the test model as the input data, an inversion with 10 iterations was carried out. The model mesh option was set to "fine", which means two rectangular cells are placed between each adjacent electrode. The model consists of 720 cells, each assigned an intrinsic resistivity. The expected value of RMS is 1 since the data difference is weighted. The final RMS at 10th iteration was 1.26 which is slightly above the expected error. Algorithm stops if the value of RMS tends to decrease below 1. This aims to prevent the algorithm from converging into the noise present in

the data. The overall inversion process for 10 iterations took 25.04 s on a PC with Pentium Dual Core processor and 4 GB RAM. The resulting inverse model resistivity section is given in Fig. 3c. Some cells in the illustrated model are faded. Fading of the model parameters is related with the value of total sensitivity of the cell, calculated by a samovar through the columns of Jacobian matrix at first iteration. The values are normalized with the maximum sensitivity and assigned as the degree of transparency for each cell. The exact locations of the model components are marked with white rectangles in Fig. 3c. The locations and size of the bodies are well defined in the reconstructed model.

4.2 Test model 2

The second model was designed to test the IP data inversion capability of ELRIS2D. The test model consisted of two embedded blocks and an overburden. The background resistivity was set to 100 ohm-m. Two embedded blocks simulate a conductive and a relatively resistive ore body where resistivities were set as 10 and 500 ohm-m, respectively. Only the embedded blocks are chargeable (250 ms); therefore, the chargeability values of background and the overburden were set to 0. The resistivity and chargeability distribution of the test model is given in Fig. 4d, h. The apparent resistivity and chargeability responses of the respective models are calculated assuming a multi-electrode system with 37 take-outs is used. The synthetic apparent resistivity and chargeability data are evaluated for dipole-dipole electrode configuration with 25 m unit electrode spacing (Fig. 4a, e). Gaussian noise of 2% was added to the simulated data.

The IP forward modeling and inversion requires a known resistivity model. Therefore, the DC resistivity response of the test model 2 is inverted first. The responses of the test model and recovered model are demonstrated in Fig. 4a, b, respectively. The recovered resistivity model is given in Fig. 4c with the exact locations of embedded bodies and the overburden marked with white rectangles. Both the conductive and resistive bodies are outlined in the inverted model section. The data misfit after 10 iterations is calculated as 2.84%. The recovered model outlines the test model. The ELRIS2D automatically switches to IP inversion once the DC resistivity inversion is complete. The recovered resistivity model and the Jacobian matrix at the last successful iteration are passed to the IP inversion routine for necessary calculations. IP data inversion is straightforward because the previously created model mesh is used and the former routines are called with small modifications. ELRIS2D enables a radio button group whenever a data file containing apparent chargeability data is selected. User can switch between resistivity and chargeability sections by selecting appropriate option
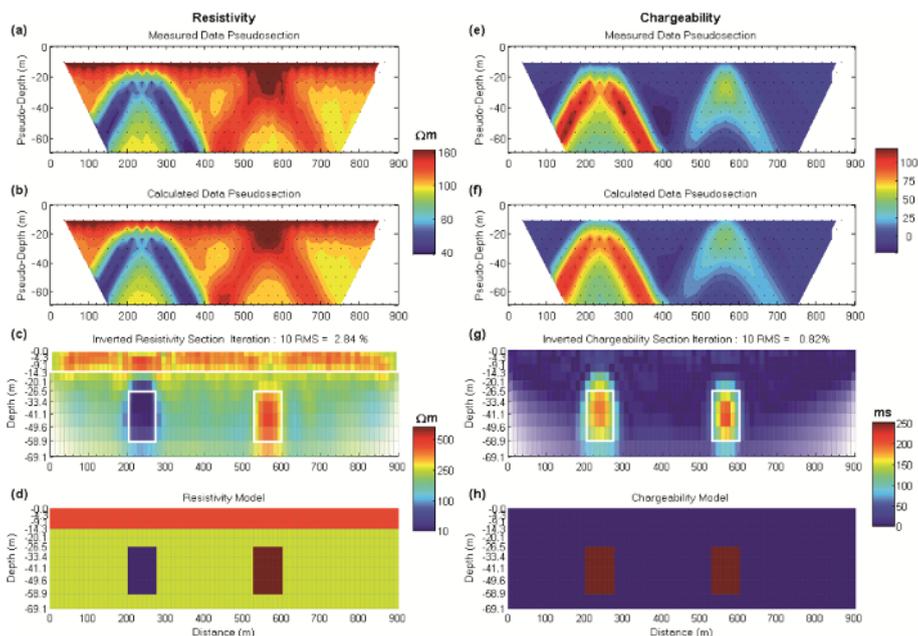
Fig. 4. The inversion results of test model 2. The left panel is for resistivity data and models, while chargeability data and models are shown on the right.

on the radio button group placed at the bottom of the sections panel. The inversion results of the IP data are demonstrated in the axes placed on the right panel of Fig. 4. The model and data space for IP inversion is linear. The recovered chargeability model is demonstrated in Fig. 4g with the exact location of embedded blocks marked with white rectangles. Remember that only the embedded blocks were chargeable and the background chargeability was set to zero. By visual comparison of the actual and recovered chargeability models it may be stated that the algorithm has the ability of outlining the position and chargeability of the anomalous bodies. This conclusion is supported with the fitness between the measured and calculated data sets. The apparent chargeability values are noise free and the data space is linear. Therefore, the data fitness is calculated as 0.82% after 10 iterations.

4.3 Field data test

The algorithm was also tested with field data sets. Field data were measured at the ancient city of Pisidian Antioch (Öztürk Akca 2011) located nearby the town of Yalvaç in southwest Turkey (Fig. 5a). The significant part of the city is still not excavated (Fig. 5b). Therefore, ERT measurements were carried out for the purpose of archaeological prospection. Measurements were
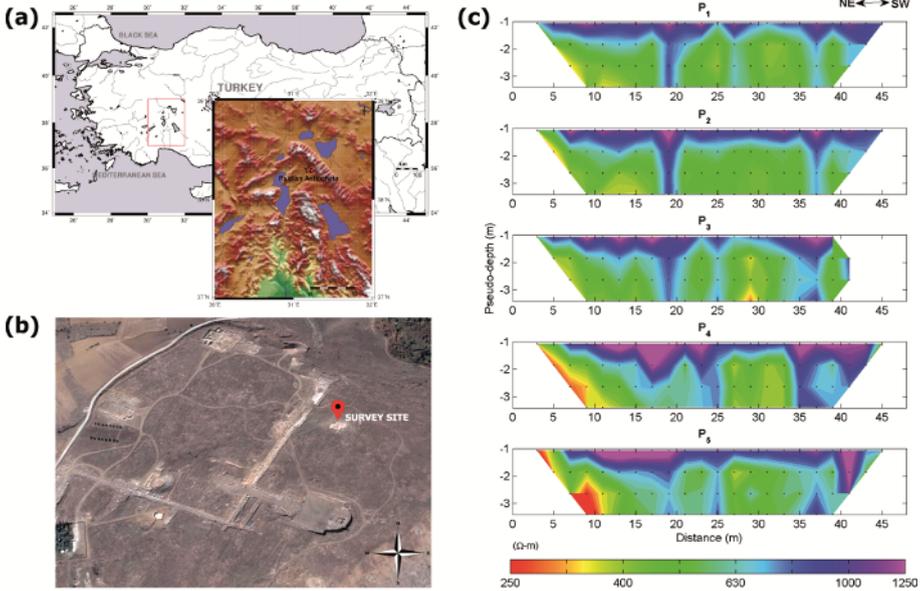
Fig. 5: (a) Location of survey area, (b) general view of the survey site, and (c) field data measured at Aedilis Hill.

taken by a multi-electrode system with 25 electrodes. The field data set was measured on the hill called "Aedilis" overlooking the city (Fig. 5b). Archaeologists expected the existence of a kind of temple due to the location of the hill. Therefore, a preliminary survey was carried out at the mentioned location. A Wenner–Schlumberger array with 2 m electrode spacing was used to measure the data on five parallel profiles 2 m apart. Profiles are numbered P1 to P5, starting from the one located at NW. The apparent resistivity pseudosections of the measured lines are demonstrated in Fig. 5c.

The data were inverted using a fine mesh consisting of 288 parameters. The number of iterations was set to 10. All data were inverted using the same settings. Data misfits varied between 5 to 11%. Inverted model resistivity sections of all profiles are illustrated in Fig. 6a with same color range and a common logarithmic scale. Two main anomalous bodies can be distinguished from the background. The first one is located at a distance of around 20 m from the beginning of the lines. The strength of the anomaly reaches to 2000 ohm-m where the background resistivity is assumed to be around 500 ohm-m. The second anomaly is located between 36 to 42 m of the lines and may be interpreted as two close embedded bodies. The overall inspection of the resistivity sections proves the existence of the remains of an ancient structure. The study was a preliminary examination of the survey area;
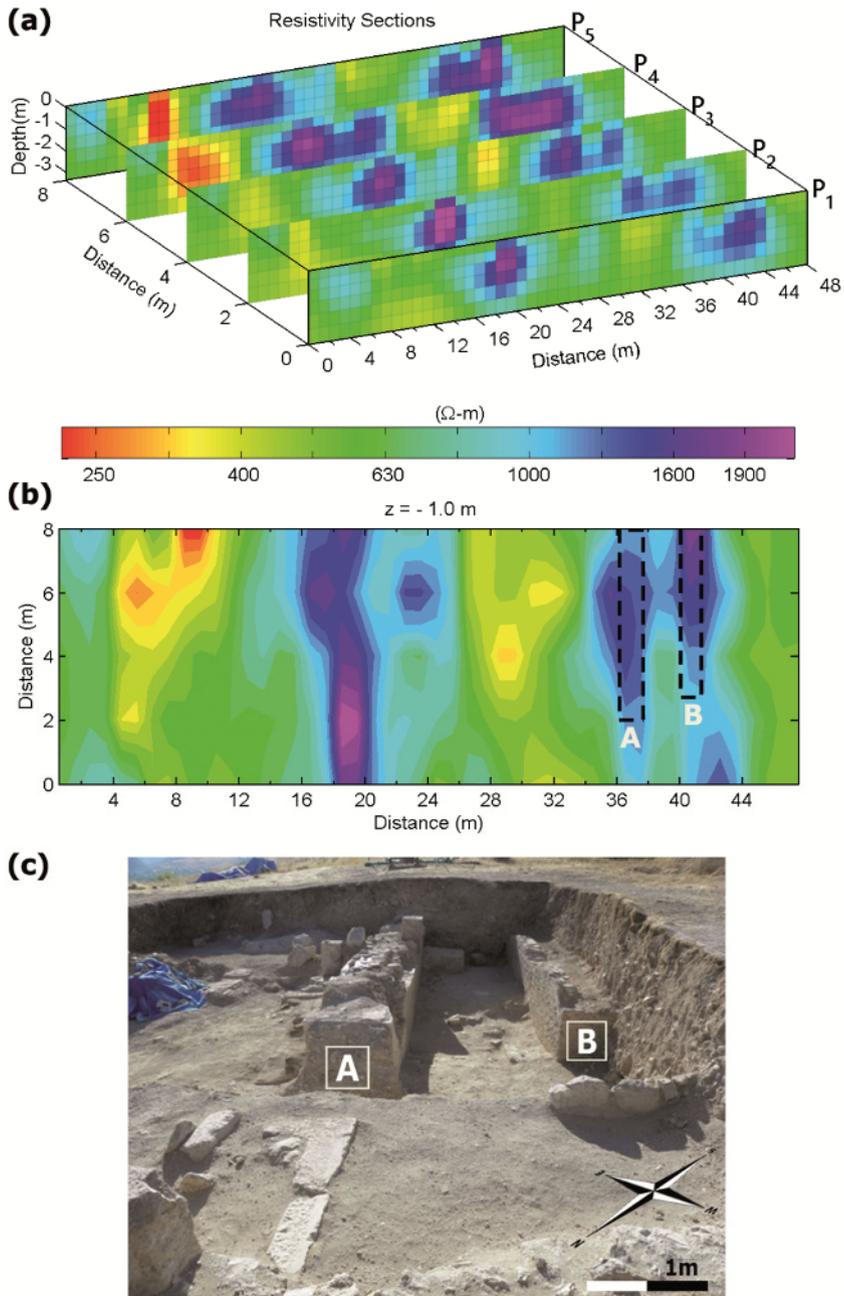
Fig. 6. Inversion results of field data: (a) model resistivity sections, (b) resistivity map of $z = -1.0$ m depth, and (c) excavation result at survey site.

therefore, the electrode and profile spacing were set to 2 m which is quite large for an archaeogeophysical survey. The survey site was excavated during the studies in 2011, which resulted in the exploration of the sixth church of the ancient city (Özhanlı, private communication). Figure 6c shows a picture of the excavation area. The anomaly located between 36 to 42 m of the lines corresponded to two walls made of limestone 2 m apart. The depth of the top of walls was 1.25 m from the surface. The dimension and locations of the walls are marked on the resistivity map of 1.0 m depth given in Fig. 6b. The other anomaly, clearly seen around 18-20 m, possibly arose from another structure, such as a wall. However, this interpretation has not verified because that part of the survey area is still unexcavated.

5. CONCLUSIONS

Two-dimensional inversion of resistivity and IP data is a well-established research area in geophysical community. Several commercial and non-commercial inversion codes are produced with good coverage of available inversion and visualization options. Emerging capabilities of measuring devices produce large amount of data in shorter times. Therefore, practical inversion programs are needed for the fast and efficient interpretation of acquired data sets. A new MATLAB program named ELRIS2D was developed for the two-dimensional inversion of DC resistivity and IP data to meet these demands. Two main new features of the program are the approach used to construct the model mesh and the practical user interface. Both features help to reduce the computation requirements and the time spent by the interpreter to process multiple data sets. The interactive file list box permits switching between data files in the current folder by just a mouse click or using arrow keys. Content of selected data files and previous inversion results (if available) are displayed immediately by appropriate mouse clicks. The inversion algorithm is based on the smoothness constrained least squares scheme. Based on the availability, program automatically inverts the apparent chargeability data whenever a resistivity data inversion is completed. A non-linear scheme is used to invert IP data. Switching between resistivity/ chargeability sections are as easy as clicking a radio button place on the user interface. The program provided satisfactory processing times (~ 1.5 seconds/ iteration for a moderate model although the Jacobian matrix is recalculated at each iteration).

ELRIS2D is an open source program which is expected to be improved by potential users and developers. Support for including topography in the inversion, simultaneous visualization of multiple sections, more visual and inversion options may be added for the next releases. The current version of the program is verified by two test models and a set of field data.

5.1 Availability and requirements

The source code and supplementary files of the program are available at https://goo.gl/PZ2tMC as a compressed folder. There is no specific setup procedure. It will be enough to unzip the contents of the compressed folder to a specific location. The program may be started by calling the main function elris.m from the MATLAB command line. The source code was developed on a PC operated by Windows 7 (64 bit) equipped with 4 GB RAM and a Pentium dual core processor. The program was tested on the MATLAB versions R2010a and R2013b resulting in a full functionality. The MATLAB version R2010a and newer must be installed on the user's computer. There is no strict hardware requirements, however it is suggested to have at least 3 GB of RAM installed. The data format and usage details are given in the user manual included in the main folder.

References

Akca, I., and A.T. Başokur (2010), Extraction of structure-based geoelectric models by hybrid genetic algorithms, *Geophysics* **75**, 1, F15-F22, DOI: 10.1190/1.3273851.

Başokur, A.T., and I. Akca (2011), Object-based model verification by a genetic algorithm approach: Application in archeological targets, *J. Appl. Geophys.* **74**, 4, 167-174, DOI: 10.1016/j.jappgeo.2011.05.004.

Bertin, J., and J. Loeb (1976), *Experimental and Theoretical Aspects of Induced Polarization*, Vols. 1 and 2, Gebrüder Borntraeger, Berlin.

Coggon, J.H. (1971), Electromagnetic and electrical modeling by the finite element method, *Geophysics* **36**, 1, 132-155, DOI: 10.1190/1.1440151.

DC2DInvRes (2014), DC2DInvRes – Tutorial, http://www. resistivity.net.

Dey, A., and H.F. Morrison (1979), Resistivity modelling for arbitrarily shaped two-dimensional structures, *Geophys. Prospect.* **27**, 1, 106-136, DOI: 10.1111/j.1365-2478.1979.tb00961.x.

EarthImager (2009) EarthImager 2D. Resistivity and IP inversion software. Instruction manual, Advanced Geosciences, Inc., Austin, USA, 139 pp.

Edwards, L.S. (1977), A modified pseudosection for resistivity and IP, *Geophysics* **42**, 5, 1020-1036, DOI: 10.1190/1.1440762.

Farquharson, C.G., and D.W. Oldenburg (1998), Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.* **134**, 1, 213-227, DOI: 10.1046/j.1365-246x.1998.00555.x.

Fink, J.B., E.O. McAlister, B.K. Sternberg, W.G. Wieduwilt, and S.H. Ward (eds.) (1990), *Induced Polarization: Applications and case histories*, Investigations in Geophysics, No. 4, Society of Exploration Geophysicists, Tulsa, DOI: 10.1190/1.9781560802594.

Günther, T. (2004), Inversion methods and resolution analysis for the 2D/3D reconstruction of resistivity structures from DC measurements, Ph.D. Thesis, Technische Universitaet Bergakademie, Freiberg, Germany.

Günther, T., and C. Rücker (2015), Boundless Electrical Resistivity Tomography BERT 2 – the user tutorial, http://resistivity.net/download/bert-tutorial.pdf.

Karaoulis, M., A. Revil, P. Tsourlos, D.D. Werkema, and B.J. Minsley (2013), IP4DI: A software for time-lapse 2D/3D DC-resistivity and induced polarization tomography, *Comput. Geosci.* **54**, 164-170, DOI: 10.1016/j.cageo.2013.01.008.

Loke, M.H. (2014), Tutorial: 2-D and 3-D electrical imaging surveys, http://www.geotomosoft.com/coursenotes.zip.

Loke, M.H., and R.D. Barker (1996), Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method, *Geophys. Prospect.* **44**, 1, 131-152, DOI: 10.1111/j.1365-2478.1996.tb00142.x.

Loke, M.H., J.E. Chambers, D.F. Rucker, O. Kuras, and P.B. Wilkinson (2013), Recent developments in the direct-current geoelectrical imaging method, *J. Appl. Geophys.* **95**, 135-156, DOI: 10.1016/j.jappgeo.2013.02.017.

Marescot, L., S. Rigobert, S.P. Lopes, R. Lagabrielle, and D. Chapellier (2006), A general approach for DC apparent resistivity evaluation on arbitrarily shaped 3D structures, *J. Appl. Geophys.* **60**, 1, 55-67, DOI: 10.1016/j.jappgeo.2005.12.003.

Mufti, I.R. (1976), Finite-difference resistivity modeling for arbitrarily shaped two-dimensional structures, *Geophysics* **41**, 1, 62-78, DOI: 10.1190/1.1440608.

Oldenburg, D.W., and Y. Li (1994), Inversion of induced polarization data, *Geophysics* **59**, 9, 1327-1341, DOI: 10.1190/1.1443692.

Öztürk Akca, C. (2011), Geoarcheological and new archaeogeopyhsical methods applied at ancient settlements: Psidia Antiocheia example, Ph.D. Thesis, Süleyman Demirel University, Isparta, Turkey.

Pelton, W.H., L. Rijo, and C.M. Swift (1978), Inversion of two-dimensional resistivity and induced-polarization data, *Geophysics* **43**, 4, 788-803, DOI: 10.1190/1.1440854.

Pidlisecky, A., and R. Knight (2008), FW2_5D: A MATLAB 2.5-D electrical resistivity modeling code, *Comput. Geosci.* **34**, 12, 1645-1654, DOI: 10.1016/j.cageo.2008.04.001.

Pidlisecky, A., E. Haber, and R. Knight (2007), RESINVM3D: A 3D resistivity inversion package, *Geophysics* **72**, 2, H1-H10, DOI: 10.1190/1.2402499.

Res2DInv (2014), Res2DInv user manual. Version 4.03, Geotomo Software, Penang, Malaysia.

Rijo, L. (1977), Modeling of electric and electromagnetic data, Ph.D. Thesis, University of Utah, Salt Lake City, USA.

Rücker, C. (2011), Advanced electrical resistivity modelling and inversion using unstructured discretization, Ph.D. Thesis, Leipzig University, Germany.

Sasaki, Y. (1994), 3-D resistivity inversion using the finite-element method, *Geophysics* **59**, 12, 1839-1848, DOI: 10.1190/1.1443571.

Seigel, H.O. (1959), Mathematical formulation and type curves for induced polarization, *Geophysics* **24**, 3, 547-565, DOI: 10.1190/1.1438625.

Shewchuk, R.J. (1997), Delaunay refinement mesh generation, Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, USA.

Si, H. (2008), Three dimensional boundary conforming Delaunay mesh generation, Ph.D. Thesis, Inst. für Mathematik, Technische Universitat Berlin, Germany.

Spitzer, K. (1998), The three-dimensional DC sensitivity for surface and subsurface sources, *Geophys. J. Int.* **134**, 3, 736-746, DOI: 10.1046/j.1365-246x.1998.00592.x.

Sumner, J.S. (1976), *Principles of Induced Polarization for Geophysical Exploration*, Developments in Economic Geology, Vol. 5, Elsevier Science Publ. Co., Amsterdam.

Tripp, A.C., G.W. Hohmann and C.M. Swift (1984), Two-dimensional resistivity inversion, *Geophysics* **49**, 10, 1708-1717, DOI: 10.1190/1.1441578.

Tsourlos, P.I., J.E. Szymanski, and G.N. Tsokas (1998), A smoothness constrained algorithm for the fast 2-D inversion of DC resistivity and induced polarization data, *J. Balkan Geophys. Soc.* **1**, 1, 3-13.

Ward, S.H. (ed.) (1990), *Geotechnical and Environmental Geophysics*, Vols. 1-3, Investigations in Geophysics, No. 5, Soc. of Explor. Geophysicists, Tulsa.

Wolke, R., and H. Schwetlick (1988), Iteratively reweighted least squares: algorithms, convergence analysis, and numerical comparisons, *SIAM J. Sci. Stat. Comput.* **9**, 5, 907-921, DOI: 10.1137/0909062.

Zhou, J., A. Revil, M. Karaoulis, D. Hale, J. Doetsch, and S. Cuttler (2014), Image-guided inversion of electrical resistivity data, *Geophys. J. Int.* **197**, 1, 292-309, DOI: 10.1093/gji/ggu001.

Water Storage Changes over the Tibetan Plateau Revealed by GRACE Mission

Jinyun GUO[1,2], Dapeng MU[1,3], Xin LIU[1], Haoming YAN[3],
Zhongchang SUN[4], and Bin GUO[1,2]

[1]College of Geodesy and Geomatics, Shandong University of Science
and Technology, Qingdao, China; e-mail: jinyunguo1@126.com

[2]State Key Laboratory of Mining Disaster Prevention and Control
Co-founded by Shandong Province and Ministry of Science and Technology,
Shandong University of Science and Technology, Qingdao, China

[3]State Key Laboratory of Geodesy and Earth's Dynamics, Institute of Geodesy
and Geophysics, Chinese Academy of Science, Wuhan, China

[4]Institute of Remote Sensing and Digital Earth, Chinese Academy of Science,
Beijing, China

A b s t r a c t

We use GRACE gravity data released by the Center for Space Research (CSR) and the Groupe de Recherches en Geodesie Spatiale (GRGS) to detect the water storage changes over the Tibetan Plateau (TP). A combined filter strategy is put forward to process CSR RL05 data to remove the effect of striping errors. After the correction for GRACE by GLDAS and ICE-5G, we find that TP has been overall experiencing the water storage increase during 2003-2012. During the same time, the glacier over the Himalayas was sharply retreating. Interms of linear trends, CSR's results derived by the combined filter are close to GRGS RL03 with the Gaussian filter of 300-km window. The water storage increasing rates determined from CSR's RL05 products in the interior TP, Karakoram Mountain, Qaidam Basin, Hengduan Mountain, and middle Himalayas are 9.7, 6.2, 9.1, –18.6, and –20.2 mm/yr, respectively.

These rates from GRGS's RL03 products are 8.6, 5.8, 10.5, –19.3 and –21.4 mm/yr, respectively.

Key words: GRACE, Tibetan Plateau, water storage, filter method, equivalent water height.

1. INTRODUCTION

The elevation of the Tibetan Plateau (TP) known as the third pole of the world is greater than 4000 m. TP has been more and more important in the study of global change and geodynamics. Due to the high elevation, the glacial area of TP is about 160 km^2 (see Fig. 1). On the background of global warming, the glaciers over TP have been sharply shrinking since the 1990s (Yao *et al.* 2013, Matsuo and Heki 2010).

Due to the limitation of data availability, it is very difficult to directly quantify the mass changes (*e.g.*, glacier, groundwater, and surface water) over TP and the Himalayas (Kääb *et al.* 2012). The glacier area and mass variations detected by satellite imagery and satellite altimetry have great differences on TP. ICESat implied that the shrinkage of glacier in the Himalayas from 2003 to 2009 was 10-30 mm/yr (Gardner *et al.* 2013). The digital elevation model showed the volume loss rate of 22 and 45 cm/yr (1999-2008) in the east and west Himalayas, respectively (Gardelle *et al.* 2013). Gardelle *et al.* (2013) used earlier time series than the GRACE data used in our analysis. Comparing the mass loss in the Himalayas, GRACE mission indicated that the water mass over TP was increasing by 7 Gt/yr during
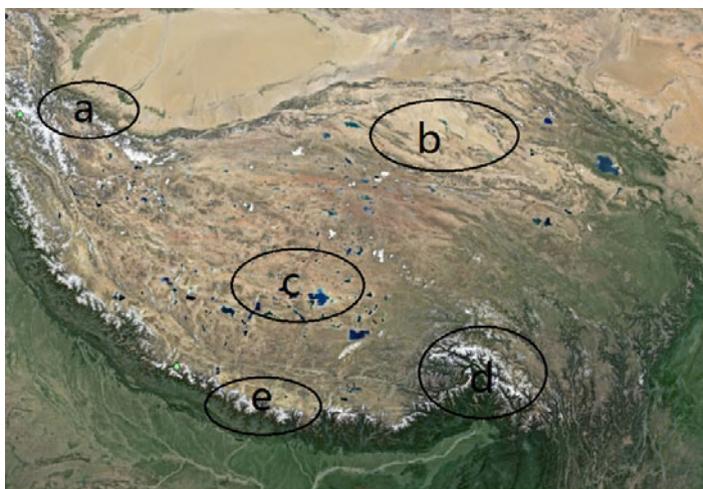
Fig. 1. Tibetan Plateau (from Google Earth): (a) Karakoram Mountain, (b) Qaidam Basin, (c) the middle TP, (d) Hengduan Mountain, and (e) the middle Himalayas are selected as black ellipsoids for further study.

2003-2010 (Jacob *et al.* 2012). ICESat also demonstrated that water levels of most lakes in the interior TP were rising (Song *et al.* 2013, Zhang *et al.* 2013). All the investigations proved that the mass changes over TP and surrounding area (*e.g.*, Himalayas) were spatially inhomogenous.

Since its successful launch, GRACE mission has been able to provide useful information of surface mass change in the large scale (Tapley *et al.* 2004). Because of different estimation methods and post-processing techniques, GRACE-derived mass changes varied (Guo *et al.* 2014, Ju *et al.* 2014, Mu *et al.* 2014), especially in regions with a weak geophysical signal. Taking the mascon approach, for example, a positive rate of 7 Gt/yr in TP was detected by the spectral domain inverse method from GRACE data (Jacob *et al.* 2012), while a positive rate of 30 Gt/yr was given by Yi and Sun (2014) with the space domain inverse method. This indicates that GRACE will deliver different estimates with different data processing strategies.

The main purpose of this paper is to investigate the water mass changes over TP by evaluating different GRACE solutions with different filter methods. Based on the time series of GRACE products (*i.e.*, we assumed that GRACE-derived signals were dominated by the linear and periodic components), a combined filter which integrates the statistical filter approach (Davis *et al.* 2008), the classical Gaussian filter (Wahr *et al.* 1998), and the decorrelation method (Swenson and Wahr 2006), is put forward to process monthly GRACE-derived Stokes coefficients estimated by the Center for Space Research (CSR). Monthly GRACE-derived geopotential coefficients provided by the Groupe de Recherches en Geodesie Spatiale (GRGS) were estimated by means of the truncated singular value decomposition (TSVD) technique which made the post-processing unnecessary (Save *et al.* 2012). To more accurately detect the water or glacier mass changes over TP, the post-glacial rebound (PGR) effect should be removed by the glacial isostatic adjustment (GIA). We also use the hydrological model to interpret GRACE-derived results.

2. DATA AND DATA PROCESSING

2.1 GRACE products of CSR

The newest GRACE monthly solution (level 2) from CSR is the RL05 model up to degree and order 60 (Bettadpur 2012). These geopotential data can be downloaded by http://isdc.gfz-potsdam.de. GRACE data through January 2003 to December 2012 are used in the paper. The missing data had been replaced by the mean value before and after the missing month with the linear interpolation.

We put forward a combined filter strategy to process CSR GRACE data, as seen in Fig. 2a. Firstly, the statistical filter approach (Davis *et al.* 2008),

(a)   (b)
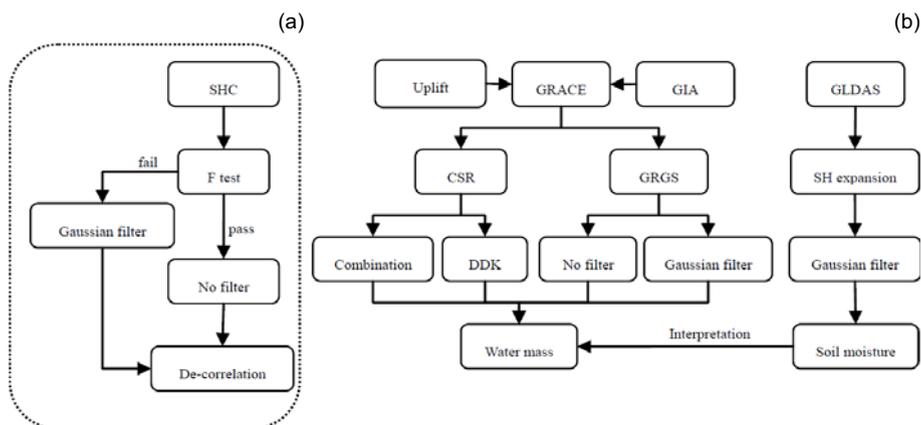

Fig. 2. Data processing designed in this paper. (a) was the combined filter strategy in which SHC represents the time series of GRACE SHC, F test was used in the statistical filter approach to examine the significance level of annual and linear components in GRACE series, and No filter means that the SHC will be kept unchanged. All Gaussian filter radius used in this paper was 300 km; (b) is the flowchart of data processing.

i.e., F test, is used to examine the significance level of annual and linear components in GRACE spherical harmonic coefficients (SHCs). If the SHCs pass the test, it would be kept unchanged. Otherwise it would be filtered by the Gaussian filter (Wahr *et al.* 1998). Then we used the de-correlation method (Swenson and Wahr 2006) to remove the correlation errors in GRACE products. Filter methods make the big effect on GRACE-derived results (Werth *et al.* 2009). As a comparison, the DDK2 filter (Kusche 2007) is used to process CSR geopotential data again. After filtering, GRACE SHCs are converted into the equivalent water height (EWH) (Wahr *et al.* 1998) with the spatial resolution of $1° \times 1°$.

2.2 GRGS regularization solutions

Recently, GRACE RL03 regularization solutions up to degree and order 80 are released by GRGS, which are available at http://grgs.obs-mip.fr/grace. Differing from CSR, GRGS employed TSVD to estimate the GRACE SHCs. It is unnecessary to filter GRGS monthly solutions while used to determine the surface mass anomalies. Even though the TSVD technique may reduce the errors in GRACE SHCs during the process of K-band range rate, we also try to use the post-processing strategy, *i.e.*, the Gaussian filter, to unify the spatial resolution of mass changes. To be consistent with CSR data in the frequency domain, we also truncated the GRGS data to degree and order 60 and then estimated the EWHs over TP.

2.3 GLDAS model

To better interpret and understand GRACE-derived results, the soil moisture is calculated from GLDAS model (Rodell *et al.* 2004) to represent the surface water change. GLDAS contains the soil moisture within the underground to a depth of 2 m. The original GLDAS data used in this paper are monthly with the spatial resolution of $1° × 1°$. In order to agree with GRACE data processing and its spatial resolution, we convert the spatial grids of GLDAS data into the spherical harmonic coefficients (Wahr *et al.* 1998), and then we use Gaussian filter with the window of 300 km to process the GLDAS coefficients. In the end, EWHs with grid of $1° × 1°$ are estimated by the filtered GLDAS coefficients.

2.4 GIA model

GRACE products are largely affected by PGR, especially in Antarctica (Velicogna and Wahr 2006), North America and North Europe, which needs to be corrected by the glacial isostatic adjustment model. Here we used ICE-5G model (Peltier 2004) which can provide the global uplift rate to estimate the correction for GRACE-derived EWHs.

3.   SPATIAL DISTRIBUTIONS OF GRACE-DERIVED EWHS

EWHs largely depended on the filter method and parameters from monthly GRACE SHCs (Werth *et al.* 2009). GRACE models released by different groups – like CSR, Jet Propulsion Laboratory, and GeoForschungsZentrum at Potsdam – will also lead to different results (Guo *et al.* 2014, Ju *et al.* 2014, Mu *et al.* 2014). So it is necessary to evaluate the different GRACE products and methods over TP. Figure 2 shows the flowchart of data processing to estimate water storages over TP. Figure 3 gives the spatial distributions of linear trends over TP determined by: (a) CSR combined filter, (b) CSR DDK2 filter (Kusche 2007), (c) GRGS without filter, and (d) GRGS with the Gaussian filter of 300-km window from 2003 to 2012, respectively. Since GRACE is particularly sensitive to PGR, we remove this effect by ICE-5G model (Peltier 2004). Due to the complicated geodynamics of TP, we have to consider the uplift of TP. Unfortunately, the uplift of TP is not well determined homogenously. Sun *et al.* (2009) showed that the uplift of whole TP was 1.2 mm/yr. We use this result to evaluate the effect of TP uplift on GRACE products.

   From Fig. 3, we can apparently see that the water mass in TP (especially in the middle and north region) was increasing, and in the Himalayas and the North India it was decreasing from 2003 to 2012, which is consistent with Rodell *et al.* (2009). In the north TP, an increasing rate of about 5 mm/yr is revealed by both the combined filter and DDK2 filter from CSR SHCs.
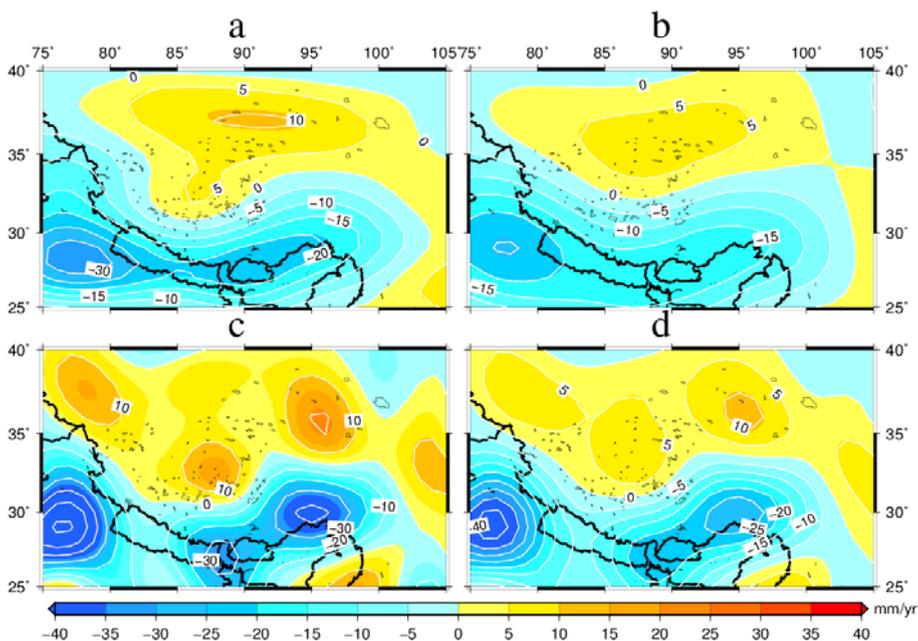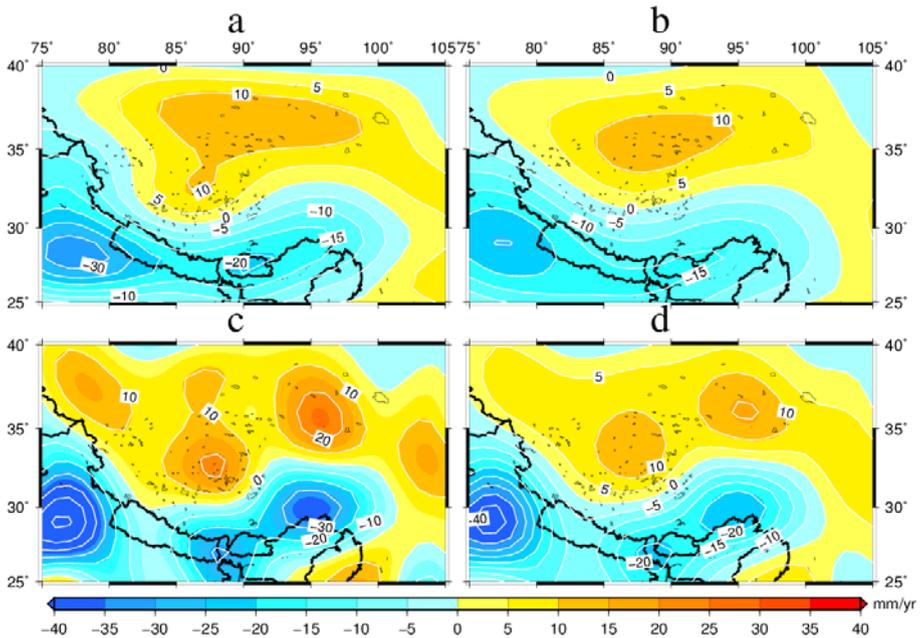
Fig. 3. Linear trends of EWHs derived from GRACE SHCs by: (a) CSR combined filter with window of 300 km, (b) CSR DDK2 filter, (c) GRGS regularization solutions truncated to degree 60 without any filter, and (d) GRGS solutions truncated to degree 60 with the Gaussian filter of 300-km window. The effect of PGR had been removed by using ICE-5G model, and the TP uplift of 1.2 mm/yr had been taken into consideration in the figure.

GRGS regularization solution gives more detailed mass change in the middle TP and the Qaidam Basin where the rate is greater than 10 mm/yr, as seen in Fig. 3.

It is clear that different filter methods will lead to different results and this is also true for different GRACE SHCs. GRGS processes its own K-band range rates to generate GARCE models instead of using those provided by the GRACE Project (Swenson and Wahr 2011). CSR RL05 data were estimated by the least squares technique whereas GRGS RL03 data were computed using TSVD which made the post-processing (*i.e.*, the filter method) unnecessary (the constraints are mainly made for the high degree; > 30).

We considered the effect of uplift (about 1.2 mm/yr based on Sun *et al*. (2009)) in TP and Himalayas. This uplift may be inappropriate because Sun *et al*. (2009) only used observation data from three stations. In some studies, the effect of uplift was excluded due to the mass loss beneath TP. It is prob-

Fig. 4. Linear trends of EWHs derived from GRACE SHCs without the uplift effect: (a) CSR combined filter with window of 300 km, (b) CSR DDK2 filter, (c) GRGS regularization solutions truncated to degree 60 without any filter, and (d) GRGS solutions truncated to degree 60 with the Gaussian filter of 300-km window. The effect of PGR had been removed by using ICE-5G model, and no TP uplift had been taken into consideration in the figure.

able that the large scale tectonic uplift would be compensated by the mass loss which consequently makes no significant contributions to GRACE results (Jacob *et al.* 2012). The uplift of 1.2 mm/yr is roughly equivalent to EHWs of 3-4 mm/yr (Wahr *et al.* 2000). Figure 4 shows the EWH change trends without the uplift effect. In the north TP, the mass increasing rate from CSR with the combined filter is more than 10 mm/yr. In the middle TP and Qaidam Basin, this rate determined by GRGS (Fig. 4c, no filter applied) is about 20 mm/yr. It is apparent that the mass change rate from GRGS became smaller through the Gaussian filtering shown in Fig. 4d. The reason we applied the Gaussian filter to GRGS SHCs is that we want to determine the detailed signal comes from high degree or low degree. Figure 4d indicates that the concentration of mass rate in Fig. 4c mainly comes from low degree.

GRACE mission can measure the total water storage changes, including soil moisture, groundwater, surface water, ice/snow, and glacier. It is necessary to interpret the geophysical mechanism of water storage increase over
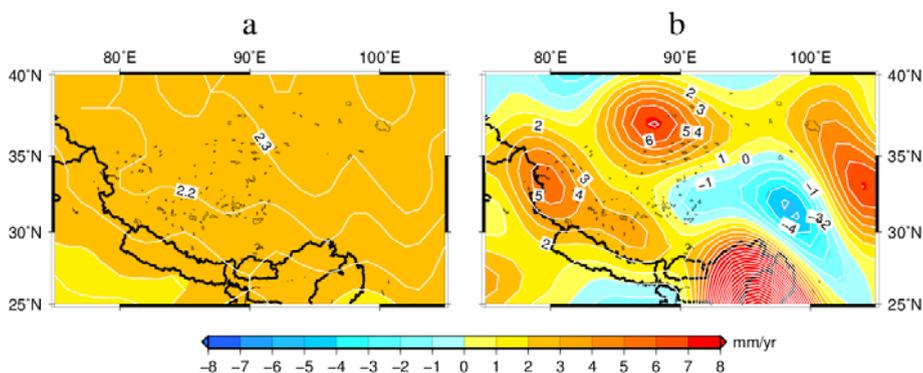
Fig. 5. Effects of PGR and soil moisture: (a) correction for GRACE by ICE-5G model, and (b) soil moisture rate from GLDAS model with the Gaussian filter of 300-km window.

TP. Since PGR may have the positive effect on GRACE-derived EWHs, we use ICE-5G model to evaluate it. In TP, PGR contributes about 2.2-2.3 mm/yr (Fig. 5a) to GRACE-derived results and this value has been corrected. Due to the limitation of data availability, we only computed the surface water using GLDAS model (Fig. 5b).

We used the 300-km Gaussian filter to process GLDAS model (see Fig. 2) to agree with the GRACE resolution. In the north TP, GLDAS had an increasing rate with 4-6 mm/yr, which was consistent with GRACE products. We can conclude that water increase in the north TP may be caused by the surface water. Of course, this conclusion is not rigorous since water changes are very complicated.

4.   REGIONAL WATER STORAGE CHANGES

EWHs of five regions selected in TP and surrounding area (see Fig. 1) will be analyzed in detail. Five series of EWHs are shown in Fig. 6 where CSR means the results derived by the combined filter and GRGS results are filtered by the Gaussian smoothing with the window of 300 km. Table 1 gives the linear and period changes determined by the least square fitting.

Region 'a' is located in the Karakoram Mountain where the mass increase rate is about 6 mm/yr and the rate difference between CSR and GRGS results is very small (see Table 1). Comparing to the mass loss in the Pamir Plateau, the mass increase in the Karakoram is called the Karakoram anomaly which is mainly caused by the glacier accumulation. Water storages over the Qaidam Basin (region 'b') and the interior TP (region 'c') are all increasing with the rates of 9.1 and 9.7 mm/yr, respectively, from the CSR results with the combined filter, see Table 1.
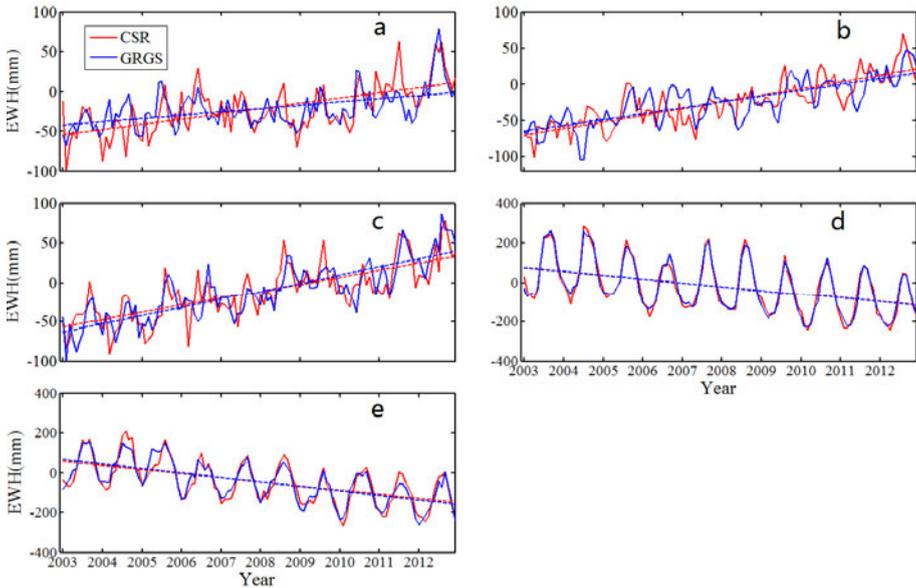
Fig. 6. EWHs in five selected regions: (a) Karakoram, (b) Qaidam, (c) the middle TP, (d) Hengduan, and (e) the middle of Himalayas. Here CSR means the combined filter results and GRGS were filtered by the Gaussian filter with window of 300 km. The data processing same as in Fig. 5.

Table 1

Least squares spectral analysis of water storages in five selected regions:
a – Karakoram, b – Qaidam, c – the middle TP, d – Hengduan,
and e – the middle of Himalayas (see Fig. 1)

Region	Data	Rate [mm/yr]	AA [mm]	AP [month]	SA [mm]	SP [month]
a	CSR	6.2	17	−1.2	4	0.2
a	GRGS	5.8	19	−1.6	5	0.3
b	CSR	9.1	18	−1.9	6	0.4
b	GRGS	10.5	16	−1.8	5	−0.1
c	CSR	9.7	16	−2.9	5	−0.3
c	GRGS	8.6	16	−2.7	12	−1.7
d	CSR	−18.6	102	−2.5	9	−1.6
d	GRGS	−19.3	95	−2.3	20	−2.1
e	CSR	−20.2	143	−2.9	31	−0.6
e	GRGS	−21.4	146	−3.1	30	−0.8

Explanations: AA stands for the annual amplitude, AP is the annual phase, SA is the semi-annual amplitude, SP is the semi-annual phase; CSR was filtered by the combined filter and GRGS was processed by the 300-km Gaussian filter.

From Fig. 6 and Table 1, we can find that the linear trends of EWH over regions 'a', 'b', and 'c' are very significant and the annual amplitudes are relatively small. AAs over regions 'a', 'b', and 'c' are less than 20 mm. Region 'b' is located in the Qaidam Basin and region 'c' is at the middle TP where there are lots of lakes. Considering the geographical environments of regions 'b' and 'c', we can infer that the water storage increase in regions 'b' and 'c' may be mainly caused by the surface water (soil moisture, lake water, and runoff) instead of the glacier, which is consistent with Song *et al.* (2013).

The annual variations and the linear trends of EWHs in regions 'd' (Hengduan Mountain) and 'e' (the middle Himalayas) are very obvious as shown in Fig. 6 and Table 1. AAs in regions 'd' and 'e' are greater than 90 and 140 mm, respectively, which indicates that the thaw and freezing of glacier are very strong. This may be related to the warming trend, especially in summer. Except for AA, the mass loss rates of regions 'd' (–18.6 and –19.3 mm/yr inferred from CSR and GRGS, respectively) and 'e' (–20.2 and –21.4 mm/yr inferred from CSR and GRGS, respectively) are also triggered by the warming trends.

One point should be noted, namely, that the regional water storage changes with the strong geophysical signal, *i.e.*, EWH anomaly, detected by different GRACE products and methods may have little differences as compared to TP investigated in this paper. On one hand, the geophysical signal (*e.g.*, water mass changes which is our interests), uplift of TP, GIA, and mass loss beneath TP, are hard to be independently separated from GRACE products without other auxiliary information. On the other hand, the filter method for SHCs and the data processing during Level-1B may also contaminate the feeble signals.

5. DISCUSSION

As revealed in this paper and others (*e.g.*, Jacob *et al.* 2012 and Yi and Sun 2014), GRACE shows that the water mass over TP was increasing for the period 2003-2012. Other independent remote data also confirmed this conclusion. In the interior TP, the satellite altimetry technique demonstrated that the water levels of most lakes were gradually increasing (Song *et al.* 2013, Zhang *et al.* 2011). Satellite imageries also revealed that the number of lakes, especially glacial lakes, is increasing (Wang *et al.* 2013). This tells us that the positive signal in TP was not only from the soil moisture but also from lakes (Zhang *et al.* 2013). This leads to another interesting question: where does the lake water or the soil moisture comes from? The previous studies demonstrated that TP had been experiencing the warming trends since the 1980s (Rangwala *et al.* 2009). The surface air warming, moistening

and solar dimming would trigger the more convective precipitation, especially over the central TP (Yang *et al.* 2014). Of course, the relation of precipitation and water increase needs to be examined by using more data.

In theory, the filter method is equal to the regularization solution (Swenson and Wahr 2011). The difference between these two techniques is to constrain the SHC before or after deriving SHC (Save *et al.* 2012, Eshagh *et al.* 2013). But in practice, different choices of parameter and solution approach will generate different results (Werth *et al.* 2009), *e.g.*, the water storage change in TP, as shown in this paper. It is easy to understand that feeble signal is very sensitive to the filter parameter and method. The merit of our combined filter is that it can keep the linear and annual signal because of the usage of statistical method (Davis *et al.* 2008). Our estimates of water increase rate for TP (5.3 Gt/yr, based on our combination filter method) agree more with Jacob *et al.* (2012; 7Gt/yr) than Yi and Sun (2014; 30Gt/yr). Given the precipitation data and other remote sensing data, the increase rate of 30 Gt/yr is hard to interpret reasonably. One possible reason for the 30 Gt/yr is that the observation grids used in Yi and Sun (2014) might be inappropriate (one degree grids over the entire Asia may overestimate the ice mass change, see more details in Yi and Sun 2014).

In the Himalaya, different GRACE products and methods agree generally better than in other regions (Fig. 4). But the EWHs estimated by GRGS are bigger than CSR (no matter filtered by combined filter or DDK filter). One possible explanation is that CSR data might have been over-smoothed. A simple test is to use the lower smooth radius (*e.g.*, 150 km Gaussian filter) to process CSR data (Jacob *et al.* 2012). If this procedure is introduced, an extra examination or comparison (with other independent data) must be done to make sure that the result is reliable.

6. CONCLUSIONS

GRACE data have revealed that the water storage over TP was increasing during 2003-2012. The combined filter strategy presented in this paper is more practical than the traditional Gaussian filter because the trend signal would be less attenuated. Compared with the mass loss of 30 Gt/yr in TP given by Yi and Sun (2014), our result of loss rate 5.3 Gt/yr was close to 7 Gt/yr (Jacob *et al.* 2012).

Different GRACE products and filter methods will lead to different results when the regional signal is relatively small, especially in TP, as shown in this paper. Compared with CSR data, GRGS regularization solutions give more detailed changes over TP (two centered mass changes in interior of TP) and the post-processing strategy is not necessary for GRGS. The geodynamics over TP is very complicated. The uplift of TP, effect of PGR, and mass

loss beneath the TP will amplify the uncertainties of GRACE results. One potential technique for better understanding the glacier variations over TP is the reconciled estimate method (Gardner *et al.* 2013), which of course depends on the data availability.

R e f e r e n c e s

Bettadpur, S. (2012), UTCSR Level-2 processing standards document for level-2 product release 05, Rev. 4.0, Center for Space Research, The University of Texas at Austin, USA.

Davis, J.L., M.E. Tamisiea, P. Elósegui, J.X. Mitrovica, and E.M. Hill (2008), A statistical filtering approach for Gravity Recovery and Climate Experiment (GRACE) gravity data, *J. Geophys. Res.* **113**, B4, B04410, DOI: 10.1029/ 2007JB005043.

Eshagh, M., J.-M. Lemoine, P. Gegout, and R. Biancale (2013), On regularized time varying gravity field models based on GRACE data and their comparison with hydrological models, *Acta Geophys.* **61**, 1, 1-17, DOI: 10.2478/s11600 -012-0053-5.

Gardelle, J., E. Berthier, Y. Arnaud, and A. Kääb (2013), Region-wide glacier mass balances over the Pamir–Karakoram–Himalaya during 1999-2011, *The Cryosphere* **7**, 6, 1263-1286, DOI: 10.5194/tc-7-1263-2013.

Gardner, A.S., G. Moholdt, J.G. Cogley, B. Wouters, A.A. Arendt, J. Wahr, E. Berthier, R. Hock, W.T. Pfeffer, G. Kaser, S.R.M. Ligtenberg, T. Bolch, M.J. Sharp, J.O. Hagen, M.R. van den Broeke, and F. Paul (2013), A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009, *Science* **340**, 6134, 852-857, DOI: 10.1126/science.1234532.

Guo, J., D. Mu, X. Liu, H. Yan, and H. Dai (2014), Equivalent water height extracted from GRACE gravity field model with robust independent component analysis, *Acta Geophys.* **62**, 4, 953-972, DOI: 10.2478/s11600-014- 0210-0.

Jacob, T., J. Wahr, W.T. Pfeffer, and S. Swenson (2012), Recent contributions of glaciers and ice caps to sea level rise, *Nature* **482**, 7386, 514-518, DOI: 10.1038/nature10847.

Ju, X.L, Y.Z. Shen, and Z.Z. Zhang (2014), GRACE RL05-based ice mass changes in the typical regions of Antarctica from 2004 to 2012, *Geodesy Geodynam.* **5**, 4, 57-67, DOI: 10.3724/SP.J.1246.2014.04057.

Kääb, A., E. Berthier, C. Nuth, J. Gardelle, and Y. Arnaud (2012), Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, *Nature* **488**, 7412, 495-498, DOI: 10.1038/nature11324.

Kusche, J. (2007), Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, *J. Geodesy* **81**, 11, 733-749, DOI: 10.1007/s001190-007-0143-3.

Matsuo, K., and K. Heki (2010), Time-variable ice loss in Asian high mountains from satellite gravimetry, *Earth Planet. Sci. Lett.* **290**, 1-2, 30-36, DOI: 10.1016/j.epsl.2009.11.053.

Mu, D.P., J.Y. Guo, Z.C. Sun, and Q.L. Kong (2014), Equivalent water height from GRACE gravity model based on principal component analysis, *Prog. Geophys.* **29**, 4, 1512-1517, DOI: 10.6038/pg20140405.

Peltier, W.R. (2004), Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE, *Ann. Rev. Earth Planet. Sci.* **32**, 111-149, DOI: 10.1146/annurev.earth.32.082503.144359.

Rangwala, I., J.R. Miller, and M. Xu (2009), Warming in the Tibetan Plateau: Possible influences of the changes in surface water vapor, *Geophys. Res. Lett.* **36**, 6, L06703, DOI: 10.1029/2009GL037245.

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll (2004), The global land data assimilation system, *Bull. Am. Meteorol. Soc.* **85**, 3, 381-394, DOI: 10.1175/BAMS-85-3-381.

Rodell, M., I. Velicogna, and J.S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in India, *Nature* **460**, 999-1002, DOI: 10.1038/nature08238.

Save, H., S. Bettadpur, and B.D. Tapley (2012), Reducing errors in the GRACE gravity solutions using regularization, *J. Geodesy* **86**, 9, 695-711, DOI: 10.1007/s00190-012-0548-5.

Song, C., B. Huang, and L. Ke (2013), Modeling and analysis of lake water storage changes on the Tibetan Plateau using multi-mission satellite data, *Remote Sens. Environ.* **135**, 25-35, DOI: 10.1016/j.rse.2013.03.013.

Sun, W., Q. Wang, H. Li, Y. Wang, S. Okubo, D. Shao, D. Liu, and G. Fu (2009), Gravity and GPS measurements reveal mass loss beneath the Tibetan Plateau: Geodetic evidence of increasing crustal thickness, *Geophys. Res. Lett.* **36**, 2, L02303, DOI: 10.1029/2008GL036512.

Swenson, S., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.* **33**, 8, L08402, DOI: 10.1029/2005GL025285.

Swenson, S.C., and J.M. Wahr (2011), Estimating signal loss in regularized GRACE gravity field solutions, *Geophys. J. Int.* **185**, 2, 693-702, DOI: 10.1111/j.1365-246X.2011.04977.x.

Tapley, B.D., S. Bettadpur, J.C. Ries, P.F. Thompson, and M.M. Watkins (2004), GRACE measurements of mass variability in the Earth system, *Science* **305**, 5683, 503-505, DOI: 10.1126/science.1099192.

Velicogna, I., and J. Wahr (2006), Acceleration of Greenland ice mass loss in spring 2004, *Nature* **443**, 7109, 329-331, DOI: 10.1038/nature05168.

Wahr, J., M. Molenaar, and F. Bryan (1998), Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.* **103**, B12, 30205-30229, DOI: 10.1029/98JB02844.

Wahr, J., D. Wingham, and C. Bentley (2000), A method of combining ICESat and GARCE satellite data to constrain Antarctic mass balance, *J. Geophys. Res.* **105**, B7, 16279-16294, DOI: 10.1029/2000JB900113.

Wang, X., F. Siegert, A.G. Zhou, and J. Franke (2013), Glacier and glacial lake changes and their relationship in the context of climate change, Central Tibetan Plateau 1972-2010, *Global Planet. Change* **111**, 246-257, DOI: 10.1016/j.gloplacha.2013.09.011.

Werth, S., A. Güntner, R. Schmidt, and J. Kusche (2009), Evaluation of GRACE filter tools from a hydrological perspective, *Geophys. J. Int.* **179**, 3, 1499-1515, DOI: 10.1111/j.1365-246X.2009.04355.x.

Yang, K., H. Wu, J. Qin, C. Lin, W. Tang, and Y. Chen (2014), Recent climate changes over the Tibetan Plateau and their impacts on energy and water cycle: A review, *Global Planet Change* **112**, 79-91, DOI: 10.1016/j.gloplacha.2013.12.001.

Yao, T.D., D.H. Qin, Y.P. Shen, L. Zhao, N.L. Wang, and A.X. Lu (2013), Cryospheric changes and their impacts on regional water cycle and ecological conditions in the Qinghai–Tibetan Plateau, *Chin. J. Nat.* **35**, 3, 179-186.

Yi, S., and W. Sun (2014), Evaluation of glacier changes in high-mountain Asia based on 10 year GARCE RL05 models, *J. Geophys. Res.* **119**, 3, 2504-2517, DOI: 10.1002/2013JB010860.

Zhang, G., H. Xie, S. Kang, D. Yi, and S.F. Ackley (2011), Monitoring lake level changes on the Tibetan Plateau using ICESat altimetry data (2003-2009), *Remote Sens. Environ.* **115**, 7, 1733-1742, DOI: 10.1016/j.rse.2011.03.005.

Zhang, G., T. Yao, H. Xie, S. Kang, and Y. Lei (2013), Increased water over the Tibetan Plateau: From lakes or glaciers? *Geophys. Res. Lett.* **40**, 10, 2125-2130, DOI: 10.1002/grl.50462.

DE GRUYTER OPEN

Climate Change Impact on Hydrological Extremes: Preliminary Results from the Polish-Norwegian Project

Renata J. ROMANOWICZ[1], Ewa BOGDANOWICZ[1], Sisay E. DEBELE[1], Joanna DOROSZKIEWICZ[1], Hege HISDAL[2], Deborah LAWRENCE[2], Hadush K. MERESA[1], Jarosław J. NAPIÓRKOWSKI[1], Marzena OSUCH[1], Witold G. STRUPCZEWSKI[1], Donna WILSON[2], and Wai Kwok WONG[2]

[1]Institute of Geophysics, Polish Academy of Sciences, Warszawa, Poland;
e-mail: romanowicz@igf.edu.pl

[2]Norwegian Water Resources and Energy Directorate, Oslo, Norway

A b s t r a c t

This paper presents the background, objectives, and preliminary outcomes from the first year of activities of the Polish–Norwegian project CHIHE (Climate Change Impact on Hydrological Extremes). The project aims to estimate the influence of climate changes on extreme river flows (low and high) and to evaluate the impact on the frequency of occurrence of hydrological extremes. Eight "twinned" catchments in Poland and Norway serve as case studies. We present the procedures of the catchment selection applied in Norway and Poland and a database consisting of near-natural ten Polish and eight Norwegian catchments constructed for the purpose of climate impact assessment. Climate projections for selected catchments are described and compared with observations of temperature and precipitation available for the reference period. Future changes based on those projections are analysed and assessed for two periods, the near future (2021-2050) and the far-future (2071-2100). The results indicate increases in precipitation and temperature in the periods and regions studied both in Poland and Norway.

Key words: climate change, EURO-CORDEX, hydro-climatic projections, catchment selection.

1. INTRODUCTION

Examination of historical observations of climate variables indicates that the evidence for anthropogenic-induced climate change is overwhelming. Following the publication of the IPCC AR3, AR4, and AR5, numerous studies at various spatial scales have considered the impact of projected changes in the climate system on hydrology and water resources. After air temperature, freshwater availability is the second most responsive environmental variable to changes in climate (IPCC 2013).

A preliminary Polish climate adaptation programme has been developed for agriculture, water management, and coastal zone management (Sadowski 2008). It is based on a selection of GCMs and IPCC sea level rise scenarios. Recently, that preliminary adaptation program has been extended to consider extreme events via the work on the adaptation to climate change by Osuch *et al.* (2012). This work resulted in strategic plans for adaptation to climate change in two time horizons, 2020 (SPA 2020) (IOŚ–PIB 2013) and 2070 (SPA 2070) (both documents are in Polish). There have, thus far, been no adaptation plans developed for managing risks from extreme events. Until a political decision is made on the level of protection of risk-prone areas required, it is not possible to estimate the possible costs of protection either at present or in the future. It is also not possible to estimate trends in the future costs of protection (Osuch *et al.* 2012).

The collaborative project "Climate Change Impact on Hydrological Extremes (CHIHE)" is part of the Research Programme of the EEA/Norway Grants Framework. The Institute of Geophysics, Polish Academy of Sciences, acts as Project Promoter and the Norwegian Water Resources and Energy Directorate acts as Project Partner. The scientific novelty of the project, amongst other things, lies in the development and adaptation of new non-stationary approaches to the analysis of changes in quantiles of hydrological extremes on a catchment scale (floods and droughts) for past events and future climate scenarios.

The project aims are:

❑ to investigate the effect of climate change on extreme flows (floods and droughts) in selected twinned catchments in Poland and Norway;

❑ to set up a preliminary benchmark dataset for Poland;

❑ to assess the uncertainty related to climate scenarios and hydrological modelling of observed and simulated extreme events and their basin characteristics;

❑ to develop and apply in practice non-stationary frequency analysis tools for high and low flow extreme events;

❑ to evaluate the impact of climate changes on the frequency of hydrological extremes;

❑ to prepare and evaluate the adaptation strategies to changes in character-
istics of extreme events in Poland and Norway, with a focus on high flow
events.

The development of an adaptation strategy to projected climate change
impacts on high flow extremes, and a foundation for the future collaboration
on the adaptation to low flow events taking into account the uncertainty of
the projections are important project deliverables.

In this paper we present some results that cover the first year of the pro-
ject. Namely, in Section 2 we present the methodology developed for the se-
lection of catchments suitable for the analysis of climate change impact on
extreme events under different geographic and climatic conditions in Poland
and Norway, which was a task in the first work package. We also describe
the selected catchments forming a database of near-natural catchments, the
so-called Reference Hydrologic Networks (RHN), suitable for climate im-
pact comparative studies. It is important to note that the selected 10 Polish
catchments represent the first such database in Poland.

Secondly, in Section 3 we present an analysis of climate model projec-
tions obtained from the EURO-CORDEX initiative (Jacob *et al*. 2014, Kot-
larski *et al*. 2014) which has developed future climate projections for the
21st century. In the project, projections from the ENSEMBLE project (van
der Linden and Mitchell 2009) were also used (*e.g.*, Osuch *et al.* 2015).
However, this is the first time that the EURO-CORDEX data have been ap-
plied in Poland.

In Section 4 we present a comparison of simulated and observed time se-
ries in the reference period 1971-2000. Section 5 describes future projections
of air temperature and precipitation obtained from seven GCM/RCM models
with the RCP4.5 emission scenario as an example. We thus present results of
an analysis of precipitation and temperature projections in three periods: ref-
erence period (1971-2000), near future period (2021-2050), and far future
period (2071-2100). Conclusions from the first year of studies in the CHIHE
project and further plans are given in Section 6.

2.   HYDROLOGICAL EXTREMES

2.1 Floods

Recently, Poland has experienced a number of catastrophic floods that have
endangered people and property. In 2001-2010 losses related to flooding in
the Upper Vistula catchment were over 14 billion PLN. Since 1990, 259 sig-
nificant floods have been observed in Central Europe, of which 165 occurred
after 2000 (Kundzewicz *et al.* 2005). This increase may be due to better in-
formation exchange and changes in land use (*e.g.*, urbanization). Non-
sequitur, it is projected that global warming will cause an increased fre-

quency of extreme events in Europe. In addition, due to the projected decrease of snowfall and following a decrease in the risk of spring floods, the character and timing of the floods are expected to change. Even though there is no evidence of a direct relationship between the increased frequency of extreme events and climate change, the observed positive trend indicates an increased probability of urban flooding and flash floods. Moreover, changes in land use, such as urbanization, a decrease of natural retention and poor water management may strongly influence the number of extreme (flood and drought) events.

Norway has experienced a number of large floods. The most costly flood occurred in June 1995, with a total loss of 1.8 billion NOK. This flood was caused by melting snow in both the mountains and lower lying areas, combined with heavy rainfall. The total losses caused by natural disasters exceeded 2.3 billion NOK in 2011, a new record for a single year. Thirty-nine flood events occurred in Norway in that year. The years 2012 to 2015 have also had a large number of flood events, some with very high losses.

According to IPCC (2013), simulation results for the year 2050 indicate decreasing trends for river discharges and soil moisture in the summer-autumn periods and a shift in the flood season from March-April to January-February, due to earlier snowmelt. The influence of climate change on extreme flows must be distinguished from human-induced changes in order to prepare adaptation tools for the future. The adaptation of water management to an uncertain future requires a comprehensive study of the current state and future predictions of the number, severity and frequency of extreme events. Wilson *et al.* (2010) examined trends in flood magnitude and timing, summer drought duration and deficit volume in the Nordic countries for the periods 1920-2005, 1941-2005, and 1965-2005. A pan-Nordic dataset of 151 stream flow records was analysed by applying the Mann–Kendall test to detect spatial and temporal changes in floods and drought. A rough differentiation of the flood generation mechanism was achieved by considering spring and autumn floods separately. The period analysed and the selection of stations influenced the regional patterns. However, a signal towards earlier spring floods and more severe droughts in southern and eastern Norway was evident in all three periods. No consistent trends were found in flood magnitudes. A qualitative comparison of the findings with available projections for daily stream flow under a future climate for the region indicates that earlier snowmelt floods are generally consistent with expected future changes. Hence, the projected changes in floods which are a consequence of increased temperatures are reflected in the observed trends, whereas changes anticipated due to increases in precipitation are not. Wong *et al.* (2011) found that despite small changes in future meteorological drought characteristics, an increase in hydrological drought can be expected in the future in the south-

ernmost and northernmost parts of Norway. These findings partly show the same signal as the trends found by Wilson *et al.* (2010).

Changes in hydrological processes, including flooding, under a future climate in Norway have been investigated by Beldring *et al.* (2008), Lawrence and Haddeland (2011), and Lawrence and Hisdal (2011). The results indicate large regional differences in the impact of climate change on flooding. Areas which are currently dominated by autumn and winter floods derived primarily from heavy rainfall are expected to experience moderate to large increases in flood magnitudes under a future climate. This includes western and parts of northern Norway, and the entire coastal region. On the other hand, large river basins in which spring snowmelt dominates peak flows, as in the inland areas of southern and northern Norway, are projected to have a decreased flood magnitude. This is due to a decrease in winter snow storage, leading to a reduction in peak flows. Flood peaks in such catchments are also projected to occur earlier in the season. These projected changes in flood hazard are in good agreement with similar studies in Sweden (Bergström *et al.* 2012), in Finland (Veijalainen *et al.* 2010), and in the Baltic region (Kriaučiūnienė *et al.* 2008).

In Poland studies on possible scale of changes in climatic processes caused by anthropogenic forcing and in the light of the need for an adaptation strategy for water resources strategy in an uncertain environment were initiated in the 1990s by Prof. Z. Kaczmarek (Kaczmarek *et al.* 1996, Kaczmarek and Napiórkowski 1996). In particular, Kaczmarek (2003) examined the role of climatic and hydrological variability in assessing the cumulative risk of flood events. In that study, flood-risk estimation is combined with a frequency analysis of extreme hydrological phenomena to evaluate flood-induced damages. If, however, the process is non-stationary, the risk of flood damage may also depend significantly on the variability of hydrological processes. In that study, it was found that spatial variation was larger than seasonal variability, which may indicate changes related to land use and water management. Kundzewicz *et al.* (2013) presented a comprehensive review of flood risk assessments in Europe. Romanowicz and Osuch (2011) applied a number of different statistical and model-based tools to separate the influence of climate and land use changes on flows in the Upper Narew catchment. The results showed that changes related to water management were suppressing any other impacts on flow magnitude.

The potential effects of long-term environmental changes on flood risk, including climate change, should be considered in conjunction with both preliminary flood risk mapping and the development of flood risk management plans. Therefore, assessments regarding likely climate change impacts on flood frequency must be developed from global and regional climate projections (IPCC 2007) so that this future effect can be considered. In Norway,

the hydrological projections for changes in flooding under a future climate presented in Lawrence and Hisdal (2011) have been used to develop regional guidance for taking account of climate change in flood risk management. Three categories for changes in flood discharge have been recommended in practise: 0, 20%, and 40%. It is advised that flood hazard maps illustrate or otherwise indicate patterns of flood inundation both in today's and under a future climate, based on these assessed changes (*e.g.*, Edvardsen and Roald 2012).

2.2 Droughts

Droughts have a substantial impact on ecosystems and the agriculture of the affected region, but also for the country/region's economy in general. Obviously, the cause and effects of floods and droughts could not be more different; surprisingly, however, many statistical methods for their description and analysis are quite similar. Droughts, as floods, can be characterised by their severity, duration, intensity, inter-arrival time, and other direct and indirect parameters. Such a multi-parameter character of the phenomena induces modellers to concentrate on one dominating parameter of the flood (important for the sake of the project's purpose) or to apply multivariate methods commonly used in Flood Frequency Analysis (FFA) and Flood Risk Analysis (FRA). Among the many possible probabilistic methods for drought analysis (see, *e.g.*, Mishra and Singh 2011) the estimation of return periods for time-dependent drought parameters and univariate or bivariate drought analysis will be the main subject of investigation in our project.

An example of a univariate technique is the fitting of the sample frequency distribution of drought characteristics using univariate probability distribution functions (*e.g.*, Shiau and Shen 2001, Cancelliere and Salas 2004). However, analyses of the drought return period (and quantiles) and frequency of low flow occurrence require long data series. The availability of such series is even less commonly available than similar series for peak flows, due to the fact that droughts are infrequent but of long duration. Thus, long time series are required to capture a sufficient distribution of events for analyses. To meet the requirements of this project, a scrutiny of drought frequency should be accompanied by an analysis of its duration (Fernández and Salas 1999a, b; Chung and Salas 2000). This directs us towards either univariate models in which the duration of the flow below an assumed threshold is a random variable (Shiau and Shen 2001, Strupczewski *et al.* 2012) or bivariate (frequency-duration) statistics with the help of copulas. The former methodology was successfully applied by Strupczewski *et al.* (2012) for duration-depth-frequency modelling developed for flood analysis, but can easily be applied to drought modelling and extended to non-stationary conditions under climate change (Hisdal and Tallaksen 2003).

The latter techniques, based on univariate marginal distributions combined with copula functions, have recently become extremely popular among hydrologists (*e.g.*, Favre *et al.* 2004, Salvadori and De Michele 2004) and seem to be a cure-all for every aspect of the hydrology of extremes. Therefore, we plan to verify the limits of this methodology within the context of its applicability to drought modelling and to evaluate whether it is possible to use copulas to join several very different components into one useful model of droughts. To the best of our knowledge the application of copulas in hydrology is a novel approach in Poland (Strupczewski *et al.* 2015).

3.   CATCHMENT SELECTION

3.1 Criteria for catchment selection

The CHIHE study is based on the analysis of a set of study catchments, both for Poland and Norway. The basis for selecting catchments suitable for the analysis of the influence of climate forcing on extreme flows (both low and high flows) applied in this study follows the rules recommended for the development of Reference Hydrologic Networks (RHN). Many countries have invested in a RHN database of streamflow gauging stations that are maintained and operated with the intention of observing how the hydrology of watersheds responds to variations in climate (Stahl *et al.* 2010, Monk *et al.* 2011, Fleig *et al.* 2013). In general, streamflow and climate records from reference hydro-climate networks are used to represent near-natural river flow regimes from catchments with varying hydro-climate characteristics, usually assumed to be representative of a broad range of regions. The networks provide time series records suitable for investigating the predominant climate and catchment processes that govern changes in regional hydrology. Data from such "reference" networks are of fundamental importance for detection and attribution studies and for the validation of large-scale climate and hydrological models (Slack and Landwehr 1992, Mishra and Coulibaly 2010, Raje and Mujumdar 2010, Hannah *et al.* 2011).

The main criteria that should be fulfilled by the catchments belonging to RHN are (Whitfield *et al.* 2012):

- ❑ Degree of basin development. Ideally, catchments should be pristine or at least have stable land-use conditions; catchments have an urban area <10%;
- ❑ Absence of significant regulations, diversions, or water use. A catchment is considered natural only if there is no substantial control structure upstream or water extraction within the basin, or diversions between basins. When regulation is present in a basin, some gauging stations may be appropriate for analyzing high flows and average flows, but not, for example, low flows;

❑ Record length. Any RHN station must have a minimum record length of 20 years. This length ensures that underrepresented climatic or geographic areas, which are characterized by minimal data availability, are also included. However, record lengths should also be as long as possible to allow decadal variability to be distinguished from long-term trends; in this project we require a minimum of 30 years of record;

❑ Active data collection. A station is included in the network if it is currently active and is expected to continue operation until it achieves the desired record length;

❑ Data accuracy. Only stations with what is considered good quality data are included in the network, *i.e.*, stations with reliable rating curves and with records tested for homogeneity. Stations with poorly or unquantified factors, such as changes in the river bed profile winter ice and spring ice jams or other difficult measuring conditions, should not be included;

❑ Adequate metadata. Adequate metadata should be available to support the above five conditions.

These criteria should be supported by additional analysis of catchment geomorphology and an analysis of the hydrological regime of the catchment (Hannaford and Marsh 2006, Marsh 2010, Thorne *et al.* 2010).

In heavily populated parts of Europe, with a long history of human disturbance, there are few pristine catchments, so some degree of disturbance must be tolerated. In the United Kingdom, for example, catchments where the net impact of abstractions and discharges are considered to be within 10% of the natural flow, flows at or in excess of the Q95 are deemed suitable for representative basin status (Hannaford and Marsh 2008). Polish catchments are characterized by a high degree of agricultural and forest land use. It may be appropriate to use these catchments in our database if agricultural practices have remained constant during the period of interest. In Norway, many areas are considered to have pristine conditions; however, changes in grazing patterns over the past century have contributed to afforestation in some areas. This effect is currently unquantified, but again, must be tolerated within the context of this study until more information is available.

The applied selection procedure can be illustrated using Norwegian and Polish catchments. The Norwegian stations both belong to the National Reference Database (Fleig *et al.* 2013) and have had their modelling capabilities tested (Lawrence *et al.* 2009). In the case of Polish catchments the calibration and validation of rainfall-runoff models have to be performed. A conceptual rainfall-runoff HBV model (Bergström *et al.* 2001) was chosen for that purpose. Results of calibration and validation procedures are used as additional criteria for catchment selection. Namely, only catchments with good calibration/validation results are selected for further analysis, as bad model

performance indicates an unresolved problem in modelling a rainfall-runoff process in the catchment (Romanowicz *et al.* 2013). Additional selection criteria are linked to specific project aims, *i.e.*, a comparison of Polish and Norwegian catchments from the point of impact of climate change on hydrological extremes. Therefore, the selected catchments were required to: (i) provide a good spatial coverage across the country; (ii) provide representative stations of the different flood and summer drought characteristics, which can be used for the pairing of Norwegian and Polish stations; and (iii) to have catchment characteristics that aid pairing of the Norwegian and Polish stations.

3.2 Evaluation of observation data sets

In Poland a set of 41 catchments was available for use within the project. Precipitation data were compiled from 79 climate stations and 58 synoptic stations across the country. Temperature data from 79 synoptic and climatic stations were also available. The synoptic data cover the full time period 1951-2010, but among climatic stations only about 50% have a full data range. Out of all the available stations, 39 gauged Polish stations were selected to provide hydrological information. On average, the length of all records spans the period 1950 to 2010. All data sets have daily data. Thiessen polygons were applied to derive precipitation data specific to each catchment. Land use data for 1990, 2000, and 2006 were collected from the EEA website (EEA 2014) and used to assess land use change.

The Norwegian gauging stations considered for use within this project have been previously scrutinized for inclusion in the Norwegian Benchmark dataset for climate change studies (Fleig *et al.* 2013). That dataset consists of 138 long term records suitable for both high and low flow studies, together with information on the characterizes of the catchments.

The availability of precipitation and temperature data was not an issue for the Norwegian gauging station selection since catchment average values derived from 1×1 km^2 gridded observation based datasets are routinely used to calibrated and run the HBV catchment models, and will therefore be used in this study. These gridded datasets are available for Norway for the period 1958 to present and are presented at www.senorge.no, with the station data having undergone various quality checks before deriving the gridded series (Tveito *et al.* 2005). The gridded precipitation dataset was also the subject of a trend analysis of moderate to strong precipitation events across Norway in Dyrrdal *et al.* (2012). However, there are issues with using gridded datasets for trend analyses since different stations used to derive gridded temperature and precipitation values during a period could result in the false identification of a trend. In addition, the Meteorological Office in Norway (MET)

does not use homogeneous time series as a basis for their gridded datasets. Preliminary analyses have shown that changes in the station network and the addition of stations have a greater impact on the quality and homogeneity of the grids locally than the use of homogeneous or non-homogeneous time series. Nevertheless, similar problems are likely to arise deriving catchment average rainfall using alternative procedures, such as Thiessen polygons. In addition, the use of meteorological data from a single station, often located outside a catchment, can be unrepresentative of the true catchment average values.

Wilson *et al.* (2010) undertook a trend analysis of a pan-Nordic dataset of 151 pristine streamflow records, which includes many of the Norwegian stations available for use in this study. In addition to standard quality control procedures at the national hydrological institutions, a visual inspection of the dataset was undertaken prior to the analyses.

3.3 Hydrological regimes

Flood regime

In Poland 41 catchments were classified into three types of flood regime categories: (i) snowmelt, (ii) rainfall, and (iii) mixed, based on their seasonal event number contribution to the total number of annual events in each station. Where the contribution of winter season events is higher than 66.67% to the annual, the flood regime is classified as a snowmelt driven regime, and if the summer season contributes more than 66.67% to annual flood, it is defined as a rainfall driven regime. The mixed regime occurs in between. We used two types of threshold values: (i) daily maximum flow of 80-90% probability as a threshold to count the number of events in a specific catchment; (ii) rainfall/snow melt regime – contribution greater than 66.7% of flow is directly caused by rainfall/snow melt, respectively; mixed regime – if the percentage contribution is in between those probability ranges.

In Norway all 138 catchments have been classified on the basis of their flood regime into the following types: (i) snowmelt, (ii) rainfall, and (iii) mixed, based on the percentage of years that a flood is caused by either snowmelt, rainfall or a contribution of both factors. Where the percentage of years that the annual maximum flood is caused by either snowmelt or rainfall is greater than 2/3, the respective mechanism is given as the characteristics of the flood regime. Where neither snowmelt nor rainfall alone accounts for more than 2/3 of the annual maxima flood events at a station, the flood regime is classified as mixed. As an illustration, the classifications of Polish and Norwegian catchments for the final selected catchments are given in Figs. 1a and 1b, respectively.
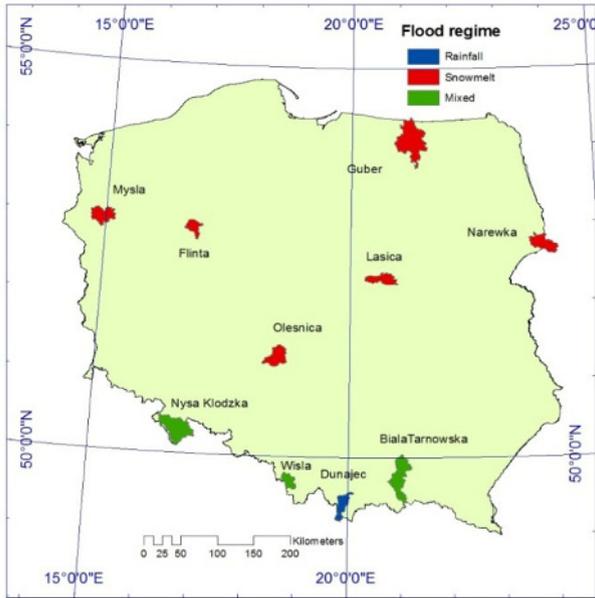
Fig. 1a. Flood regime classification in selected Polish catchments.
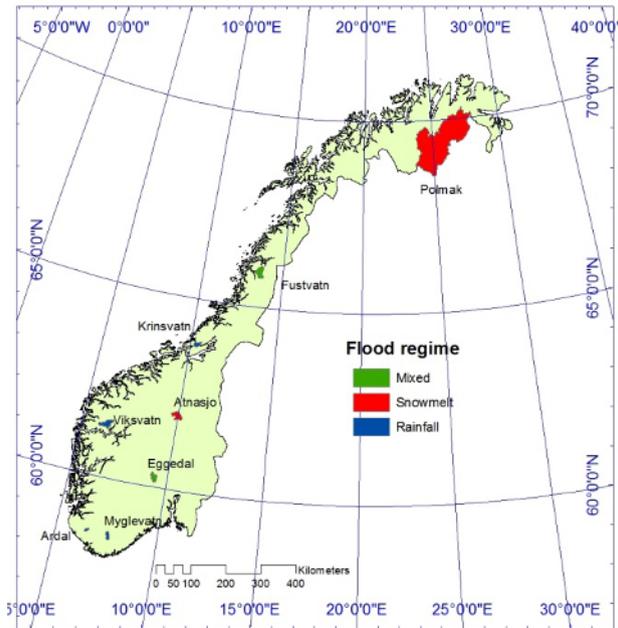

Fig. 1b. Flood regime classification for the Norwegian catchments based on the percentage of the events comprising the annual maximum series which are driven by > 60% rainfall (*versus* snowmelt) input.

3.4 Homogeneity testing

Given that floods and droughts are the focus of this study, the following parameters have been selected to establish the homogeneity of the data series at each station and to assess their suitability for use in this project:

❑ mean daily flow values,
❑ median annual flow,
❑ annual total (the sum of daily streamflow over a year),
❑ annual maxima daily peak flow,
❑ annual Q5 (flow with exceedance probability 0.05),
❑ annual summer minima, with a 7-day moving average (see definition of summer below),
❑ annual summer Q80 (flow with exceedance probability 0.8), with 7-day moving average.

Homogeneity testing has been undertaken for each of the parameters detailed above. A data series is homogeneous if it represents the natural discharge from a catchment (Astrup 2000). However, sudden inhomogeneity can occur due to, for example, changes in instrumentation, observation method, regulation, changes in gauge location and/or surrounding conditions. This study does not consider gradual changes or trends in the identification of inhomogeneity in each series, since this is the focus of a later assessment. Some inhomogeneity may affect the whole data series, whilst others may affect only high or low water levels. There are various homogeneity tests available, but it can be difficult to specify the actual date of an inhomogeneity, and it is therefore beneficial to compare the results of several tests in order to specify more accurately the date of a break (Astrup 2000). Figure 2 illustrates the homogeneity testing procedure used in this study, with further details about each of the tests given below:

❑ Visual inspection – visual inspection of a time series plot is arguably the most effective method for validating streamflow data (Gustard and Demuth 2009), and is used to identify individual potentially erroneous values or jumps in a data series;
❑ Bayesian analysis – this procedure is used to infer the probability that a data series is homogeneous, as well as identifying the probability that a jump exists within a dataset. This analysis is based on both prior understanding of the probability (P) of homogeneity within a data series (*i.e.*, before the data are available) and the actual data. For all of the homogeneity assessments in this study, the prior probability for homogeneity has been set to 50% for annual values, based on user experience for the Norwegian data series (Reitan and Petersen-Øverlier 2005). In determining whether a series is homogeneous, a threshold of homogeneity was taken to be 50%. Where $P > 50\%$, this is taken as evidence of homogeneity, and
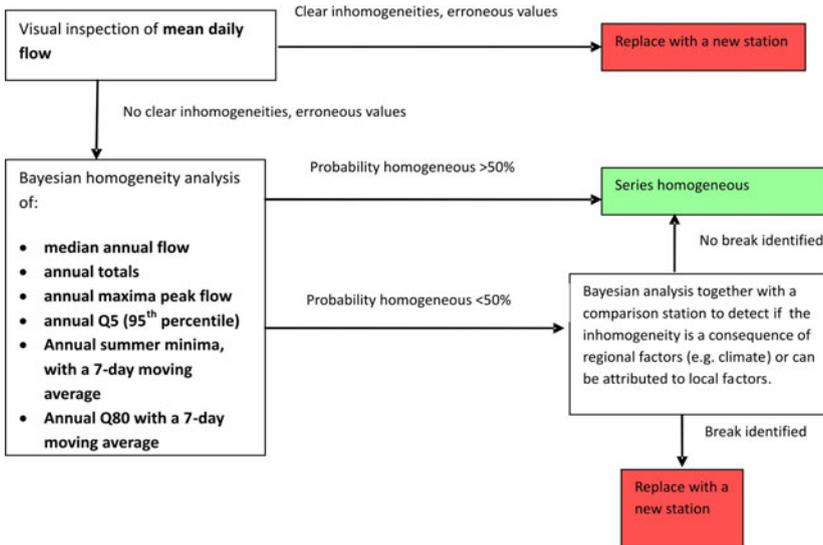
Fig. 2. Flow diagram illustrating the procedure undertaken to establish the homogeneity of the flow record at each gauging station.

where $P < 50\%$ this is taken as evidence of inhomogeneity. The strength of the evidence depends on how much P increases or decreases. For these analyses, NVE's in-house Hydra II software "Check Homogeneity" (Reitan and Petersen-Øverleir 2005) has been used.

❑ Bayesian analysis using a comparison station – this procedure seeks to detect a break in a station's data series, identified by comparison with the series from another nearby station. The aim of this homogeneity test differs from the Bayesian analysis detailed above, since it seeks to detect whether a jump in a series can be explained by regional climate variations which affect several stations, rather than an inhomogeneity being caused by changes in a local parameter at a station (*e.g.*, the gauging station being moved). This is undertaken by analyzing a new series which is the ratio between the subject series and the comparison series:

New series = subject series / comparison series

If homogeneities are influenced by regional factors, rather than local changes at a station, any jump should be detectable in both station series. A sudden jump in the new series indicates a possible inhomogeneity in only one of the series, but in which series the inhomogeneity lies cannot be determined based on this test alone. This analysis procedure relies on a single station or group of nearby stations with similar climate and catchment characteristics that can be used to help establish the homogeneity of

the selected station. Different parameters can also be compared (*e.g.*, precipitation and streamflow). However, if it is extremely wet or dry, two catchments may respond differently in comparison to a normal year.

Hydrological modeling

The HBV model was applied to rainfall-runoff modelling in both Norway and Poland. As mentioned in Section 3.1, the HBV model has already been tested for the Norwegian catchments (Lawrence *et al.* 2009). However, the HBV model had to be calibrated and validated for all selected Polish catchments.

As an input to the HBV models, precipitation, temperature, and streamflow data from 39 catchments were prepared. Observations for the period 1971-2000 were applied during the calibration stage, and the period 2001-2010 was used for validation (Table 1).

T a b l e 1

Results of calibration and validation (the Nash–Sutcliffe coefficient)
of ten selected catchments in Poland

Station number	Gauging station	River	1971-2000	2001-2010
			Calibration	Verification
1	Koszyce Wielkie	Biała Tarnowska	0.7874	0.7558
2	Nowy Targ Kowaniec	Dunajec	0.7678	0.7990
3	Skoczów	Wisła	0.6556	0.7541
4	Kłodzko	Nysa Kłodzka	0.6412	0.6039
5	Niechmirów	Oleśnica	0.7166	0.5309
6	Władysławów	Łasica	0.6030	0.5345
7	Narewka	Narewka	0.6876	0.5597
8	Prosna	Guber	0.5043	0.5270
9	Ryczywół	Flinta	0.7001	0.5911
10	Myślibórz	Myśla	0.7074	0.4989

The HBV model parameters were optimized using the DEGL (Differential Evolution with Global and Local neighbors) method (Storn and Price 1997). As an objective function the Nash–Sutcliffe coefficient was used (Nash and Sutcliff 1970). All Polish catchments marked in Fig. 1 were calibrated and validated during the selection procedure.

3.5 Final catchment selection

The selected study areas include ten catchments in Poland and eight catchments in Norway (Figs. 1a and 1b). These catchments have diverse hydroclimatic conditions. The watershed area ranges from 77.25 to 14 161 km^2.

The catchments are covered mostly by forest and are near-natural. The flood regimes of all selected catchments are driven either by rainfall and/or snow-melt (Table 2).

Table 2

List of selected catchments in Poland and Norway
for streamflow and precipitation variables

Country	Station	Catchment	Area [km^2]	Flood regime	PM [mm]	QM [m^3/s]
Poland	1	Biała Tarnowska	966.9	Mixed	2.1	9.0
	2	Dunajec	681.1	Rainfall	3.1	14.5
	3	Wisła	296.5	Mixed	2.6	6.1
	4	Nysa Kłodzka	1061.5	Mixed	2.1	12.9
	5	Oleśnica	583.5	Snow melt	1.6	2.5
	6	Łasica	629.3	Snow melt	1.5	1.1
	7	Narewka	635.3	Snow melt	1.8	3.0
	8	Guber	1554.5	Snow melt	1.6	8.5
	9	Flinta	813.4	Snow melt	1.5	0.7
	10	Myśla	586.9	Snow melt	1.5	1.3
Norway	1	Polmak	14161.4	Snow melt	1.4	179.2
	2	Årdal	77.25	Rainfall	9.3	12.5
	3	Fustvatn	525.69	Mixed	6.6	32. 7
	4	Krinsvatn	206.61	Rainfall	6.8	12.6
	5	Atnasjø	463.2	Snow melt	2.3	10.2
	6	Myglevatn	182.2	Rainfall	6.1	8.0
	7	Viksavtn	508.13	Rainfall	10.5	44.8
	8	Eggedal	309.77	Mixed	3.8	6.9

Explanations: *QM* is an average value of streamflow daily, *PM* is an average value of precipitation daily.

The temporal patterns of observed annual sums of precipitation and annual mean air temperature in ten Polish catchments for the period 1970-2000 are shown in Fig. 3. The catchments are located in varying hydro-climatic conditions of Poland: Nysa Kłodzka, Wisła, Dunajec, and Biała Tarnowska are in the highlands (south); Myśla and Flinta are in the western part; Łasica and Oleśnica are in the central; Narewka and Guber are located in the north and east part of the country. There is a large difference in precipitation amounts between the northern, southern, and western parts of Poland. Precipitation time series do not show clear trends in any of the catchments. The lowest precipitation is observed in Myśla, Flinta, and Oleśnica, situated in north-west Poland.
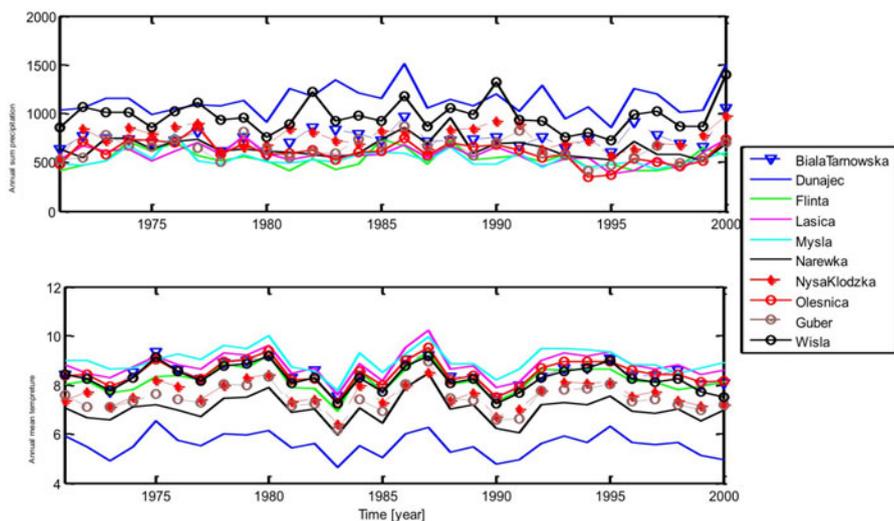
Fig. 3. Annual sums of precipitation [mm] (upper panel) and annual mean air temperature [°C] (lower panel), over the period 1971-2000 in the selected ten catchments in Poland.
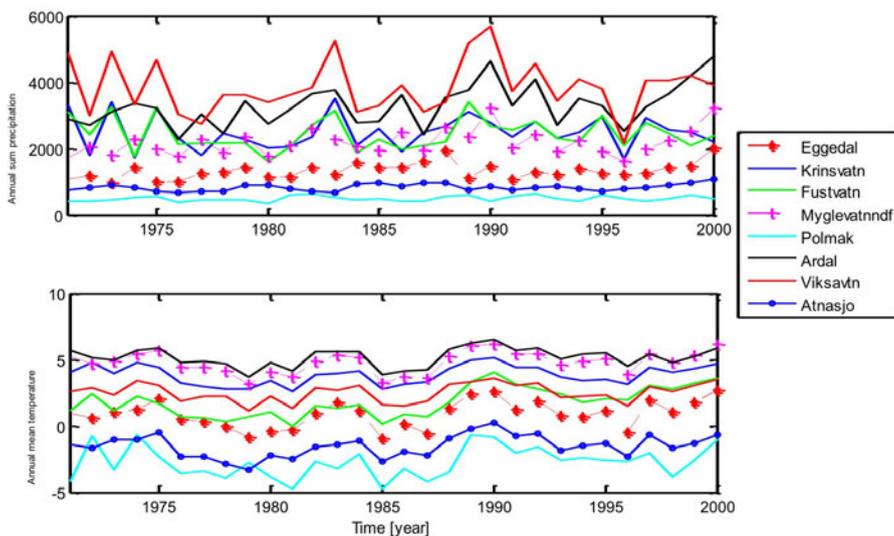

Fig. 4. Annual sum of precipitation [mm] (upper panel) and annual mean air temperature [°C] (lower panel), over the period 1971-2000 in the selected eight catchments in Norway.

The annual sums of precipitation and annual mean air temperature in the eight catchments from Norway are presented in Fig. 4. Amongst the Norwegian catchments, Myglevatn and Årdal are located not far from the southern

and southwestern coast of Norway, respectively. Viksvatn and Krinsvatn are located on the western coast of southern and middle Norway. All of these four catchments have relatively wet and warm climate conditions, by Norwegian standards. Fustvatn is also influenced by coastal conditions, with high precipitation volumes, although it is farther north and is therefore somewhat cooler than the other four. The remaining catchments, Eggedal, Atnasjø and Polmak, are all characterized by inland climate conditions with cold winters leading to lower annual temperatures and overall drier conditions throughout much of the year. The most northern catchment, Polmak, is most pronounced in this regard.

4. CLIMATE PROJECTIONS

To assess the impact of climate change on flood and drought conditions in Poland, we will use the newest source of climate projections available from the EURO-CORDEX initiative (Jacob *et al.* 2014, Kotlarski *et al.* 2014). The EURO-CORDEX simulations use the new Representative Concentration Pathways (RCPs) scenario products which are defined by Moss *et al.* (2010) and applied in the fifth IPCC Report. Four emission scenarios: RCP2.6, RCP4.5, RCP6, and RCP8.5, were created and named after a possible range of radiative forcing values in the year 2100 relative to pre-industrial values (+2.6, +4.5, +6.0, and +8.5 W/m^2, respectively) (Moss *et al.* 2010). RCP2.6 is a very low emission scenario requiring a considerable reduction from 2020, zero emissions from 2080, and removal of climate gasses from the atmosphere. This is the only scenario with a global warming of less than 2 degrees compared to 1850-1900. The RCP4.5 is an "intermediate pathway" in which radiative forcing is stabilized after 2100 (Clarke *et al.* 2007). The resulting global warming is estimated to be around 2.5 degrees towards the end of the century. RCP6.0 is also an intermediate pathway and very similar to RCP4.5 until the middle of the century. RCP8.5 is the highest emission scenario, also referred to as "business as usual", and implies a global warming of more than 4 degrees in 2100. According to Clarke *et al.* (2007), the new RCP scenarios have many advantages over the older SRES scenarios (Nakicenovic *et al.* 2000) in that they provide more information for understanding the climate system and carbon-cycle behaviour. The new scenarios also include more detailed information than provided by SRES: aerosol emissions, geographically explicit descriptions of land use and emissions, and detailed specification of emissions by source type (Edenhofer *et al.* 2010, van Vuuren *et al.* 2007).

The climatic variables to be used in this project have been obtained from EURO-CORDEX initiative in the form of time series of daily sums of precipitation and mean daily air temperature derived from seven different combinations of global and regional climate models: CNRM-CM5-CCLM4-8-17,

EC-EARTH-CCLM4-8-17,   EC-EARTH-HIRHAM5,   EC-EARTH-
RACMO22E, EC-EARTH-RCA4, MPI-ESM-LR-CCLM4-8-17, and MPI-
ESM-LR-RCA4 (Jacob *et al.* 2014, Kotlarski *et al.* 2014). The selected
combination of models consists of three GCMs: EC-EARTH, MPI-ESM-LR,
and CNRM-CM5 and four RCMs (Table 3). The selected available RCM/
GCMs provide projections of climatic variables up to the year 2100 at a
resolution of 12.5 km. Analyses of hydro-meteorological conditions were
conducted for the whole 1971-2100 period. For purposes of comparison,
three periods are used: 1971-2000, the so-called "reference" period, and two
future periods: "near future" (2021-2050) and "far future" (2071-2100) period.

Table 3

List of GCM/RCM climate models applied

GCM\RCM	RCA4	HIRHAM5	CCLM4-8-17	RACMO22E
EC-EARTH	✔	✔	✔	✔
MPI-ESM-LR	✔	–	✔	–
CNRM-CM5	–	–	✔	–

The quality of climate projections of precipitation and air temperature is
very important for reliable and accurate future extreme value predictions.
The RCM/GCMs simulations are generally biased and cannot be used as
forcing variables in flood and drought prediction at the catchment-scale
without some form of prior bias correction. Several studies have been per-
formed in which a bias correction method was applied to RCM data (*e.g.*,
Berg *et al.* 2012, Li *et al.* 2012, Teng *et al.* 2012, Lafon *et al.* 2013, Cloke *et
al.* 2013). In this study, quantile mapping methods (*e.g.*, Piani *et al.* 2010,
Gudmundsson *et al.* 2012) have been used for each catchment to correct
simulated precipitation and air temperature time series. Both empirical and
distribution-based methods have been applied. The methods are based on
correcting the RCM data such that the resulting distribution function better
matches that of the observed during the reference period. Empirical methods
use the empirical distribution function of the observed and RCM data for this
purpose, whereas distribution-based methods fit a particular function (*e.g.*,
a gamma function) to both the observed data and the RCM data. Further de-
tails of these procedures can be found in Gudmundsson *et al*. (2012). On this
basis, transformation of the quantiles of simulated time series relative to the
observed series is derived, and this correction is then applied to the simulat-
ed time series. Due to problems with realistic simulations of the seasonal
pattern of precipitation by the RCM models, these transformations were de-
rived independently for each month. This can sometimes lead to problems
during very dry summer months simulated by some of the RCMs, as a suffi-

cient number of dry days are not available to develop a robust correction. The transformations derived in the reference time period are also applied to correct RCM data for the future periods, thus assuming that the corrections are invariant with respect to climate change. For correcting the bias in the air temperature series from the RCM, empirical quantile mapping was used on the residuals, after removing the difference in the air temperature between the reference and the future period. This must be done to ensure that bias correction does not remove the climate change signal for temperature.

5. COMPARISON OF SIMULATED AND OBSERVED CLIMATE TIME SERIES IN THE REFERENCE PERIOD

In the first step of the analysis, a comparison of the observed and simulated precipitation and air temperature time series for the reference period (1971-2000) was performed. Figure 5a shows a comparison of historical monthly sums of precipitation, uncorrected precipitation, and corrected precipitation in the reference period for seven climate models for two Polish catchments, Dunajec (upper panel) and Narewka (lower panel). It is seen that all uncorrected simulations overestimate monthly sums of precipitation, except for the months of June, July, and August, when precipitation sums are underestimated by two models (m1 − CNRM-CM5-CCLM4-8-17 and m2 − EC-EARTH-CCLM4-8-17). Differences between climate model simulations are generally large. A comparison of the uncorrected precipitation simulations with observations indicates consistency in most catchments. The worst results are obtained for the two mountainous catchments: Nysa Kłodzka and Wisła. In general, the uncorrected precipitation simulations from EURO-CORDEX are significantly better than climate simulations from the ENSEMBLES project (Osuch *et al.* 2015). The application of bias correction by the double gamma method (Yang *et al.* 2010) significantly improved the simulated monthly total precipitation for the selected catchments in Poland.

A comparison of the observed and simulated monthly air temperatures in the 1971-2000 period for two studied catchments, Dunajec and Narewka, is presented in Fig. 5b. The results of comparison between uncorrected simulations (black lines) and observations (red line) indicate that most of the climate models underestimate the observed mean monthly air temperature. Only the simulations from CNRM-CM5-CCLM4-8-17 overestimate air temperature in summer months. As in the case of precipitation, the application of bias correction significantly improves air temperature simulations.

For Norway, Sorteberg *et al.* (2014) compared different bias correction methods for precipitation and temperature, including both Empirical Quantile Mapping and single and double gamma functions for a limited set of locations in Norway, based on the EURO-CORDEX RCM data available at that time. A range of multi-objective indices was compared to evaluate
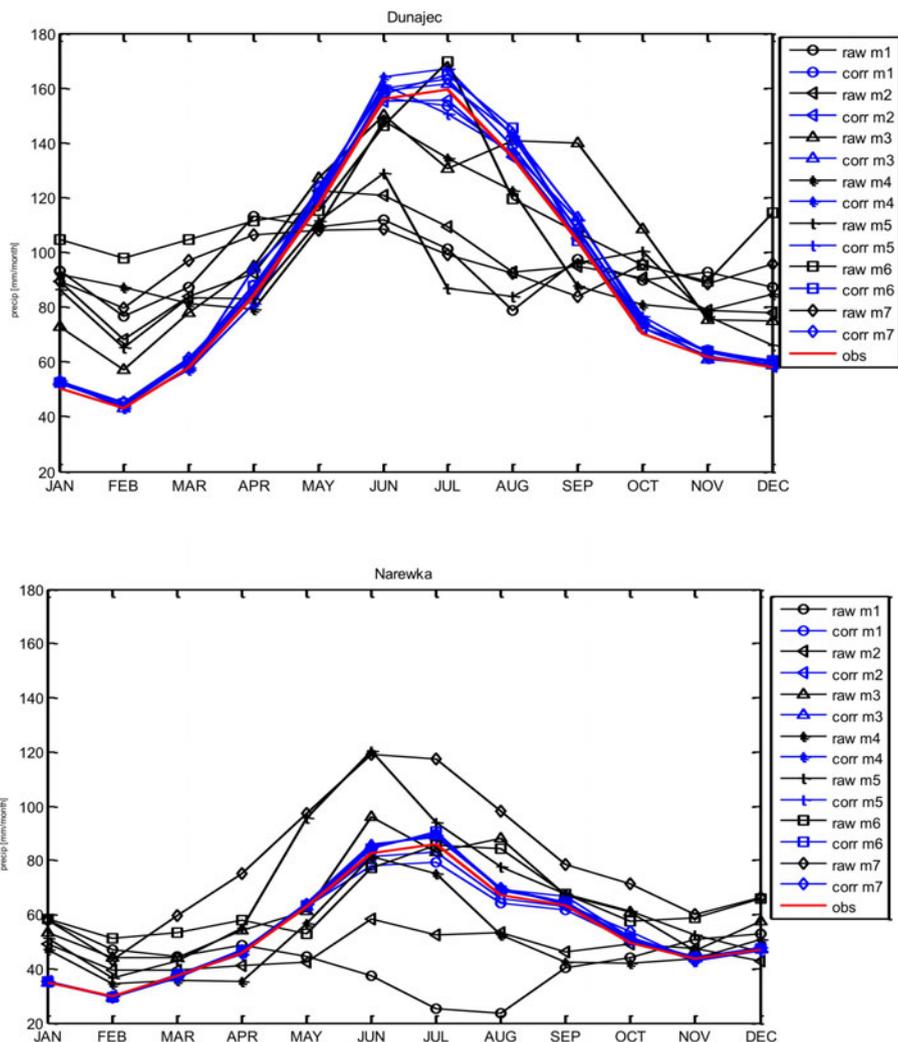
Fig. 5a. Comparison of historical monthly precipitation totals (red color), uncorrected precipitation (black color), and corrected precipitation (blue color) in the reference period (1971-2000) for seven climate models (m1 − CNRM-CM5-CCLM4-8-17, m2 − EC-EARTH-CCLM4-8-17, m3 − EC-EARTH-HIRHAM5, m4 − EC-EARTH-RACMO22E, m5 − EC-EARTH-RCA4, m6 − MPI-ESM-LR -CCLM4-8-17, m7 − MPI-ESM-LR -SMHI-RCA4) for two catchments located in Poland, Dunajec (upper panel) and Narewka (lower panel).
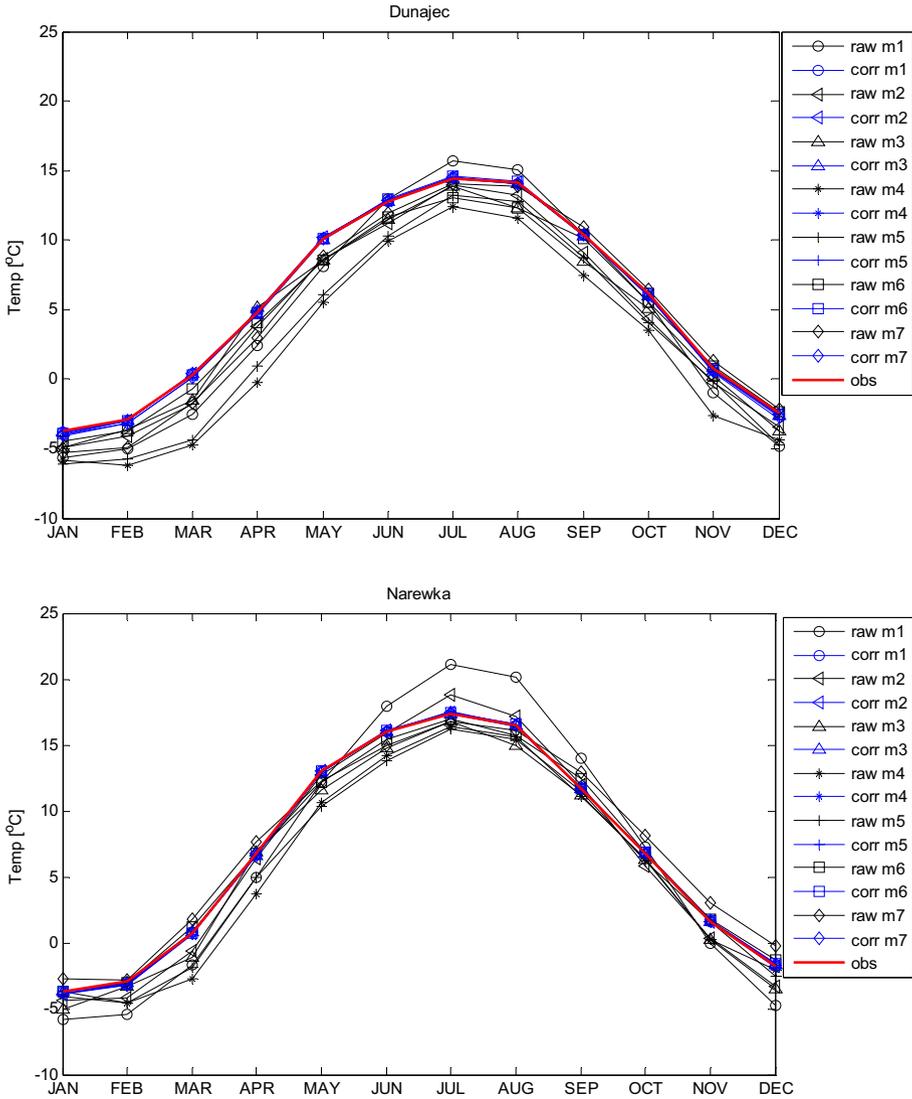
Fig. 5b. Comparison of historical monthly mean temperatures (red color) with simulated uncorrected temperatures (blue color) and corrected temperatures (black color) in the reference period (1971-2000) for seven climate models (m1 − CNRM-CM5-CCLM4-8-17, m2 − EC-EARTH-CCLM4-8-17, m3 − EC-EARTH-HIRHAM5, m4 − EC-EARTH-RACMO22E, m5 − EC-EARTH-RCA4, m6 − MPI-ESM-LR-CCLM4-8-17, m7 − MPI-ESM-LR -SMHI-RCA4) for two catchments located in Poland, Dunajec (upper panel) and Narewka (lower panel).

the performance of the methods for a test period. This comparison and evaluation of the need for computational resources formed the basis for the selection of bias correction methods to be used for the 1×1 km^2 gridded data set. Although the study concluded that distribution-based methods can perform better, practical limitations with respect to computing time led to the use of empirical quantile mapping for national work in Norway. In this project, however, bias correction was performed relative to spatially-averaged values of precipitation and temperature for the eight Norwegian catchments, and this bias correction was performed using both empirical quantile mapping and the use of a double gamma function. In this way, direct comparisons could be made between the results for both Poland and Norway.

6. PRECIPITATION AND TEMPERATURE CHANGES IN NEAR AND FAR FUTURE IN THE POLISH AND NORWEGIAN CATCHMENTS

A comparison of the change in annual precipitation in three periods: 1971-2000, 2021-2050, and 2071-2100, in ten Polish catchments is presented in
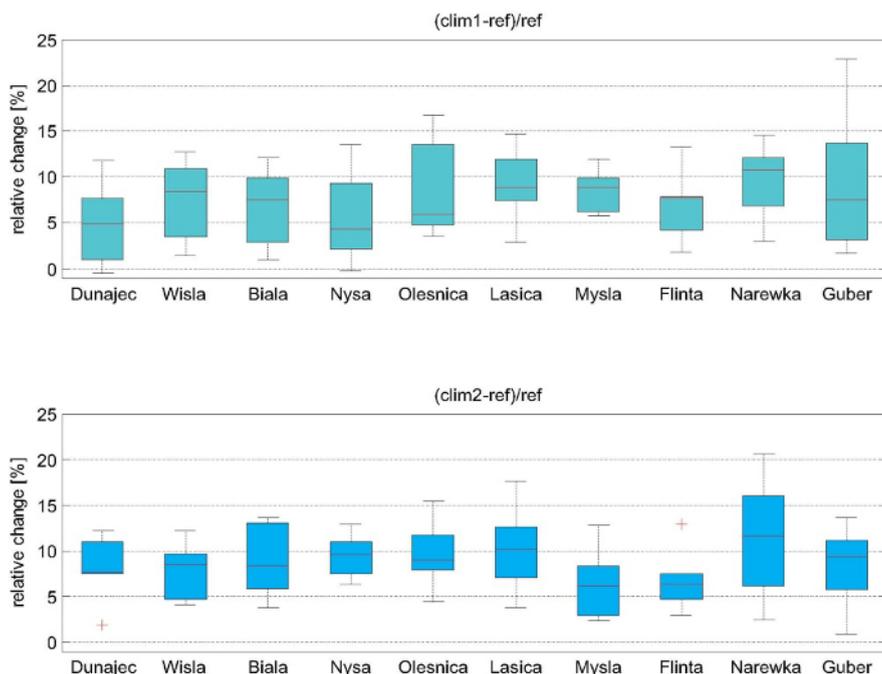
Fig. 6a. A comparison of the relative changes in annual precipitation totals in two future periods: 2021-2050 – upper panel, and 2071-2100 – lower panel *versus* reference period for Polish catchments. In each box, the central red mark denotes a median from seven climate model simulations, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers, and outliers are plotted individually in the form of red crosses.
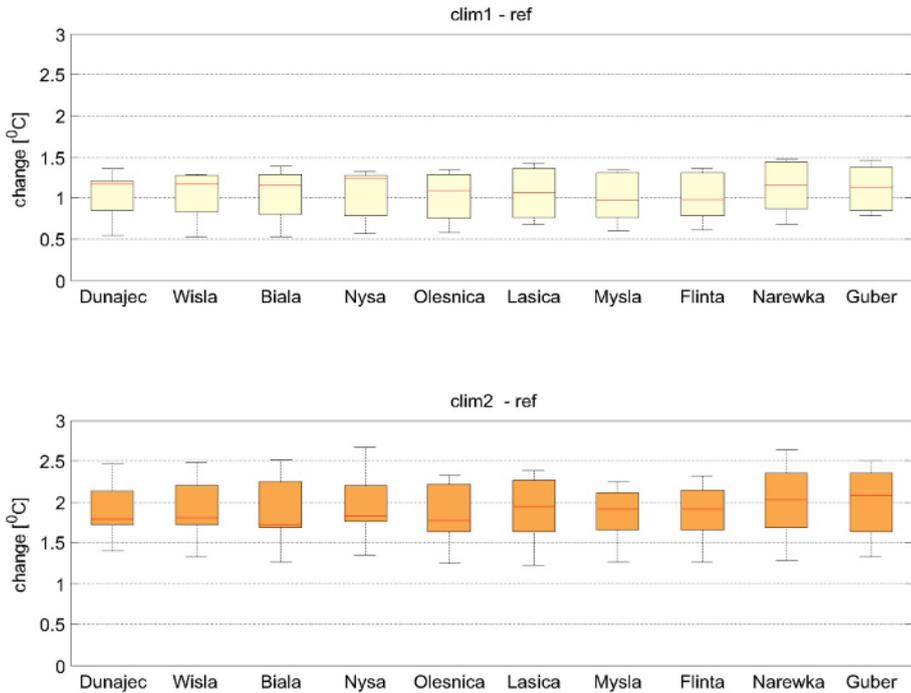
Fig. 6b. A comparison of the absolute changes in annual mean temperature [°C] in two future periods: 2021-2050 – upper panel, and 2071-2100 – lower panel *versus* reference period for Polish catchments. In each box, the central red mark denotes a median from seven climate model simulations, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers, and outliers are plotted individually in the form of red crosses.

Figs. 6a and b for emission scenario RCP4.5. Relative changes in annual precipitation in two future periods as compared with the reference period are shown in Fig. 6a. The estimated tendency of changes depends on the analysed period and catchment. In the case of an annual change between the near future and the reference periods, the climate model simulations project an increase in the annual precipitation sums of less than 10% for all catchments for the near future, except Narewka. The tendency of changes between the far future and the reference periods is also similar for all catchments, but relative changes are larger than 10% for Łasica and Narewka catchments. A similar analysis was performed for changes in air temperature. In that case the results are presented in the form of absolute changes expressed in °C (Fig. 6b). In all cases, an increase in air temperature is projected but there are

differences in the magnitude of the changes between catchments. An analysis of the tendency of changes in mean annual air temperature between the two future and the reference periods indicates an increase of about 1°C for the near future (2021-2050) and 2°C for the far future (2071-2100) periods in comparison with the reference period 1971-2000.

A comparison of the change in annual precipitation and temperature in the two periods 2021-2050 and 2071-2100 relative to 1971-2000, for eight Norwegian catchments is presented in Figs. 7a (precipitation) and 7b (temperature).

For the Norwegian catchments, changes in annual precipitation are positive, but smaller than for the Polish catchments. For the near-future, changes exceed 5% in 5 catchments (Eggedal, Futsavatn, Krinsvatn, Myglevatn, Polmak, Viksvatn), whilst in the far-future changes are over 10% in Eggedal, Krinsvatn, and Polmak. The latter is the northernmost catchment in this study and is located north of the Arctic Circle. Interestingly, Polmak is also the catchment where the highest temperature changes occurred both in the near- and far-future.
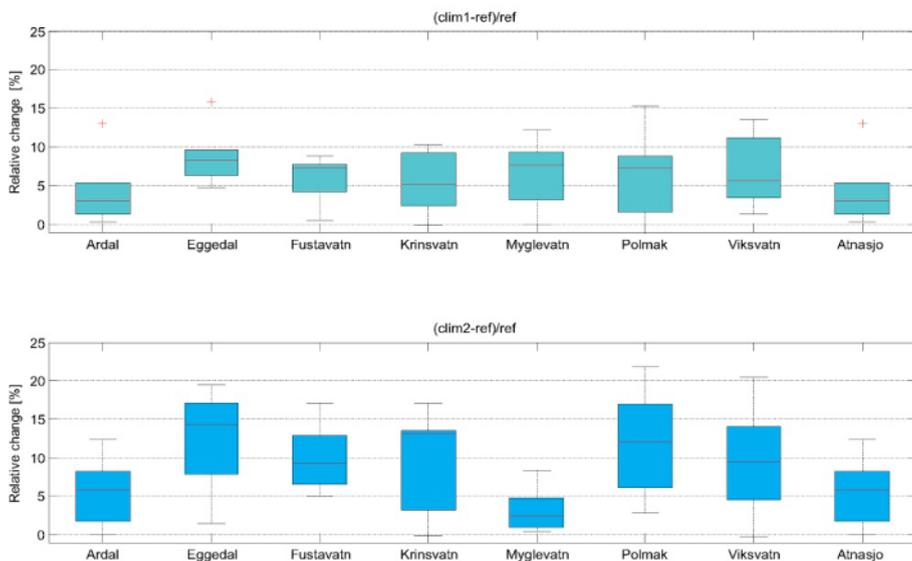
Fig. 7a. A comparison of the relative changes in the annual sum of precipitation in two future periods (2021-2050 and 2071-2100) *versus* the reference period for the Norwegian catchments. In each box, the central red mark denotes a median from seven climate model simulations, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers, and outliers are plotted individually in the form of red crosses.
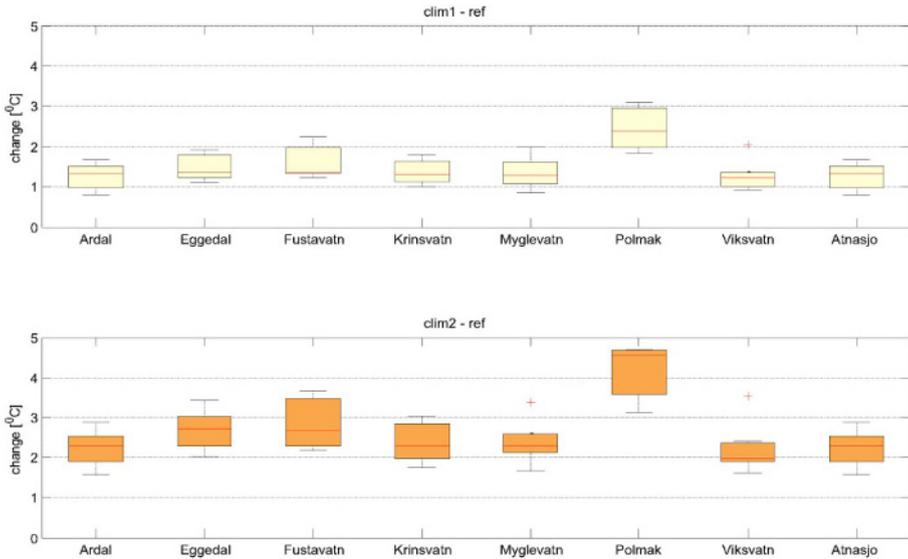
Fig. 7b. A comparison of the absolute changes in annual mean temperature [°C] in two future periods (2021-2050 and 2071-2100) *versus* the reference period for the Norwegian catchments. In each box, the central red mark denotes a median from seven climate model simulations, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers, and outliers are plotted individually in the form of red crosses.

7. CONCLUSIONS

This paper presents the background, objectives, and preliminary outcomes of the first year of activities of the project Climate Change Impact on Hydrological Extremes. Climate change affects temperature and precipitation, which act as driving forces for rainfall-runoff processes on a catchment scale. However, a catchment is affected not only by natural events, but also by human-induced changes. Therefore, in order to minimise other anthropogenic impacts on hydrological extremes, the choice of catchments for a climate impact study should follow well thought-through rules for catchment selection.

The methodology of catchment selection used here can be applied in similar studies. We list the main criteria that should be followed, of which the most important are the length of the observation series, their quality and homogeneity, the smallest number of human induced factors such as reservoirs, and river regulations. The degree of correlation between rainfall and catchment discharge measured by its modelling ability could also assist in the selection of catchments. Additional important criteria related to the na-

ture of the project are the spatial locations of the catchments and their flood regimes. Based on these criteria, a project database was formed consisting of ten nearly-natural Polish and eight nearly-natural Norwegian catchments constructed for the purpose of climate impact assessment. This is the first study establishing a foundation for a Reference Hydrologic Networks (RHN) database in Poland.

As a climatic forcing indicator, we have used the newest source of climate projections available from the EURO-CORDEX initiative. An "intermediate pathway" RCP4.5 was presented in this paper, as an example of a medium high emission scenario. The RCP8.5 pathway is also being evaluated in the project. Seven RCM/GCMs were selected to provide daily projections of climatic variables up to the year 2100 at a resolution of 12.5 km. The numerical tools were developed to efficiently handle those very large amounts of data that had to be downloaded and processed.

A comparison of simulations with observations in the reference period can be performed only by comparing distributions, as climate projections represent possible realisations of climate variables and not historical events. The comparison shows that those distributions differ, with larger discrepancies for precipitation than for temperature data. Two variants of the *QM* method were chosen to correct the biases in the RCM simulations. These were a theoretically-based double gamma function (Yang *et al.* 2010) for quantile mapping for precipitation and empirical quantile mapping for the residuals of air temperature.

A comparison of changes of projected temperature and precipitation series in two future periods was performed, a near-future period spanning from 2021-2050 and a far-future period 2071-2100. The results show increases in precipitation and temperature for the study catchments in both countries.

The results reaffirm the critical nature of the uncertainties involved in projections of future catchment meteorological forcing; these will have implications in the assessment of impact of climate changes on hydrological extremes.

Ongoing work in the project is evaluating low and high flow indices in the selected catchments, based both on observations and on projections to the end of the 21st century. The analysis of trends of future changes in hydrological extremes will be incorporated into uncertainty and adaptation studies as outlined by Doroszkiewicz and Romanowicz (2014).

within statutory activities No. 3841/E-41/S/2016 of the Ministry of Science and Higher Education of Poland. The hydro-meteorological data for Polish catchments were provided by the Institute of Meteorology and Water Management (IMGW), Poland.

R e f e r e n c e s

Astrup, E. (2000), Homogeneity testing of hydrological data, NVE Rapport 7-2000, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway (in Danish).

Beldring, S., T. Engen-Skaugen, E.J. Førland, and L.A. Roald (2008), Climate change impacts on hydrological processes in Norway based on two methods for transferring regional climate model results to meteorological station sites, *Tellus A* **60**, 3, 439-450, DOI: 10.1111/j.1600-0870.2008. 00306.x.

Berg, P., H. Feldmann, and H.-J. Panitz (2012), Bias correction of high resolution regional climate model data, *J. Hydrol.* **448-449**, 80-92, DOI: 10.1016/ j.jhydrol.2012.04.026.

Bergström, S., B. Carlsson, M. Gardelin, G. Lindström, A. Pettersson, and M. Rummukainen (2001), Climate change impacts on runoff in Sweden – assessments by global climate models, dynamical downscaling and hydrological modelling, *Clim. Res.* **16**, 101-112, DOI: 10.3354/cr016101.

Bergström, S., J. Andréasson, and L.P. Graham (2012), Climate adaptation of the Swedish guidelines for design floods for dams. **In:** *Proc. 24th ICOLD Congress on Large Dams, 6-8 June 2012, Kyoto, Japan*, Q94.

Cancelliere, A., and J.D. Salas (2004), Drought length properties for periodic-stochastic hydrologic data, *Water Resour. Res.* **40**, 2, W02503, DOI: 10.1029/2002WR001750.

Chung, C., and J.D. Salas (2000), Drought occurrence probabilities and risks of dependent hydrologic processes, *J. Hydrol. Eng.* **5**, 3, 259-268, DOI: 10.1061/(ASCE)1084-0699(2000)5:3(259).

Clarke, L.E., J.A. Edmonds, H.D. Jacoby, H.M. Pitcher, J.M. Reilly, and R.G. Richels (2007), Scenarios of greenhouse gas emissions and atmospheric concentrations, Sub-report 2.1a, U.S. Climate Change Science Program. Synthesis and Assessment Product 2.1a, CCSP, Washington, 154 pp.

Cloke, H.L., F. Wetterhall, Y. He, J.E. Freer, and F. Pappenberger (2013), Modelling climate impact on floods with ensemble climate projections. *Quart. J. Roy. Meteorol. Soc.* **139**, 671, 282-297, DOI: 10.1002/qj.1998.

Doroszkiewicz, J., and R. Romanowicz (2014), WP4 internal report on identifying flood management measures and constructions, overview of policy and law

for adaptation process in Poland, Report CHIHE/WP4/I, IGP PAS, Warsaw, Poland.

Dyrrdal, A.V., K. Isaksen, H.O. Hygen, and N.K. Meyer (2012), Changes in meteorological variables that can trigger natural hazards in Norway, *Clim. Res*. **55**, 2, 153-165, DOI: 10.3354/cr01125.

Edenhofer, O., B. Knopf, T. Barker, L. Baumstark, E. Bellevrat, B. Chateau, P. Criqui, M. Isaac, A. Kitous, S. Kypreos, M. Leimbach, K. Lessmann, B. Magné, S. Scrieciu, H. Turton, and D.P. van Vuuren (2010), The economics of low stabilization: model comparison of mitigation strategies and costs, *The Energy J.* **31**, 1, 11-48, DOI: 10.5547/ISSN0195-6574-EJ-Vol31-NoSI-2.

Edvardsen, S.-M., and C.M. Roald (2012), Flood zone map Naustdal, NVE Rapport 23/2012, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway (in Norwegian).

EEA (2014), Corine Land Cover 2006 raster data, European Environment Agency, http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster.

Favre, A.C., S. El Adlouni, L. Perreault, N. Thiémonge, and B. Bobée (2004), Multivariate hydrological frequency analysis using copulas, *Water Resour. Res*. **40**, 1, W01101, DOI: 10.1029/2003WR002456.

Fernández, B., and J.D. Salas (1999a), Return period and risk of hydrologic events: I: Mathematical formulation, *J. Hydrol. Eng.* **4**, 4, 297-307, DOI: 10.1061/(ASCE)1084-0699(1999)4:4(297).

Fernández, B., and J.D. Salas (1999b), Return period and risk of hydrologic events: II. Applications, *J. Hydrol. Eng*. **4**, 4, 308-316, DOI: 10.1061/(ASCE)1084-0699(1999)4:4(308).

Fleig, A.K., L.M. Andreassen, E. Barfod, J. Haga, L.E. Haugen, H. Hisdal, K. Melvold, and T. Saloranta (2013) Norwegian Hydrological Reference Dataset for climate change studies, NVE Report 2-2013, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway.

Gudmundsson, L., J.B. Bremnes, J.E. Haugen, and T. Engen-Skaugen (2012), Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, *Hydrol. Earth Syst. Sci.* **16**, 3383-3390, DOI: 10.5194/hess-16-3383-2012.

Gustard, A., and S. Demuth (eds.) (2009), Manual on low-flow estimation and prediction, Operational Hydrology Report No. 50,WMO-No. 1029, World Meteorological Organization, Geneva, Switzerland.

Hannaford, J., and T. Marsh (2006), An assessment of trends in UK runoff and low flows using a network of undisturbed catchments, *Int. J. Climatol*. **26**, 9, 1237-1253, DOI: 10.1002/joc.1303.

Hannaford, J., and T.J. Marsh (2008), High-flow and flood trends in a network of undisturbed catchments in the UK, *Int. J. Climatol*. **28**, 10, 1325-1338, DOI: 10.1002/joc.1643.

Hannah, D.M., S. Demuth, H.A.J. van Lanen, U. Looser, C. Prudhomme, G. Rees, K. Stahl, and L.M. Tallaksen (2011), Large-scale river flow archives: importance, current status and future needs, *Hydrol. Process*. **25**, 7, 1191-1200, DOI: 10.1002/hyp.7794.

Hisdal, H., and L.M. Tallaksen (2003), Estimation of regional meteorological and hydrological drought characteristics: a case study for Denmark, *J. Hydrol.* **281**, 3, 230-247, DOI: 10.1016/S0022-1694(03)00233-6.

IOŚ–PIB (2013), Strategic adaptation plan for areas vulnerable for climate change 2020 prespective, Institute of Environmental Protection – National Research Institute (IOŚ–PIB), Warszawa, Poland (in Polish).

IPCC (2007), *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, New York.

IPCC (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T.F. Stocker *et al.* (eds.), Cambridge University Press, New York, 1535 pp.

Jacob, D., J. Petersen, B. Eggert, A. Alias, O.B. Christensen, L.M. Bouwer, A. Braun, A. Colette, M. Déqué, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Kröner, S. Kotlarski, A. Kriegsmann, E. Martin, E. van Meijgaard, C. Moselely, S. Pfeifer, S. Preuschmann, C. Rader-macher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J.-F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber, and P. Yiou (2014), EURO-CORDEX: new high-resolution climate change projections for European impact research, *Reg. Environ. Change* **14**, 2, 563-578, DOI: 10.1007/s10113-013-0499-2.

Kaczmarek, Z. (2003), The impact of climate variability on flood risk in Poland, *Risk Anal.* **23**, 3, 559-566, DOI: 10.1111/1539-6924.00336.

Kaczmarek, Z., and J. Napiórkowski (1996), Water resources adaptation strategy in an uncertain environment. **In:** J.B. Smith, N. Bhatti, G.V. Menzhulin, R. Benioff, M. Campos, B. Jallow, F. Rijsberman, M.I. Budyko, and R.K. Dixon (eds.), *Adapting to Climate Change*, Springer, New York, 211-224, DOI: 10.1007/978-1-4613-8471-7_18.

Kaczmarek, Z., J. Napiórkowski, and K.M. Strzepek (1996), Climate change impacts on the water supply system in the Warta river catchment, Poland, *Int. J. Water Resour. Dev.* **12**, 2, 165-180, DOI: 10.1080/07900629650041939.

Kotlarski, S., K. Keuler, O.B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen, D. Jacob, D. Lüthi, E. van Meijgaard, G. Nikulin, C. Schär, C. Teichmann, R. Vautard, K. Warrach-Sagi, and V. Wulfmeyer (2014), Regional climate modeling on European scales: a joint standard evaluation

of the EURO-CORDEX RCM ensemble, *Geosci. Model Dev.* **7**, 4, 1297-1333, DOI: 10.5194/gmd-7-1297-2014.

Kriaučiūnienė, J., D. Meilutyte-Barauskienė, E. Rimkus, J. Kažys, and J. Vincevičius (2008), Climate change impact on hydrological processes in Lithuanian Nemunas river basin, *Baltica* **21**, 1-2, 51-61.

Kundzewicz, Z.W., D. Graczyk, T. Mauer, I. Pińskwar, M. Radziejewski, C. Svensson, and M. Szwed (2005), Trend detection in river flow series: 1. Annual maximum flow, *Hydrol. Sci. J.* **50**, 5, 797-810, DOI: 10.1623/hysj.2005.50.5.797.

Kundzewicz, Z.W., I. Pińskwar, and G.R. Brakenridge (2013), Large floods in Europe, 1985-2009, *Hydrol. Sci. J.* **58**, 1, 1-7, DOI: 10.1080/02626667.2012.745082.

Lafon, T., S. Dadson, G. Buys, and C. Prudhomme (2013), Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods, *Int. J. Climatol.* **33**, 6, 1367-1381, DOI: 10.1002/joc.3518.

Lawrence, D., and I. Haddeland (2011), Uncertainty in hydrological modelling of climate change impacts in four Norwegian catchments, *Hydrol. Res*. **42**, 6, 457-471, DOI: 10.2166/nh.2011.010.

Lawrence, D., and H. Hisdal (2011), Hydrological projections for floods in Norway under a future climate, NVE Report 5-2011, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway.

Lawrence, D., I. Haddeland, and E. Langsholt (2009), Calibration of HBV hydrological models using PEST parameter estimation, NVE Report 1-2009, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway, 44 pp.

Li, C.Z., L. Zhang, H. Wang, Y.Q. Zhang, F.L. Yu, and D.H. Yan (2012), The transferability of hydrological models under nonstationary climatic conditions, *Hydrol. Earth Syst. Sci.* **16**, 1239-1254, DOI: 10.5194/hess-16-1239-2012.

Marsh, T. (2010), The UK Benchmark Network – designation, evolution and application. **In:** *Fifth Int. Conf. on Water Resources and Environment Research, 5-7 July 2010, Quebec City, Canada.*

Mishra, A.K., and P. Coulibaly (2010), Hydrometric network evaluation for Canadian watersheds, *J. Hydrol*. **380**, 3-4, 420-437, DOI: 10.1016/j.jhydrol.2009.11.015.

Mishra, A.K., and V.P. Singh (2011), Drought modeling – A review, *J. Hydrol*. **403**, 1-2, 157-175, DOI: 10.1016/j.jhydrol.2011.03.049.

Monk, W.A., D.L. Peters, R.A. Curry, and D.J. Baird (2011), Quantifying trends in indicator hydroecological variables for regime-based groups of Canadian rivers, *Hydrol. Process*. **25**, 19, 3086-3100, DOI: 10.1002/hyp.8137.

Moss, R.H., J.A. Edmonds, K.A. Hibbard, M.R. Manning, S.K. Rose, D.P. van Vuuren, T.R. Carter, S. Emori, M. Kainuma, T. Kram, G.A. Meehl,

J.F.B. Mitchell, N. Nakicenovic, K. Riahi, S.J. Smith, R.J. Stouffer, A.M. Thomson, J.P. Weyant, and T.J. Wilbanks (2010), The next generation of scenarios for climate change research and assessment, *Nature* **463**, 747-756, DOI: 10.1038/nature08823.

Nakicenovic, N., J. Alcamo, G. Davis, B. De Vries, J. Fenhann, S. Gaffin, K. Gregory, A. Grübler, T.Y. Jung, T. Kram, E.L. La Rovere, L. Michaelis, S. Mori, T. Morita, W. Pepper, H. Pitcher, L. Price, K. Raihi, A. Roehrl, H.-H. Rogner, A Sankovski, M. Schlesinger, P. Shukla, S. Smith, R. Swart, S. van Rooijen, N. Victor, and Z. Dadi (2000), *Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, 599 pp.

Nash, J.E., and J.V. Sutcliffe (1970), River flow forecasting through conceptual models. Part I – A discussion of principles, *J. Hydrol.* **10**, 3, 282-290, DOI: 10.1016/0022-1694(70)90255-6.

Osuch, M., J. Kindler, R.J. Romanowicz, K. Berbeka, and A. Banrowska (2012), Strategia adaptacji Polski do zmian klimatu w zakresie sektora "Zasoby i gospodarka wodna", KLIMADA Project, Institute of Environmental Protection, National Research Institute, Warsaw, Poland (in Polish).

Osuch, M., R.J. Romanowicz, D. Lawrence, and W.K. Wong (2015), Assessment of the influence of bias correction on meteorological drought projections for Poland, *Hydrol. Earth Syst. Sci. Discuss*. **12**, 10331-10377, DOI: 10.5194/hessd-12-10331-2015.

Piani, C., J.O. Haerter, and E. Coppola (2010), Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.* **99**, 1, 187-192, DOI: 10.1007/s00704-009-0134-9.

Raje, D., and P.P. Mujumdar (2010), Reservoir performance under uncertainty in hydrologic impacts of climate change, *Adv. Water Resour.* **33**, 3, 312-326, DOI: 10.1016/j.advwatres.2009.12.008.

Reitan, T., and A. Petersen-Øverleir (2005), Evaluating the homogeneity of hydrological time series with the help of Bayesian regression, NVE Report 5-2005, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway (in Norwegian).

Romanowicz, R.J., and M. Osuch (2011), Assessment of land use and water management induced changes in flow regime of the Upper Narew, *Phys. Chem. Earth* **36**, 13, 662-672, DOI: 10.1016/j.pce.2011.04.012.

Romanowicz, R.J., M. Osuch, and M. Grabowiecka (2013), On the choice of calibration periods and objective functions: A practical guide to model parameter identification, *Acta Geophys*. **61**, 6, 1477-1503, DOI: 10.2478/s11600-013-0157-6.

Sadowski, M. (2008), An approach to adaptation to climate changes in Poland, *Climate Change* **90**, 4, 443-451, DOI: 10.1007/s10584-008-9394-0.

Salvadori, G., and C. De Michele (2004), Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, *Water Resour. Res.* **40**, 12, W12511, DOI: 10.1029/2004WR003133.

Shiau, J., and H. Shen (2001), Recurrence analysis of hydrologic droughts of differing severity, *J. Water Resour. Plan. Manage* **127**, 1, 30-40, DOI: 10.1061/(ASCE)0733-9496(2001)127:1(30).

Slack, J.R., and J.M. Landwehr (1992), Hydro-climatic data network (HCDN); A U.S. Geological Survey streamflow data set for the United States for the study of climatic variations, 1874-1988, Open-file Report 92-129, U. S. Geol. Survey, Reston, USA.

Sorteberg, A., I. Haddeland, J.E. Haugen, S. Sobolowski, and W.K. Wong (2014), Evaluation of distribution mapping based bias correction methods, Norwegian Centre for Climate Services, NCCS Report 1/2014, 23 pp.

Stahl, K., H. Hisdal, J. Hannaford, L.M. Tallaksen, H.A.J. van Lanen, E. Sauquet, S. Demuth, M. Fendekova, and J. Jódar (2010), Streamflow trends in Europe: evidence from a dataset of near-natural catchments, *Hydrol. Earth Syst. Sci.* **14**, 2367-2382, DOI: 10.5194/hess-14-2367-2010.

Storn, R., and K. Price (1997), Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.* **11**, 4, 341-359, DOI: 10.1023/A:1008202821328.

Strupczewski, W.G., K. Kochanek, E. Bogdanowicz, W. Feluch, and I. Markiewicz (2012), *A Two-Level Method of Estimation of Non-stationary Flood Quantiles*, Monografie Komitetu Gospodarki Wodnej PAN, Warszawa, No. 35, 109-124 (in Polish).

Strupczewski, W.G., E. Bogdanowicz, and S. Debele (2015), The estimation of trends in quantiles by means of selected pdf fitted to seasonal, annual or partial duration (named also Peaks Over Threshold – POT) flow series, Report CHIHE/WP3/I, IGP PAS, Warsaw, Poland.

Teng, J., J. Vaze, F.H.S. Chiew, B. Wang, and J.-M. Perraud (2012), Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff, *J. Hydrometeorol.* **13**, 1, 122-139, DOI: 10.1175/JHM-D-11-058.1.

Thorne, J.H., R. Boynton, L. Flint, A. Flint, and T.-N. Le (2010), Development and application of downscaled hydroclimatic predictor variables for use in climate vulnerability and assessment studies, Publication No. CEC-500-2012-010, California Energy Commission, Davis, USA.

Tveito, O.E., I. Bjørdal, A.O. Skjelvåg, and B. Aune (2005), A GIS-based agro-ecological decision system based on gridded climatology, *Meteorol. Appl.* **12**, 1, 57-68, DOI: 10.1017/S1350482705001490.

van der Linden, P., and J.F.B. Mitchell (2009), ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, Exeter, UK.

van Vuuren, D.P., P.L. Lucas, and H. Hilderink (2007), Downscaling drivers of global environmental change: Enabling use of global SRES scenarios at the national and grid levels, *Global Environ. Change* **17**, 1, 114-130, DOI: 10.1016/j.gloenvcha.2006.04.004.

Veijalainen, N., E. Lotsari, P. Alho, B. Vehviläinen, and J. Käyhkö (2010), National scale assessment of climate change impacts on flooding in Finland, *J. Hydrol.* **391**, 3-4, 333-350, DOI: 10.1016/j.jhydrol.2010.07.035.

Whitfield, P.H., D.H. Burn, J. Hannaford, H. Higgins, G.A. Hodgkins, T. Marsh, and U. Looser (2012), Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends, *Hydrol. Sci. J.* **57**, 8, 1562-1579, DOI: 10.1080/02626667. 2012.728706.

Wilson, D., H. Hisdal, and D. Lawrence (2010), Has streamflow changed in the Nordic countries? – Recent trends and comparisons to hydrological projections, *J. Hydrol.* **394**, 3-4, 334-346, DOI: 0.1016/j.jhydrol.2010. 09.010.

Wong, W.K., S. Beldring, T. Engen-Skaugen, I. Haddeland, and H. Hisdal (2011), Climate change effects on spatiotemporal patterns of hydroclimatological summer droughts in Norway, *J. Hydrometeorol.* **12**, 6, 1205-1220, DOI: 10.1175/2011JHM1357.1.

Yang, W., J. Andréasson, L.P. Graham, J. Olsson, J. Rosberg, and F. Wetterhall (2010), Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, *Hydrol. Res.* **41**, 3-4, 211-229, DOI: 10.2166/nh.2010.004.

Changes in the Large-Scale Atmospheric Circulation over Romania Between 1961 and 2010 on Seasonal Basis

Nicu BARBU[1,2], Cristina BURADA[3], Sabina STEFAN[1], and Florinela GEORGESCU[2]

[1]University of Bucharest, Faculty of Physics, Magurele, Bucharest, Romania
e-mail: nicubarbu1982@yahoo.com (corresponding author)

[2]National Meteorological Administration, Bucharest, Romania

[3]National Meteorological Administration – Regional Meteorological Center, Craiova, Romania

A b s t r a c t

The aim of this paper is to investigate the trends and shifts of the circulation types over Romania for 50-year period (1961-2010) on seasonal basis. In order to achieve this, two objective catalogues, namely GWT and WLK, from COST733 Action were employed. Daily circulation types were grouped according to the cyclonicity and anticyclonicity and were used to calculate the seasonal occurrence frequency of cyclonic and anticyclonic types. The trend of seasonal time series was investigated by using Mann–Kendall test and the shifts points were determined by using Pettitt test. The results reveal that the occurrence frequency of anticyclonic types increases in summer and winter seasons and the occurrence frequency of cyclonic ones decreases for the summer season (for alpha = 0.05).

Key words: circulation types, GWT, WLK, Mann–Kendall test, Pettitt test.

1. INTRODUCTION

The atmospheric circulation is the principal factor that determines the climate variability in Europe and therefore in Romania. A variety of subjective or objective methodological approaches were applied for atmo-spheric circulation classification (Huth *et al.* 2008, Philipp *et al.* 2010). The subjective (so-called manual) classification of atmospheric circulations (Hess and Brezowski 1952, Lamb 1972, Perret 1987) has no longer been used for atmospheric circulation classification; however it has not lost its scientific importance (Yarnal *et al.* 2001).

Changes in the atmospheric circulation over the North Atlantic European region have been studied in the last decades by several authors (Bárdossy and Caspary 1990, Jacobeit *et al.* 2001) employing various methods for different periods of time and most of them have noted changes in circulation type frequencies. A few recent studies have reported considerable changes in the 1980s and 1990s in the persistence of air circulation over Europe for zonal circulation in winter (Werner *et al.* 2000) and for all circulation types in summer (Kyselý 2002). One important application of circulation type classification is to quantify past, present, and future in the atmospheric circulation in terms of frequency and persistence (Huth *et al.* 2008).

In Romania, Georgescu and Stefan (2010) have investigated the link between cyclonic activity over Romania and air circulation types by using LWT2 catalogue. According to the same study the anticyclonic circulation types seem to be the most frequent and cyclonic activities are correlated to the eastern and southeastern circulation types.

The objective of this study is to analyze the temporal variability of air circulation over Romania on seasonal basis. In order to achieve this, two objective classifications, namely Grosswetter-Types (GWT) and WetterLagen-Klassifikation (WLK) of atmospheric circulation from COST733Action for a domain centered over Romania, were employed. The trend of the circulation types is analyzed by using Mann–Kendall test and the Pettitt test is used to investigate the change point in the mean circulation types.

This paper is organized as follows: In Section 2 the data sets used for determining circulation types and statistical tests are introduced. Relevant results and discussions are presented in Section 3. Section 4 is dedicated to conclusions.

2. DATA AND METHODS

Re-analysis data sets for 50 years period (1961-2010), built up by the NCEP/NCAR (Kalnay *et al.* 1996) were used with a spatial resolution of 2.5° (latitude) by 2.5° (longitude) and consists of daily mean sea level pressure, geopotential high at 925 and 500 hPa, u and v component of the wind vector at 700 hPa and precipitable water content for the entire atmosphere column.
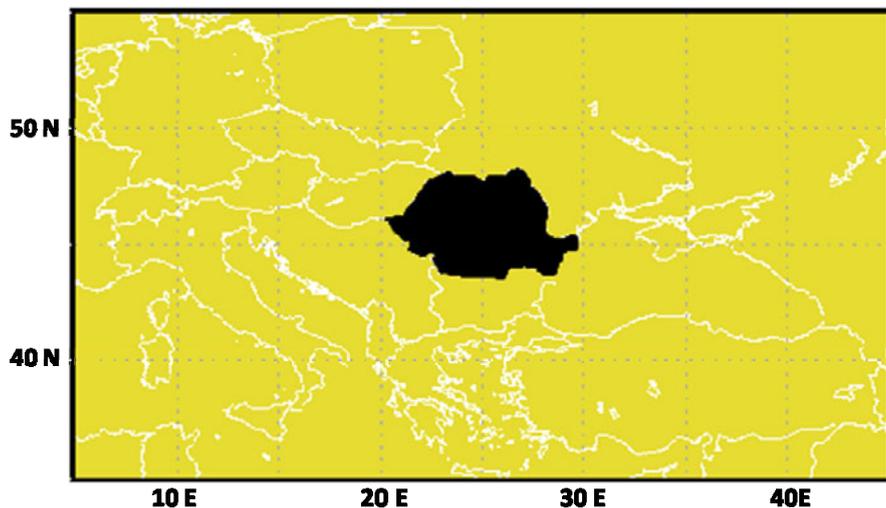
Fig. 1. Spatial domain used to determine the circulation types. Romania is highlighted.

For the large-scale air circulation classification, two objective catalogues, namely GWT and WLK from the COST733Action, were employed. The spatial domain used (Fig. 1) for building up the GWT and WLK catalogues is delimited by 5° E, 35° N and 45° E, 55° N, and includes Romania in its center (Barbu *et al.* 2014).

The GWT Catalogue uses as its input data set the sea level pressure (SLP) field (Beck *et al.* 2007), and it is based on the idea of characterizing the air circulation by the coefficients of zonality (Z), meridionality (M), and vorticity (V).

The WLK catalogue is based on the classification of weather types developed by Dittmann *et al.* (1995) and Bissolli and Dittmann (2004) originally including 40 different types. This catalogue uses information from three basic tropospheric levels: 925, 700, and 500 hPa, and information of the precipitable water content over the entire tropospheric column.

For the present study, we have been used the GWT Catalogue with 18 circulation types and the WLK Catalogue with 40 circulation types.

The trend of air circulations was investigated by using Mann–Kendall (Mann 1945, Kendall 1975) statistical method. Mann–Kendall test is a statistical test widely used for the analysis of trend in climatology (Mavromatis and Stathis 2011). More theoretical details about this test can be found in a lot of studies (*e.g.*, Salarijazi *et al.* 2012, Marin *et al.* 2014).

In this study, the non-parametric approach developed by Pettitt (1979) in order to detect change-points in the circulation types was employed. It is a

rank-based and distribution-free test for detecting a significant change in the mean of a time series and it is particularly useful when no hypothesis is required to be made about the location of the change point. More theoretical details about Pettitt test can be found in the study of Salarijazi *et al.* (2012).

3. RESULTS AND DISCUSSION

For the evaluation of air circulation over Romania, 18 262 cases that correspond to daily circulation types between 1 January 1961 and 31 December 2010, were analyzed. First, the seasonal occurrence frequency of circulation types was computed and thereafter the values obtained were averaged over the entire study period.

Figure 2 presents the occurrence frequency of GWT (left side) and WLK (right side) circulation types grouped by cyclonicity and anticyclonicity for each season over the entire study period. It can be seen that for the GWT circulation types, the occurrence frequency of anticyclonic circulations is higher than that of the cyclonic ones for all seasons, except of winter when cyclonic circulations were more frequent (around 58%). It is well known that during summer the anticyclonic activity is pronounced and in winter the cyclonic activity is more frequent. In the transition seasons (spring and autumn), since the generation of baroclinic instabilities determines the rapid succession of baric systems related to a decrease of meridional temperature gradient reduces the persistence of the atmospheric circulation (Geng and Sugi 2003) and leads to small differences between the occurrence frequency of cyclonic and anticyclonic circulation types. For the WLK circulation types the highest occurrence frequency of cyclonic types was obtained for spring (around 45%) and the highest occurrence frequency of anticyclonic types was obtained for autumn (around 50%).
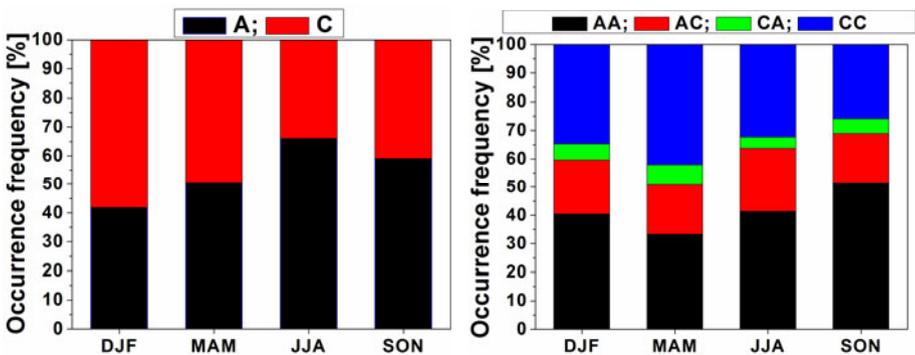

Fig. 2. Occurrence frequency of GWT (left side) and WLK (right side) circulation types for all seasons (winter – DJF, spring – MAM, summer – JJA, and autumn – SON). Please see the text for the explanation circulation type's abbreviation.
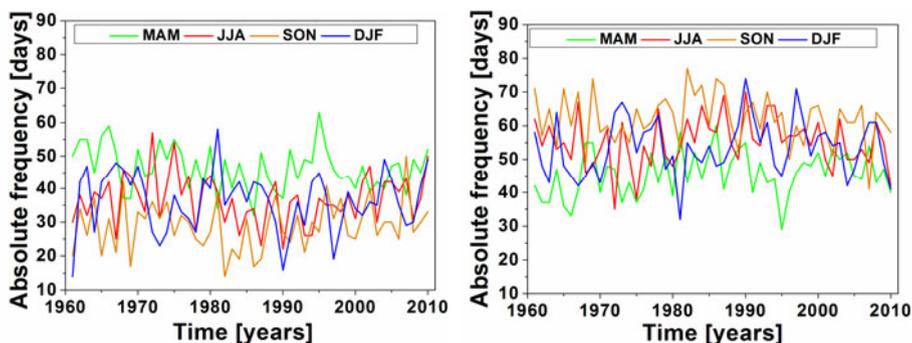
Fig. 3. Time series of absolute frequency of GWT circulation types grouped due to the cyclonicity (C – left side) and anticyclonicity (A – right side) for spring (MAM – green), summer (JJA – red), autumn (SON – orange), and winter (DJF – blue).

In order to investigate the temporal variations for each season, the circulation types of GWT Catalogue were grouped by cyclonicity (Fig. 3 left side) and anticyclonicity (Fig. 3 right side) thereby resulting in two time series. After that, the Mann–Kendall trend test and Pettitt change point test are applied to these two time series. The results of both tests are not statistically significant, at 0.05 significance level. Also, the Pettitt test indicates that there is no shift in both time series.

For the investigation of temporal variations of WLK Catalogue, circulation types were grouped by cyclonicity (C) and anticyclonicity (A) at 925 and 500 hPa, thereby resulting in four time series as follows: AA – anticyclone at 925 and 500 hPa; AC – anticyclone at 925 hPa, and cyclone at 500 hPa; CA − cyclone at 925 hPa, and anticyclone at 500 hPa; CC – cyclone at 925 and 500 hPa. In Figure 4, the absolute frequency of these four circulation types mentioned above, for seasonal time scales are presented.

The trend of the time series was determined by using the Mann–Kendall test and Pettitt test was applied in order to detect shifts in the time series. In Table 1 the Mann–Kendall statistics of the time series grouped by cyclonicity (C) or anticyclonicity (A) at 925 and 500 hPa that are statistically significant at the 0.05 level are presented (the up arrow indicates upward trend and the down arrow indicates downward trend). From Table 1 it can be seen that the anticyclonic regime increases for 925 and 500 hPa levels in summer (JJA) and winter (DJF), while cyclonic regime decreases for the same geopotential levels in summer (JJA). This result is in agreement with those obtained by Wang *et al.* (2013). According to this study, the occurrence frequency of cyclones increases in the high latitude North Atlantic in cold seasons, with decreases in the mid-latitude North Atlantic – southern Europe.
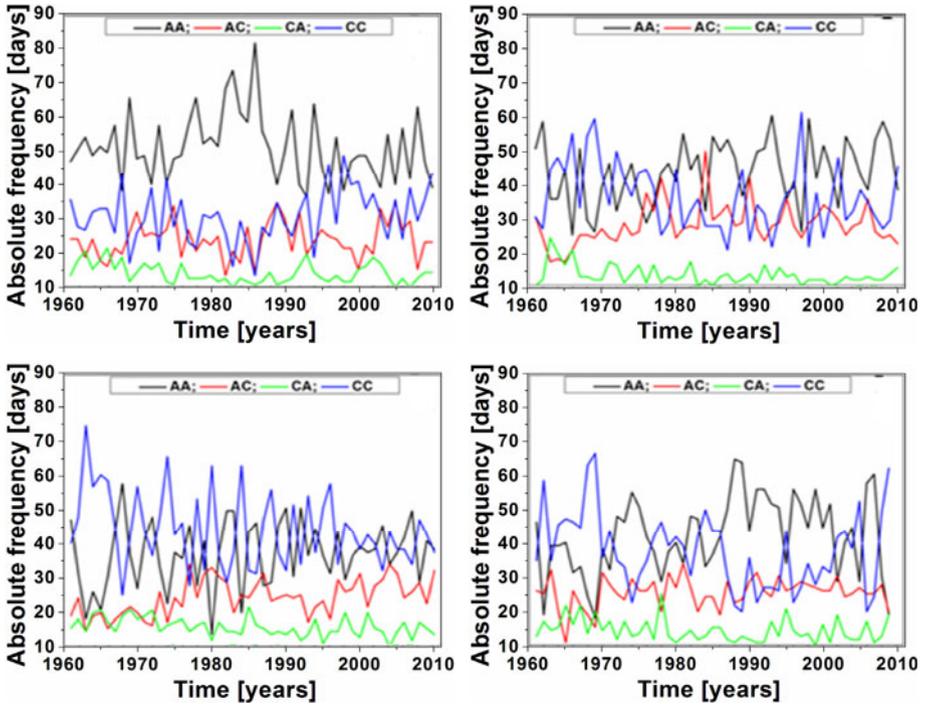
Fig. 4. Time series of absolute frequency of WLK circulation types grouped by cyclonicity (C – left side) and anticyclonicity (A – right side) at 925 and 500 hPa levels for spring (left upper panel), summer (right upper panel), autumn (left lower panel), and winter (right lower panel). Please see the text for the explanation circulation type's abbreviation.

Table 1

The trend of the time series grouped
by cyclonicity (C) or anticyclonicity (A) at 925 and 500 hPa

Type	MAM	JJA	SON	DJF
AA	–	↑	–	↑
AC	↑	↑	–	–
CA	↓	–	–	↓
CC	–	↓	–	–

Explanations: the up arrow indicates upward trend, and the down arrow indicates downward trend; only the results that are significant at 0.05 level are presented.

Increasing frequency of anticyclonic types found in this study can be linked to increasing frequency of regional climate extremes. For example, concerning the study conducted by Birsan and Dumitrescu (2014) related to

climate change in Romania, the upward trend found for summer mean temperature can be linked to the increasing frequency of anticyclonic types. According to Barbu *et al.* (2014) Azores Ridge and North-African Ridge are the main synoptic patterns that produce prolonged and very warm days in Romania.

Table 2

The shifts in the time series grouped
by cyclonicity (C) or anticyclonicity (A) at 925 and 500 hPa

Type	MAM	JJA	SON	DJF
AA	–	1981↑	1992↑	1987↑
AC	1977↑	1976↑	–	–
CA	1980↓	–	1964↓	1979↓
CC	–	1978↓	1995↓	1987↓

Note: only the results that are significant at 0.05 level are presented.

In Table 2 the shift in the time series grouped by the cyclonicity (C) or anticyclonicity (A) at 925 and 500 hPa that are statistically significant at the 0.05 level are presented (the number represents the year of change point, the up arrow indicates an upward shift, and the down arrow indicates a downward shift). In order to find a plausible explanation for the shift in the time series in '80s, we have taken a look of the North Atlantic Oscillation (NAO – station based) time series (Hurrell and Deser 2010). By applying Pettitt test to the annual NAO was found an upward shift (significant for alpha = 0.1) around 1982, and this may be the reason for the increasing frequency of anticyclonic types over Romania, compensated by decreasing frequency of cyclonic ones.

4. SUMMARY AND CONCLUDING REMARKS

In this study, we presented a statistical analysis of daily air circulation types over Romania during the period 1961-2010 for the lower and middle troposphere. From our knowledge, this is the first study of the temporal variations of air circulation types at the synoptic scale over Romania using objective methods. The study of the circulation types for lower and middle troposphere is useful because, on the one hand, in the lower troposphere the trajectory of air masses is modified by the orography, in our case the Carpathians, which are very important for Romanian's climate, and, on the other hand, for the mid-troposphere the orography no longer influences the air flow direction from this level. We have investigated the trend of the time series of both, GWT (with 18 circulation types) and WLK (with 40 circulation types) catalogues, using Mann–Kendall test and shifts in the time series

using Pettitt test. Both above-mentioned tests were applied to the circulation types grouped by cyclonicity (C) or anticyclonicity (A) for the entire 50 year period of the study, during 1961-2010.

The most important changes in the occurrence frequency of the anticyclonic and cyclonic types were noted for summer, autumn, and winter seasons. For these seasons, according to the WLK Catalogue, a positive trend was found for the anticyclonic types and negative trends for the cyclonic ones. These results are consistent with those previously reported that highlights that the occurrence frequency of anticyclonic types at mid-latitudes increases (Kyselý 2002, Jacobeit *et al.* 2003). The general negative trend of the cyclonic types is due to the decrease in mid-latitude cyclone frequency and increase in high-latitude cyclone frequency (McCabe *et al.* 2001). In the context of climate change, the Mediterranean cyclones have a decreasing trend of occurrence frequency (Nissen *et al.* 2014). The extension of the Hadley cell under global warming (Kang and Lu 2012) may cause a northern shift of storm tracks over North Hemisphere mid-latitude (Paciorek *et al.* 2002, Zhang *et al.* 2004).

The present study is very important due to the changes in occurrence frequency of circulation types that can be linked to the increasing of climatic extreme events (*e.g.*, heat waves, droughts, blizzards, floods, and flash floods).

References

Barbu, N., F. Georgescu, V. Ştefanescu, and S. Ştefan (2014), Large-scale mechanisms responsible for heat waves occurrence in Romania, *Rom. J. Phys.* **59**, 9-10, 1109-1126.

Bárdossy, A., and H.J. Caspary (1990), Detection of climate change in Europe by analyzing European atmospheric circulation patterns from 1881 to 1989, *Theor. Appl. Climatol.* **42**, 3, 155-167, DOI: 10.1007/BF00866871.

Beck, C., J. Jacobeit, and P.D. Jones (2007), Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in Central Europe since 1780, *Int. J. Climatol*. **27**, 4, 473-491, DOI: 10.1002/joc.1410.

Birsan, M.V., and A. Dumitrescu (2014), Snow variability in Romania in connection to large-scale atmospheric circulation, *Int. J. Climatol*. **34**, 1, 134-144, DOI: 10.1002/joc.3671.

Bissolli, P., and E. Dittmann (2004), Objective weather types. **In:** *Klimastatusbericht 2003*, Deutscher Wetterdienst, Offenbach, 101-107 (in German).

Dittmann, E., S. Barth, J.G. Lang, and J.G. Müller-Westermeier (1995), Objective weather type classification, Tech. Report 197, Deutscher Wetterdienst, Offenbach (in German).

Geng, Q., and M. Sugi (2003), Possible change of extratropical cyclone activity due to enhanced greenhouse gases and sulfate aerosols – study with a high-resolution AGCM, *J. Climate* **16**, 13, 2262-2274, DOI: 10.1175/1520-0442(2003)16<2262:PCOECA>2.0.CO;2.

Georgescu, F., and S. Stefan (2010), Cyclonic activity over Romania in connection with the air circulation types, *Rom. Rep. Phys*. **62**, 4, 878-886.

Hess, P., and H. Brezowski (1952), *Catalog of the European Large-Scale Weather Types*, Tech. Report 33, Deutscher Wetterdienst, Offenbach, 39 pp.

Hurrell, J.W., and C. Deser (2010), North Atlantic climate variability: The role of the North Atlantic Oscillation, *J. Marine Syst*. **79**, 3-4, 231-244, DOI: 10.1016/j.jmarsys.2009.11.002.

Huth, R., C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kyselý, and O.E. Tveito (2008), Classifications of atmospheric circulation patterns, *Ann. N. Y. Acad. Sci.* **1146**, 105-152, DOI: 10.1196/annals.1446.019.

Jacobeit, J., P. Jönsson, L. Bärring, C. Beck, and M. Ekström (2001), Zonal indices for Europe 1780-1995 and running correlations with temperature, *Climatic Change* **48**, 1, 219-241, DOI: 10.1023/A:1005619023045.

Jacobeit, J., R. Glaser, J. Luterbacher, and H. Wanner (2003), Links between flood events in central Europe since AD 1500 and large-scale atmospheric circulation modes, *Geophys. Res. Lett*. **30**, 4, 1172, DOI: 10.1029/2002GL016433.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph (1996), The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Am. Meteorol. Soc.* **77**, 3, 437-471, DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kang, S.M., and J. Lu (2012), Expansion of the Hadley cell under global warming: winter versus summer, *J. Climate* **25**, 24, 8387-8393, DOI: 10.1175/JCLI-D-12-00323.1.

Kendall, M.G. (1975), *Rank Correlation Methods*, 4th ed., Charles Griffin, London.

Kyselý, J. (2002), Temporal fluctuations in heat waves at Prague-Klementinum, the Czech Republic, from 1901-1997, and their relationships to atmospheric circulation, *Int. J. Climatol.* **22**, 1, 33-50, DOI: 10.1002/joc.720.

Lamb, H.H. (1972), *British Isles Weather Types and Register of Daily Sequence of Circulation Patterns, 1861-1971*, Geophysical Memoir, Vol. 116, Her Majesty's Stationery Office, London, 85 pp.

Mann, H.B. (1945), Non-parametric tests against trend, *Econometrica* **13**, 3, 245-259, DOI: 10.2307/1907187.

Marin, L., M.-V. Birsan, R. Bojariu, A. Dumitrescu, D.M. Micu, and A. Manea (2014), An overview of annual climatic changes in Romania: trends in air temperature, precipitation, sunshine hours, cloud cover, relative humidity and wind speed during the 1961-2013 period, *Carpath. J. Earth Environ. Sci.*. **9**, 4, 253-258.

Mavromatis, T., and D. Stathis (2011), Response of the water balance in Greece to temperature and precipitation trends, *Theor. Appl. Climatol.* **104**, 1-2, 13-24, DOI: 10.1007/s00704-010-0320-9.

McCabe, G.J., M.P. Clark, and M.C. Serreze (2001), Trends in Northern Hemisphere surface cyclone frequency and intensity, *J. Climate* **14**, 12, 2763-2768, DOI: 10.1175/1520-0442(2001)014<2763:TINHSC>2.0.CO;2.

Nissen, K.M., G.C. Leckebusch, J.G. Pinto, and U. Ulbrich (2014), Mediterranean cyclones and windstorms in a changing climate, *Reg. Environ. Change* **14**, **5**, 1873-1890, DOI: 10.1007/s10113-012-0400-8.

Paciorek, C.J., J.S. Risbey, V. Ventura, and R.D. Rosen (2002), Multiple indices of northern hemisphere cyclone activity, winters 1949-99, *J. Climate* **15**, 13, 1573-1590, DOI: 10.1175/1520-0442(2002)015<1573:MIONHC>2.0.CO;2.

Perret, R. (1987), *A Classification of Meteorological Situations for Use in Prediction*, Institut Suisse de Meteorologie, Zürich (in French).

Pettitt, A.N. (1979), A non-parametric approach to the change-point problem, *Appl. Statist.* **28**, 2, 126-136, DOI: 10.2307/2346729.

Philipp, A., J. Bartholy, C. Beck, M. Erpicum, P. Esteban, X. Fettweis, R. Huth, P. James, S. Jourdain, F. Kreienkamp, T. Krennert, S. Lykoudis, S.C. Michalides, K. Pianko-Kluczynska, P. Post, D. Rasilla Álvarez, R. Schiemann, A. Spekat, and F.S. Tymvios (2010), COST733cat – A database of weather and circulation type classifications, *Phys. Chem. Earth A/B/C* **35**, 9-12, 360-373, DOI: 10.1016/j.pce.2009.12.010.

Salarijazi, M., A.-M. Akhond-Ali, A. Adib, and A. Daneshkhah (2012), Trend and change-point detection for the annual stream-flow series of the Karun River at the Ahvaz hydrometric station, *Afr. J. Agric. Res.* **7**, 32, 4540-4552, DOI: 10.5897/AJAR12.650.

Wang, X.L.,Y. Feng, G.P. Compo, V.R. Swail, F.W. Zwiers, R.J. Allan, and P.D. Sardeshmukh (2013), Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of twentieth century reanalysis, *Clim. Dynam*. **40**, 11-12, 2775-2800, DOI: 10.1007/s00382-012-1450-9.

Werner, P.C., F.-W. Gerstengarbe, K. Fraedrich, and H. Oesterle (2000), Recent climate change in the North Atlantic/European sector, *Int. J. Climatol*. **20**, 5, 463-471, DOI: 10.1002/(SICI)1097-0088(200004)20:5<463::AID-JOC 483>3.0.CO;2-T.

Yarnal, B., A.C. Comrie, B. Frakes, and D.P. Brown (2001), Developments and prospects in synoptic climatology, *Int. J. Climatol*. **21**, 15, 1923-1950, DOI: 10.1002/joc.675.

Zhang, X.D., J.E. Walsh, J. Zhang, U.S. Bhatt, and M. Ikeda (2004), Climatology and interannual variability of arctic cyclone activity: 1948-2002, *J. Climate* **17**, 12, 2300-2317, DOI: 10.1175/1520-0442(2004)017<2300:CAIVOA> 2.0.CO;2.

Application of Gravity Gradients
in the Process of GOCE Orbit Determination

Andrzej BOBOJĆ

University of Warmia and Mazury in Olsztyn,
Department of Astronomy and Geodynamics, Olsztyn, Poland
e-mail: andrzej.bobojc@gmail.com

A b s t r a c t

The possibility of improving the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission satellite orbit using gravity gradient observations was investigated. The orbit improvement is performed by a dedicated software package, called the Orbital Computation System (OCS), which is based on the classical least squares method. The corrections to the initial satellite state vector components are estimated in an iterative process, using dynamic models describing gravitational perturbations. An important component implemented in the OCS package is the 8th order Cowell numerical integration procedure, which directly generates the satellite orbit. Taking into account the real and simulated GOCE gravity gradients, different variants of the solution of the orbit improvement process were obtained. The improved orbits were compared to the GOCE reference orbits (Precise Science Orbits for the GOCE satellite provided by the European Space Agency) using the root mean squares (RMS) of the differences between the satellite positions in these orbits. The comparison between the improved orbits and the reference orbits was performed with respect to the inertial reference frame (IRF) at J2000.0 epoch. The RMS values for the solutions based on the real gravity gradient measurements are at a level of hundreds of kilometers and more. This means that orbit improvement using the real gravity gradients is ineffective. However, all solutions using simulated gravity gradients have RMS values below the threshold determined by the RMS

values for the computed orbits (without the improvement). The most promising results were achieved when short orbital arcs with lengths up to tens of minutes were improved. For these short arcs, the RMS values reach the level of centimeters, which is close to the accuracy of the Precise Science Orbit for the GOCE satellite. Additional research has provided requirements for efficient orbit improvement in terms of the accuracy and spectral content of the measured gravity gradients.

Key words: GOCE satellite orbit, gravity gradients.

1. INTRODUCTION

The Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission terminated in November 2013 was dedicated to the study of the Earth's gravitational field. For the first time, a satellite with a gradiometer on board was successfully placed in an almost circular orbit with an inclination of 96.7° and a mean altitude of around 254.9 km (Rummel *et al.* 2009). The gradiometer, consisting of three pairs of identical accelerometers, forming three arms, each around 50 cm long, provided the measurements of the second spatial derivatives of the geopotential, also known as gravity gradients, or gravity gradient tensor (GGT) measurements. The second instrument, besides the gradiometer installed on board the GOCE satellite, was a Global Positioning System (GPS) receiver, which was the source of code and phase observations of the GPS satellite constellation, called Satellite-to-Satellite Tracking high-low mode observations (SST-hl) or simply GPS observations (ESA 1999). Both mentioned types of measurements, collected during the four years of the GOCE mission, are the basic data serving to implement the main scientific aim, *i.e.*, the estimation of new models of the Earth's static gravity field. These models in the form of sets of coefficients of geopotential expansion into a spherical harmonic series will find application in a number of sciences such as solid Earth physics, oceanography, and geodesy (Johannessen *et al.* 2003). An important factor in the determination of these models is the geolocation of observation data, performed by satellite orbit estimation. This estimation is carried out using SST-hl measurements, which are composed of the above-mentioned dual-frequency code and phase observations with respect to GPS satellites.

The determined GOCE satellite orbit occurs in two versions. The first version is based on parameters describing the field of forces determining the satellite's motion, including, first of all, the gravity field model. Orbit determination starts with the preparation of data on GPS satellites: orbital information and their clock corrections. Parameters describing Earth rotation are also taken into account (Bock *et al.* 2007, 2011). The approximate orbit is generated in the first stage using pseudo-range measurements. This orbit is

described by six osculating Keplerian elements at the initial epoch and nine empirical parameters modeling the constant and periodically changing additional satellite acceleration. The approximate orbit serves to synchronize the clock of the GPS receiver onboard the satellite with the GPS time system for all observation epochs. In the next step, the aforementioned orbit is improved by means of an iterative procedure using phase measurements with a 30-s sampling interval. This time, the obtained orbit has a richer parameterization; the above-mentioned six osculating Keplerian elements at the initial epoch and nine empirical parameters are additionally supplemented by three components of the pseudo-stochastic acceleration vector, estimated for successive 15-min intervals (Bock *et al.* 2007). The next stage is another orbit improvement using phase observations with an even higher time resolution than before, which is now at 10 s. The final orbit, obtained after an appropriate number of iterations, is defined by the six osculating elements at the initial epoch, three empirical constants of satellite acceleration in the radial direction, along the direction of motion and in the direction perpendicular to the instantaneous plane of the orbit, respectively, and satellite accelerations constant at successive 6-min intervals along the directions listed above. The GPS measurements mentioned here are non-differentiated observations of the carrier phase, hence the need for additional determination of the ambiguities of the total number of cycles and the receiver clock corrections. The obtained satellite orbit is presented as a time series of position and velocity components at 10-s intervals relative to the International Terrestrial Reference Frame 2005 (ITRF2005) reference frame (ESA 2010). The first version of the GOCE orbit described above is referred to as the reduced-dynamic orbit because it partly depends on the estimated pseudo-stochastic vectors of the satellite's empirical acceleration. For the second version of the GOCE orbit, also called the kinematic orbit, the determination is based solely on geometric data (code and phase measurements); no parameters characterizing the dynamic state of the satellite are taken into account. As a result, only the satellite position vectors for successive epochs are determined together with the ambiguity of the total number of cycles and the receiver clock corrections. The final solution is a time series of the satellite's coordinates in the ITRF2005 reference frame with 1-s time resolution (Bock *et al.* 2007).

To assess external accuracy, both versions of the GOCE orbit were compared with the orbit determined based on the Satellite Laser Ranging (SLR) measurements. The obtained differences show the accuracy of determining the reduced-dynamic orbit and the kinematic orbit at a level of a few centimeters (Bock *et al.* 2011).

As shown by the above description, the process of determining both versions of the GOCE satellite orbit is based on the use of GPS observations, while the second type of measurements from the GOCE mission (gravity

gradients) finds application in the estimation of gravity field models. However, gravity gradients as the second spatial derivatives of the geopotential also contain information about the satellite's position. The possibility of using this information was noted by Rummel and Colombo (1985), where the corrections to the spherical harmonic coefficients were determined together with the corrections to the initial satellite state vector components. Bobojć and Drożyner (2003) showed the possibility of using the gravity gradient tensor components to improve simulated low Earth-orbiting (LEO) satellite orbits. An interesting idea of the application of dynamical aspect of gravity gradients for a satellite orbit determination is given by Eshagh (2014). The satellite acceleration vector is derived from gravity gradients in order to determine an orbit by the numerical integration (Eshagh 2014).

The aim of this paper is to examine the possibility and conditions of efficient GOCE satellite orbit determination using the geometrical information contained in gravity gradient observations.

2. THEORETICAL BASIS

The satellite's motion in the Earth's gravity field can be described by an equation in the following general form (Palacios *et al.* 1992):

$$\ddot{\boldsymbol{r}} = -\frac{GM}{r^3} r + f \quad , \tag{1}$$

where *r* and $\ddot{\boldsymbol{r}}$ are the satellite position and acceleration vectors, respectively, *r* is the distance between the satellite and the Earth's mass center, GM is the product of the gravity constant and Earth's mass, *f* is the perturbing acceleration vector. The first term of the right-hand side of Eq. 1 represents the satellite's Keplerian motion in the central gravity field, the other takes into account the effect of perturbing forces, which can be divided into two basic groups. The first of them includes gravitational perturbations generated by the geopotential, Earth tides, ocean tides, the Earth's rotational deformation caused by changes in the angular velocity vector, the third body effect and the relativity effects. The other group contains non-gravitational perturbations: atmospheric drag, solar radiation pressure, the albedo effect, and the effect of magnetic forces (Eshagh and Najafi-Alamdari 2007). The resultant perturbing acceleration vector *f* can also be composed (besides accelerations resulting from the above-mentioned forces) of pseudo-stochastic empirical acceleration vectors constant at specific time intervals, which were estimated during the determination of the reduced-dynamic orbit in the case of the GOCE satellite (Bock *et al.* 2011).

It is obvious that the dominant effect on the satellite's motion is exerted by the gravity field, whose potential *V* is expressed by (Heiskanen and Moritz 1967):

$$V(r,\theta,\lambda) = W(r,\theta,\lambda) - Z(r,\theta,\lambda)$$

$$= \frac{GM}{r} \sum_{n=0}^{N_{max}} \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} \left(\overline{C}_{nm} \cos m\lambda + \overline{S}_{nm} \sin m\lambda\right) \overline{P}_{nm}(\cos\theta) \ . \qquad (2)$$

In Eq. 2 r, θ, and λ are the geocentric coordinates of a given point; r is the distance from Earth's center, $\theta = 90° - \varphi$ is the colatitude and φ the geocentric latitude, λ is the geocentric longitude, $W(r, \theta, \lambda)$ is the gravity potential containing centrifugal potential, $Z(r, \theta, \lambda)$ is the centrifugal potential, a is the equatorial radius of the Earth ellipsoid, $\overline{C}_{nm}$, $\overline{S}_{nm}$ are the spherical harmonic coefficients (Stokes' coefficients) of degree n ($n = 0, 1, ..., N_{max}$; N_{max} is maximum degree of the spherical harmonic expansion) and order m ($m = 0, 1, ..., n$), and $\overline{P}_{nm}(\cos\theta)$ is the normalized associated Legendre function of degree n and order m.

As already noted, the gravity gradients V_{ij} ($i, j = 1, 2, 3$) provided by the GOCE mission are the second spatial derivatives of the geopotential V. Hence, one can write

$$V_{ij} = \frac{\partial^2 V(r,\theta,\lambda)}{\partial x_j \partial x_i} \ , \qquad (3)$$

where the geocentric coordinates r, θ, λ are functions of the Cartesian coordinates x_i, x_j ($i, j = 1, 2, 3$ and $x_1 = x$, $x_2 = y$, $x_3 = z$) in the adopted Earth fixed reference frame. The gravity gradients form a second-order tensor field. This tensor, also called the gravity gradient tensor (GGT), or the Marussi tensor, assumes the form (Rummel *et al.* 2011)

$$\boldsymbol{M} = \begin{bmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{yx} & V_{yy} & V_{yz} \\ V_{zx} & V_{zy} & V_{zz} \end{bmatrix} , \qquad (4)$$

where, for example, $V_{xy} = \partial^2 V / \partial y \partial x$. The matrix $\boldsymbol{M}$ is symmetric and traceless, satisfying in the space external to the Earth the Laplace equation

$$V_{xx} + V_{yy} + V_{zz} = 0 \ . \qquad (5)$$

Taking into account the form of Eq. 2, which describes the geopotential, dependence on the coordinates x, y, z can be found also for the gravity gradients V_{ij}, forming the tensor $\boldsymbol{M}$. Therefore, V_{ij} measurements performed by the gradiometer onboard the satellite contain usable orbital information.

3.   GOCE SATELLITE ORBIT IMPROVEMENT

Assuming as the data the gravity gradient tensor components, the satellite orbit determination process can be based on the following observation equation:

$$V_{ijk}^o + v_{ijk} = V_{ijk}^c + \frac{\partial V_{ijk}^c}{\partial \mathbf{r}} \frac{\partial \mathbf{r}}{\partial (\mathbf{r}_o, \dot{\mathbf{r}}_o)} \begin{bmatrix} \Delta \mathbf{r}_o \\ \Delta \dot{\mathbf{r}}_o \end{bmatrix} . \tag{6}$$

In Eq. 6:

V_{ijk}^o, V_{ijk}^c – observed and computed gravity gradient at epoch k,

$(i, j = x, y, z)$, respectively,

v_{ijk} – correction to the observed q of V_{ijk}^o,

$\dfrac{\partial V_{ijk}^c}{\partial \mathbf{r}}$ – partial derivative of the computed gravity gradient V_{ijk}^c with respect

to the satellite position vector $\mathbf{r} = [x, y, z]^T$ at epoch k,

$\mathbf{r}_o, \dot{\mathbf{r}}_o$ – initial satellite position vector and initial satellite velocity vector,

respectively, at initial epoch t_o; $\mathbf{r}_o = [x_o, y_o, z_o]^T$ and $\dot{\mathbf{r}}_o = [\dot{x}_o, \dot{y}_o, \dot{z}_o]^T$,

$\dfrac{\partial \mathbf{r}}{\partial (\mathbf{r}_o, \dot{\mathbf{r}}_o)}$ – partial derivative of the position vector with respect to the initial

state vector $\mathbf{p}_o = [\mathbf{r}_o, \dot{\mathbf{r}}_o]^T$ interpolated at epoch k,

$\Delta \mathbf{r}_o, \Delta \dot{\mathbf{r}}_o$ – vectors of corrections to the initial satellite position vector and the initial satellite velocity vector; $\Delta \mathbf{r}_o = [\Delta x_o, \Delta y_o, \Delta z_o]^T$ and $\Delta \mathbf{r}_o = [\Delta \dot{x}_o, \Delta \dot{y}_o, \Delta \dot{z}_o]^T$.

The gravity gradient V_{ijk}^c at epoch k from Eq. 6 is determined as the second-order partial derivative of geopotential (Eq. 2) with respect to the rectangular coordinates x, y, and z of the satellite in the Earth-fixed reference frame. On the other hand, the quantity $\partial V_{ijk}^c / \partial \mathbf{r}$ at epoch k, consisting of the components $\partial V_{ijk}^c / \partial x$, $\partial V_{ijk}^c / \partial y$, $\partial V_{ijk}^c / \partial z$, is computed as the third-order spatial derivatives of the geopotential, taking into account the satellite position in the Earth-fixed reference frame. These derivatives form the third-order tensor. The detailed formulas used for the computation of V_{ijk}^c and $\partial V_{ijk}^c / \partial \mathbf{r}$ are taken from Metris *et al.* (1999), where the problem of computing the spatial derivatives of the geopotential of any order was solved. The partial derivative of the position vector with respect to the initial state vector, occurring also in Eq. 6, forms the following matrix:

$$\frac{\partial \boldsymbol{r}}{\partial \left(\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o \right)} = \begin{bmatrix} \dfrac{\partial x}{\partial x_o} & \dfrac{\partial x}{\partial y_o} & \dfrac{\partial x}{\partial z_o} & \dfrac{\partial x}{\partial \dot{x}_o} & \dfrac{\partial x}{\partial \dot{y}_o} & \dfrac{\partial x}{\partial \dot{z}_o} \\[2mm] \dfrac{\partial y}{\partial x_o} & \dfrac{\partial y}{\partial y_o} & \dfrac{\partial y}{\partial z_o} & \dfrac{\partial y}{\partial \dot{x}_o} & \dfrac{\partial y}{\partial \dot{y}_o} & \dfrac{\partial y}{\partial \dot{z}_o} \\[2mm] \dfrac{\partial z}{\partial x_o} & \dfrac{\partial z}{\partial y_o} & \dfrac{\partial z}{\partial z_o} & \dfrac{\partial z}{\partial \dot{x}_o} & \dfrac{\partial z}{\partial \dot{y}_o} & \dfrac{\partial z}{\partial \dot{z}_o} \end{bmatrix}, \tag{7}$$

whose elements are propagated in the numerical integration of GOCE orbit using the 8th order Cowell method, starting from the initial values $\partial_i / \partial_j = \delta_{ij}$ ($i = x, y, z$; $j = x_o, y_o, z_o$; δ_{ij} – the Kronecker delta).

The application of the classical least squares method with an iterative approach leads to the determination of the vector $\Delta \boldsymbol{p}_{oi}$ of unknown corrections to the initial satellite state vector for the ith iteration using the following formula:

$$\Delta \boldsymbol{p}_{o_i} = \left[\Delta \boldsymbol{r}_{o_i}, \Delta \dot{\boldsymbol{r}}_{o_i} \right]^{\mathrm{T}} = \left(\mathbf{A}_i^{\mathrm{T}} \mathbf{A}_i \right)^{-1} \mathbf{A}_i^{\mathrm{T}} \boldsymbol{l}_i \,, \tag{8}$$

in which the matrix $\mathbf{A}_i$ contains coefficients of observation equations equal to the partial derivatives of the gravity gradients with respect to the initial state vector, and the components of the vector $\boldsymbol{l}_i$ are differences between the observed and computed quantities ("O"-"C"). Thus, the matrix $\mathbf{A}_i$ and the vector $\boldsymbol{l}_i$ can be presented as:

$$A_i = \left[\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n \right]^{\mathrm{T}}, \tag{9}$$

$$l_i = \left[l_1 l_2 \cdots l_n \right]^{\mathrm{T}}, \tag{10}$$

where submatrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$, and subvectors $l_1, l_2, \ldots, l_n$ correspond to the successive observation epochs, and n is the total numer of observation epochs. The submatrix $\mathbf{A}_k$ and subvector l_k at the given observation epoch k ($k = 1, 2, \ldots, n$) are expressed by:

$$\mathbf{A}_k = \begin{bmatrix} \dfrac{\partial V_{xx}^c}{\partial \boldsymbol{r}} \cdot \dfrac{\partial \boldsymbol{r}}{\partial \left(\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o \right)} \\[3mm] \dfrac{\partial V_{yy}^c}{\partial \boldsymbol{r}} \cdot \dfrac{\partial \boldsymbol{r}}{\partial \left(\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o \right)} \\[3mm] \dfrac{\partial V_{zz}^c}{\partial \boldsymbol{r}} \cdot \dfrac{\partial \boldsymbol{r}}{\partial \left(\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o \right)} \\[3mm] \dfrac{\partial V_{xy}^c}{\partial \boldsymbol{r}} \cdot \dfrac{\partial \boldsymbol{r}}{\partial \left(\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o \right)} \\[3mm] \dfrac{\partial V_{xz}^c}{\partial \boldsymbol{r}} \cdot \dfrac{\partial \boldsymbol{r}}{\partial \left(\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o \right)} \\[3mm] \dfrac{\partial V_{yz}^c}{\partial \boldsymbol{r}} \cdot \dfrac{\partial \boldsymbol{r}}{\partial \left(\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o \right)} \end{bmatrix}, \qquad l_k = \begin{bmatrix} V_{xx}^o - V_{xx}^c \\[2mm] V_{yy}^o - V_{yy}^c \\[2mm] V_{zz}^o - V_{zz}^c \\[2mm] V_{xy}^o - V_{xy}^c \\[2mm] V_{xz}^o - V_{xz}^c \\[2mm] V_{yz}^o - V_{yz}^c \end{bmatrix}. \tag{11}$$

For example:

$$\frac{\partial V_{xx}^c}{\partial \boldsymbol{r}} \frac{\partial \boldsymbol{r}}{\partial (\boldsymbol{r}_o, \dot{\boldsymbol{r}}_o)} = \left[\frac{\partial V_{xx}^c}{\partial x_o} \; \frac{\partial V_{xx}^c}{\partial y_o} \; \frac{\partial V_{xx}^c}{\partial z_o} \; \frac{\partial V_{xx}^c}{\partial \dot{x}_o} \; \frac{\partial V_{xx}^c}{\partial \dot{y}_o} \; \frac{\partial V_{xx}^c}{\partial \dot{z}_o} \right] . \tag{12}$$

Subsequent elements of the above expression can be presented analogously to the first element development, *i.e.*

$$\frac{\partial V_{xx}^c}{\partial x_o} = \frac{\partial V_{xx}^c}{\partial x} \frac{\partial x}{\partial x_o} + \frac{\partial V_{xx}^c}{\partial y} \frac{\partial y}{\partial x_o} + \frac{\partial V_{xx}^c}{\partial z} \frac{\partial z}{\partial x_o} . \tag{13}$$

In order to realize the aim of this work, the satellite orbit is improved by dedicated software called Orbital Computation System (OCS). It was created as an expansion of the Toruń Orbit Processor (TOP) package (Drożyner 1995), which determines an orbit in the field of gravitational and nongravitational forces. In the OCS package, as already mentioned, the corrections to the initial dynamic state vector are estimated in successive iterations using the gravity gradient observations by means of the least squares method (Eq. 8). This means that the initial state vector $\boldsymbol{p}_{o_i-1}$ determined in the previous iteration is corrected within the current iteration to its actual value $\boldsymbol{p}_{o_i}$, which can be written as

$$\boldsymbol{p}_{o_i} = \boldsymbol{p}_{o_{i-1}} + \Delta \boldsymbol{p}_{o_i} , \tag{14}$$

where $\Delta \boldsymbol{p}_{o_i}$ denotes the correction vector obtained in the current iteration. This iterative process ends after meeting the convergence criteria, taking into account the RMS difference between the observed gravity gradient V_{ijk}^o and the computed gravity gradient V_{ijk}^c for the previous iteration *i*-1 and the current iteration *i*. The convergence criteria are satisfied if

$$\left| \frac{\mathrm{RMS}_i - \mathrm{RMS}_{i-1}}{\mathrm{RMS}_i} \right| < 10^{-2} . \tag{15}$$

3.1 Algorithm for orbit determination

In the frame of the OCS package, the orbit improvement process in a given iteration *i* consists of the following three basic parts:

(1) The first part includes reading initial data, *i.e.*, the initial conditions − initial satellite state vector and options characterizing a given adjustment variant, the time series of gravity gradient observations, dynamic model parameters − coefficients of geopotential expansion into a spherical harmonic

series, tidal model coefficients, third body position vectors, optionally the parameters of the empirical acceleration model, the time series of the elements of orientation of the ITRF2005 and the GOCE satellite's gradiometer reference frame (GRF) (ESA 2008) relative to the Cartesian inertial reference frame (IRF), which is realized based on the mean equator and vernal equinox of the J2000.0 standard epoch (ESA 2008). The time series of gravity gradient measurements and the above-mentioned orientation elements were acquired from the GOCE mission through the European Space Agency (ESA).

(2) The second part leads to determine the normal matrix $(\mathbf{A}_i^\mathrm{T}\mathbf{A}_i)^{-1}$ and the vector $\mathbf{A}_i^\mathrm{T}\mathbf{l}_i$ (Eq. 8). It is a process that takes place in a closed loop containing a number of repetitions equal to the number of observation epochs. The given repetition starts with the determination of transformation matrices at the given observation epoch between the ITRF2005 and GRF reference frames and between the ITRF2005 and IRF reference frames. This determination uses the aforementioned orientation elements which are given in terms of quarternions (ESA 2010).

Next, the computation of the satellite state vector (position and velocity vector) for the measurement epoch using numerical integration of equations of motion by the 8th order Cowell method and position interpolation for the measurement epoch is performed. A mathematical model of the forces acting on the satellite is created for the given epoch during the computations (Drożyner 1995). It takes into account gravitational perturbing forces and additionally optionally empirical accelerations. This model includes accelerations generated by the geopotential, ocean tides, Earth tides, the third body effect, and the relativity effects. Non-gravitational forces were compensated by the Drag-Free and Attitude Control System (DFACS) installed onboard the GOCE satellite, in which ion propulsion played an important role (Rebhan *et al.* 2000). The geopotential is represented by the ITG-GRACE2010S model (Mayer-Gürr *et al.* 2010). In this place, the derivatives of the vector of position with respect to the components of the initial state vector are also numerically propagated and interpolated for the observation epoch.

In the following step of the described part, the *a priori* values of measurements of gravity gradients (the second-order tensor) and the components of the third-order tensor of spatial derivatives $\partial V_{ijk}^c / \partial \mathbf{r}$ are computed in the ITRF2005 Earth-fixed reference frame, taking into account the determined position vector. This vector was transformed into the ITRF2005 frame from the IRF frame by means of the corresponding transformation matrix. Then, the observed tensor of gravity gradients, the computed (*a priori*) tensor of

gravity gradients and the third-order tensor are transformed into the IRF reference frame, where the orbit improvement is performed.

The next step includes the calculation of residuals, *i.e.*, the differences between the observed and *a priori* values of gravity gradients at the given measurement epoch. All of the above calculations allow to determine the partial values of the normal matrix $(\mathbf{A}_i^\mathrm{T}\mathbf{A}_i)^{-1}$ and the vector $\mathbf{A}_i^\mathrm{T}\mathbf{l}_i$. These values are summed in the successive repetitions. After the last repetition, the full normal matrix $(\mathbf{A}_i^\mathrm{T}\mathbf{A}_i)^{-1}$ and the full vector $\mathbf{A}_i^\mathrm{T}\mathbf{l}_i$ are obtained.

(3) The third part includes the determination of correction vector to the initial satellite state vector $\Delta\boldsymbol{p}_{o_i} = \left[\Delta x_0, \Delta y_0, \Delta z_0, \Delta \dot{x}_0, \Delta \dot{y}_0, \Delta \dot{z}_0\right]^\mathrm{T}$. This makes it possible to improve the initial state vector in the iteration *i* using Eq. 14. In order to assess an internal accuracy of the orbit improvement process, RMS difference between the observed and computed (*a priori*) values of gravity gradient, the observation error of unit weight and standard deviations of corrected components of the vector $\boldsymbol{p}_{o_i}$ are estimated.

It should be noted that the *a priori* gravity gradients determined in the next iteration $i + 1$ are computed along the satellite orbit improved in the previous iteration *i*.

3.2 Parameter for the estimation of orbit determination accuracy

For all orbital arcs, the initial state vectors before starting the orbit improvement process were equal to the initial state vectors of the corresponding reference orbit arcs. The corrected initial satellite state vector $\boldsymbol{p}_o$ allows to obtain an finally improved satellite orbit, through the previously-mentioned numerical integration by the 8th order Cowell method using the dynamic models listed above. The improved GOCE orbit is then compared with the reference orbit, to determine the key parameter for the assessment of the quality of the obtained solution. This parameter, denoted as RMS$_{\mathrm{OUT}}$, characterizes the accuracy of the determined orbit. It is computed using the formula

$$\mathrm{RMS}_{\mathrm{OUT}} = \sqrt{\sum_{i=1}^{3}\left(\mathrm{RMS}_i\right)^2} \ , \tag{16}$$

with the quantities of $\mathrm{RMS}_i (i = 1, 2, 3)$ obtained using the following expression:

$$\mathrm{RMS}_i = \sqrt{\frac{\sum_{j=1}^{n}\left[\left(\mathrm{x}_i\right)_j - \left(x_i\right)_{j\,\mathrm{REF}}\right]^2}{n}} \ , \tag{17}$$

where $(x_i)_j$, $(x_i)_{j\,REF}$ ($i = 1, 2, 3$, $x_1 = x$, $x_2 = y$, $x_3 = z$) mean the satellite's Cartesian coordinates for epoch j in the inertial reference frame (IRF), in the improved orbit and the reference orbit, respectively, and n represents the to-tal number of epochs of the examined orbital arcs. The value of RMS_{OUT} can be interpreted as the average spatial distance between the points of both or-bits for the same epoch or as the measure of the fit of the determined orbit to the reference orbit. The reference orbit adopted for RMS_{OUT} parameter com-putation is provided by ESA as a Level 2 GOCE mission product in the pre-viously-mentioned versions: kinematic and reduced-dynamic. The accuracy of this orbit reaches the level of 2 cm (Bock *et al.* 2011). An important indi-cator in this aspect is also the quantity RMS_O, which corresponds to the RMS_{OUT} value for the approximate orbit, computed without the use of grav-ity gradient measurements. The approximate orbit is determined in the nu-merical integration process using appropriate dynamic models. The initial state vector of this orbit is equal to the initial state vector of the correspond-ing reference orbit. The quantity RMS_O is the threshold of orbit improve-ment effectiveness. The orbit improvement is effective when the obtained RMS_{OUT} value is less than the corresponding RMS_O value.

4. NUMERICAL TESTS – USING SIMULATED MEASUREMENTS

In order to check the possibility of GOCE orbit improvement using gravity gradient data, realistic simulations of the gravity gradient observations were first used. These simulations consisted in the computation of the gravity gra-dient values along the GOCE reference orbit. As already mentioned, this or-bit corresponds to the Precise Science Orbit (PSO) of the GOCE satellite, provided as an L2 product by the European Space Agency (ESA 2010). The gravity gradients were determined using the formulas derived by Metris *et al.* (1999) based on expansion into a spherical harmonic series. Originally computed in the ITRF2005 reference frame, the gravity gradients were then transformed to the IRF reference frame, using the time series of the elements of orientation of the ITRF2005 reference frame with respect to the inertial reference frame (IRF) (ESA 2010). The geopotential model ITG-GRACE2010S complete to the degree and order 180 of spherical harmonic coefficients was used for the computations.

The improvement subjects were three 1-day orbital arcs, denoted as vari-ants A, B, C, spanning the following epochs: 6 November 2009, 23:59:45.00 UTC – 7 November 2009, 23:59:45.00 UTC (variant A); 2 December 2009, 23:59:45.00 UTC – 3 December 2009, 23:59:45.00 UTC (variant B); 18 December 2009, 23:59:45.00 UTC – 19 December 2009, 23:59:45.00 UTC (variant C). The corresponding time series of gravity gra-dients is assigned to each of these variants. The initial state vectors undergo-ing the improvement process are equal to the initial state vectors of the

corresponding reference orbit arcs (Table 1). These vectors come from the reduced-dynamic reference orbit (Bock *et al.* 2011, ESA 2010).

Table 1

Components of the initial state vectors of the GOCE satellite used
in the orbit improvement process w.r.t. the J2000.0 inertial reference frame (IRF)

No.	x_0 [km]	y_0 [km]	z_0 [km]	$\dot{x}_0$ [km/s]	$\dot{y}_0$ [km/s]	$\dot{z}_0$ [km/s]
1[A]	145.262655	880.057268	–6591.346902	5.901115	–4.956285	–0.533774
2[B]	6423.724497	–1652.533840	–29.552512	–0.198946	–0.872715	7.705914
3[C]	1039.146802	787.067728	–6509.416262	7.641320	0.064912	1.229579

[A]Variant A – initial epoch t_o: 6 November 2009, 23:59:45.00 UTC,
[B]Variant B – initial epoch t_o: 2 December 2009, 23:59:45.00 UTC,
[C]Variant C – initial epoch t_o: 18 December 2009, 23:59:45.00 UTC.

4.1 Computation results

Using of the initial data presented above in the OCS software made it possible to obtain a series of solutions of orbit determination process. Table 2 presents the accuracies of these solutions obtained using the described simulations of the full gravity gradient tensor measurements. Each of the variants A, B, C associated with the given initial state vector (Table 1) is divided into

Table 2

Accuracy of the GOCE orbit improvement solutions
using the simulated observations of the full gravity gradient tensor

No.	Orbital arc length	RMS$_O$/RMS$_{OUT}$		
		Variant A	Variant B	Variant C
1	5.6 min	143.734 m / 1.3 cm	4.2 cm / 4.8 mm	2.5 cm / 3.9 mm
2	11.2 min	274.514 m / 1.2 cm	7.1 cm / 8.2 mm	3.8 cm / 4.9 mm
3	22.5 min	478.906 m / 1.5 cm	13.5 cm / 2.2 cm	4.9 cm / 1.4 cm
4	45.0 min	499.384 m / 7.7 cm	48.9 cm / 4.8 cm	22.5 cm / 7.2 cm
5	90.0 min	499.314 m / 12.5 cm	1.357 m / 10.7 cm	59.7 cm / 12.0 cm
6	6.0 h	498.802 m / 37.4 cm	5.284 m / 34.2 cm	2.300 m / 27.5 cm
7	12.0 h	492.079 m / 7.054 m	10.024 m / 56.9 cm	3.476 m / 56.7 cm
8	18.0 h	489.077 m / 6.926 m	13.521 m / 1.049 m	3.595 m / 1.154 m
9	24.0 h	487.575 m / 6.704 m	14.813 m / 2.759 m	3.993 m / 2.960 m

nine sub-variants, differing in the length of the improved orbital arc. Therefore, Table 2 contains the characteristics of the accuracy of 27 variants of the solution of the orbit improvement process. A given solution variant can be denoted by providing the arc variant A, B, C and the number corresponding to the given arc length. For the solutions A1 to A9, the simulations of gravity gradient measurements are determined with a sampling interval of 1 s and the improved orbital arcs are compared with the corresponding arcs of the kinematic reference orbit. The other solution variants, B1-B9 and C1-C9, use simulations of gravity gradient observations with a 10-s interval. The improved orbital arcs for these variants are compared with the reduced-dynamic reference orbit. Comparison of the RMS_O and RMS_{OUT} values in Table 2 indicates effective orbit improvement for all solution variants. The fit of the improved orbit to the reference orbit increased compared with the fit of the approximate orbit from around 1.3 times (variant C9) even to around 320 times (variant A3). The highest orbit improvement effectiveness (value of the RMS_O/RMS_{OUT} ratio) is shown by variants A1-A9, representing the fit of the improved orbit to the kinematic reference orbit. Relatively high RMS_O values of the fit of the approximate orbit were obtained in this case, which was caused by the adoption of the initial state vector from the reduced-dynamic orbit for the computations. Comparing the RMS_{OUT} values, better accuracies can also be observed for the solution variants related to the reduced-dynamic reference orbit than the accuracies of the variants where the kinematic orbit was the reference orbit. For example, variants B7 and C7 related to the reduced-dynamic orbit, showing the RMS_{OUT} values of 56.9 and 56.7 cm, respectively, have a much better fit than variant A7, where the fit to the kinematic reference orbit is at around 7 m. It appears that the better fit of the obtained solutions to the reduced-dynamic orbit may be connected with its higher coherence – a kind of "smoothness", determined by the presence of a strong dynamic aspect in its determination process. Unlike the reduced-dynamic orbit, the kinematic orbit is a strictly geometric solution, covering only the time series of the satellite position vectors. Hence, the kinematic orbit depends more strongly than the reduced-dynamic orbit on the changing quality of GPS observations, which may sometimes lead to abrupt changes in the position vector. Therefore, the kinematic orbit as the reference orbit shows lower consistency with the improved orbit than the reduced-dynamic orbit, which is particularly seen for longer orbital arcs (variants A7-A9 compared with variants B7-B9 and C7-C9).

However, regardless of the type of the orbit used as the reference orbit, the accuracies of the solutions shown in Table 2 decrease with increasing length of the improved orbital arc (except for variants A7-A9, where a slight increase in accuracy can be observed). This effect is caused by different parameterization of the improved orbital arcs (six components of the initial

state vector) and a different set of used dynamic models compared with the reduced-dynamic reference orbit. Assuming an accuracy of the kinematic orbit and the reduced-dynamic orbit of 2 cm (Bock *et al.* 2011), an arc can be indicated among the orbital arcs presented in Table 2 for which the fit of the improved orbit to the reference orbit expressed by the RMS_{OUT} values reaches the level of aforementioned accuracy. This arc is the arc with the length of 22.5 min, which is around one-fourth of the satellite period. Moreover, the improvement accuracy for two shorter arcs with the lengths of 5.6 and 11.2 min is better than in the case of arc with the length of 22.5 min. This also indicates no need to use more extensive parameterization than this, comprising six components of the initial state vector, for orbital arc lengths lower or equal to 22.5 min.

4.1.1 Sensitivity of orbit determination to spectral content of measurements

Table 3 lists the accuracies of a number of variants of improving the 12-hour orbital arc of the GOCE satellite, in the form of the values of the RMS_{OUT} parameter, obtained using the time series of the simulations of the full gravity gradient tensor. These gradients were determined using the geopotential model JYY_GOCE02S (Yi *et al.* 2013), along the reduced-dynamic reference orbit, with the initial state vector corresponding to variant C from Table 1 (initial epoch: 18 December 2009, 23:59:45.00 UTC). As a result, eleven sets of gravity gradients were obtained. The gravity gradients of given set have a specified spectral content because they were generated us-

Table 3

Accuracy of solution variants for orbit determination
based on the simulated gravity gradient measurements
with different spectral content

Degree and order of geopotential model truncation for the simulation of the gravity gradient observations	RMS_{OUT} [cm]
1. only C_{00}	55.7
2. only C_{00} and C_{20}	55.4
3. 2×2	55.4
4. 5×5	55.4
5. 20×20	55.4
6. 40×40	55.4
7. 70×70	55.4
8. 120×120	55.4
9. 160×160	55.4
10. 200×200	55.4
11. 230×230 (full model)	55.4

ing the aforementioned geopotential model truncated at selected degree and order values of spherical harmonic coefficients. Using these sets of gravity gradients in the orbit determination process allowed to obtain eleven solution variants which accuracies are at the same level of 55.4 cm, with simultaneous approximate orbit accuracy of 10.049 m, computed using the full geopotential model JYY_GOCE02S. Variant 1 with the RMS_{OUT} value of 55.7 cm has a slightly lower accuracy than the others. The gradient observations in this variant are determined based on the main Newtonian term of the model JYY_GOCE02S. Starting from variant 2, the accuracy is at the above-mentioned constant level. This means clear dependence of orbit improvement accuracy on the effect of the long-wavelength part of the gravity field. As can be seen from Table 3, already the inclusion in the observations of only this long-wavelength part of the gravity field, which is connected with the C_{00} and C_{20} coefficients, suffices to maintain an accuracy of 55.4 cm. Adding the next coefficients in further variants to generate measurements, corresponding to medium- and short-wavelength part of the gravity field does not bring any change in the orbit determination accuracy. In other words, orbit improvement is sensitive to this component of gravity gradient measurements which results from the influence of long-wavelength part of the gravity field. This is consistent with the well-known fact of the dominant effect of this part of the gravity field on the satellite orbit itself.

4.1.2 Sensitivity of orbit determination to random error modification of observations

The subject of research was also the improvement of the accuracy for the orbital arc with a length of around 11.2 min depending on the standard deviation of random errors with the normal distribution, which modified the values of simulated gravity gradient observations. To perform this modification, a procedure "gasdev" generating the Gaussian noise was used (Press *et al.* 1986-1992). The corrected initial state vector was the vector 1^A from Table 1 in epoch: 6 November 2009, 23:59:45.00 UTC. Both the approximate arc (computed without improvement) and the improved arcs were compared with the corresponding arc of the reduced-dynamic reference orbit. Hence, the threshold RMS_O value for the computed orbit was around 4.7 cm. An RMS_{OUT} value of the same order for the improved orbit at 5.9 cm appeared when the simulated observations were modified by random errors with the standard deviation of 10^{-13} s^{-2}, which corresponds to 10^{-4} EU (EU − Eötvös unit, 1 EU $= 10^{-9}$ s^{-2}). For the standard deviation of 10^{-14} s^{-2} (10^{-5} EU), the orbit improvement effectiveness threshold was exceeded because the obtained RMS_{OUT} value fell to 7.8 mm. This value is within the accuracy of the reduced-dynamic reference orbit itself, also called the Precise Science Orbit (PSO) (Bock *et al.* 2011).

5. NUMERICAL TESTS – USING REAL MEASUREMENTS

Taking into account promising results for the simulated gravity gradients, the time series of gravity gradient observations V_{xx}, V_{yy}, V_{zz}, V_{xy}, V_{xz}, V_{yz}, from the GOCE mission (Bouman *et al.* 2011) have been used in the GOCE orbit determination. The gravity gradients used belong to the Level 2 product; *i.e.*, they are determined in a static gravity field. The auxiliary data were the elements of orientation of the GRF and ITRF2005 reference frames with respect to the IRF reference frame. The improvement subject were the same 1-day orbital arcs, denoted as variants A, B, C, as in the case of use of simulated gravity gradient observations (Section 4). The corresponding time series of real gravity gradients were assigned to each of these variants. The initial state vectors undergoing the improvement process are equal to the initial state vectors of the corresponding reference orbit arcs (Table 1). As already mentioned, these vectors come from the reduced-dynamic reference orbit (Bock *et al.* 2011, ESA 2010). The values of RMS_O (the fit of the approximate orbit to the reference orbit) for the described variants of the orbital arcs A, B, and C are: 10.02, 14.81, and 3.99 m, respectively.

Table 4 contains the RMS_{OUT} values for 21 variants of solutions of the improvement process for the three listed 1-day GOCE orbital arcs. The improved orbital arcs were compared with the corresponding arcs of the reduced-dynamic reference orbit. As can be seen in column two, the time series of the diagonal components V_{xx}, V_{yy}, V_{zz} and the time series of individual gravity gradient tensor components were used for orbit determination. These series were transformed from the GRF reference frame to the inertial reference frame (IRF). The sampling interval of the used observations is at 1 s. The obtained results are characterized by RMS_{OUT} values at a level of

Table 4

Accuracy of the GOCE orbit improvement solutions
using the real gravity gradient measurements

No.	Observations	RMS_{OUT} [km]		
		Variant A	Variant B	Variant C
1	V_{xx}, V_{yy}, V_{zz}	7732.638	272.101	142.329
2	V_{xx}	636.924	1729.157	903.993
3	V_{yy}	3726.838	1412.344	128.650
4	V_{zz}	700.668	374.939	227.052
5	V_{xy}	867.812	360.509	644.726
6	V_{xz}	391.541	1246.821	339.446
7	V_{yz}	12856.773	1.014×10^9	9135.146

of hundreds, and even thousands of kilometers which, compared with the corresponding RMS_O values given in meters, clearly rules out the possibility of effectively using the real gravity gradient measurements from the GOCE mission for orbit improvement. This is caused by the relatively low accuracy of the measured gravity gradients in the time domain, as indicated by an analysis of a time series including 2 584 174 measurement epochs with a 1-s sampling interval, contained within the time span from epoch 1 November 2009, 00:49:15.937778 UTC to epoch 30 November 2009, 22:38:47.175033 UTC. The absolute value of the average sum of the diagonal components is around 10^{-6} s^{-2}. For comparison, the absolute value of the sum of these components determined analytically is in the order of 10^{-38} s^{-2} while for error-free measurement values it is 0.

Ignoring the orbital aspect in this case, the results presented in Table 4 can be useful to some degree in comparing the quality of the used time series of observations. Comparison of the RMS_{OUT} values for individual variants of improved orbital arcs would indicate the best quality of measurements for arc C (the epochs: 18 December 2009, 23:59:45.00 UTC – 19 December 2009, 23:59:45.00 UTC) and/or maybe the better quality of orientation elements used to the transformation of observations from the GRF to IRF reference frame. Among the gravity gradients, the best result was obtained after the application of V_{yy} (RMS_{OUT} is 128.65 km for arc C). The solutions obtained using the diagonal components V_{xx}, V_{yy}, V_{zz} (RMS_{OUT} – 142.329 km), the component V_{zz} (RMS_{OUT} – 227.052 km) and the component V_{xz} (RMS_{OUT} – 339.446 km) also stand out. The relatively small RMS_{OUT} for the solution C1 with the components V_{xx}, V_{yy}, V_{zz} can be connected with their nominally higher accuracy and better conditioning of the orbit determination process. The solutions obtained based on the gravity gradient V_{yz} show by far the lowest accuracy. This is approximately consistent with the predicted division of the measured gravity gradient tensor components into the nominally more accurate V_{xx}, V_{yy}, V_{zz}, V_{xz} and less accurate V_{xy}, V_{yz} (Bouman et $al.$ 2011). This division results from a design limitation of the gradiometer, in which the precision of measurement along one of the three accelerometer axes is lower by two orders of magnitude than the precision of measurement along the two others (ESA 2008). In this context, the distinguishing RMS_{OUT} value of 360.509 km for the solution using the component V_{xy} in the arc B improvement is also a certain surprise. This has not been yet well understood. On the other hand, the remaining two solutions, A5 and C5, based on the component V_{xy} have the RMS_{OUT} values of 867.812 and 644.726 km, respectively, which generally indicates a smaller accuracy than the solutions based on the nominally more accurate gravity gradients.

6.   SUMMARY AND CONCLUSIONS

The time series of simulated and real gravity gradient observations were used in the GOCE satellite orbit improvement process, applying the classical least squares method. In case of using of simulated gravity gradient observations, the GOCE orbit improvement is effective in all tested cases. The accuracy of the improved orbit variants ranges from the level of millimeters (for the arc lengths of 5.6 and 11.2 min), through the level of centimeters (for the arc lengths of 22.5 and 45.0 min), decimeters (the arc lengths: 90.0 min, 6.0 h, 12.0 h), up to the level of meters (the arc lengths of 18.0 and 24.0 h). At the same time, the orbital arc with the length of 22.5 min is the longest among those tested for which improvement accuracy, from 1.4 to 2.2 cm depending on the variant, is within the accuracy of the used reference orbit (PSO orbit of the GOCE satellite). Hence, for the orbital arc of this length it is sufficient to correct only the initial state vector, without the need to introduce additional dynamic parameters, to maintain the reference orbit accuracy. The improved orbital arcs with lengths from 45.0 min to 24.0 h can be treated as more-or-less accurate approximate orbits.

Further research for the 12-hour arc showed practically complete dependence of the orbit improvement accuracy on this component of gravity gradient measurements, which is generated by the long-wavelength part of the gravity field; it is sufficient to include only the C_{00} and C_{20} coefficients in the observation model.

Disturbing the sets of simulated measurements by random errors indicates that the value of the standard deviation of the measurements in the order of 10^{-5} EU is necessary to achieve an accuracy of the improved orbital arc with the length of 11.2 min under 1 cm, *i.e.*, close to the reference orbit accuracy.

When real observations were used, the obtained solutions had an accuracy of hundreds and even thousands of kilometers. This clearly excludes the possibility of using the real gradient measurements from the GOCE mission for satellite orbit improvement. This is caused by the low accuracy of these observations in the time domain, which is connected with the low accuracy in the low frequency part of spectrum. An additional factor lowering the orbit estimation accuracy is the need to transform gravity gradients from the gradiometer reference frame to the inertial reference frame, where the orbit improvement process is performed.

The obtained results show effective improvement of different GOCE orbital arcs, on condition that a time series of measurements of the full gravity gradient tensor with appropriate accuracy in the long-term frequency range is applied. It is also best for the tensor to be measured directly in the orbit improvement reference frame, *i.e.*, in the inertial reference frame. Of course,

the application of gravity gradients in satellite orbit determination should be treated as an additional possibility for validating the orbit determined using the basic data such as GPS observations. The use of gravity gradient observations for improvement of short orbital arcs, for example, can also be considered, especially where GPS measurements show lower accuracy for different reasons.

R e f e r e n c e s

Bobojć, A., and A. Drożyner (2003), Satellite orbit determination using satellite gravity gradiometry observations in GOCE mission perspective, *Adv. Geosci.* **1**, 109-112, DOI: 10.5194/adgeo-1-109-2003.

Bock, H., A. Jäggi, D. Švehla, G. Beutler, U. Hugentobler, and P. Visser (2007), Precise orbit determination for the GOCE satellite using GPS, *Adv. Space Res.* **39**, 10, 1638-1647, DOI: 10.1016/j.asr.2007.02.053.

Bock, H., A. Jäggi, U. Meyer, P. Visser, J. van den Ijssel, T. van Helleputte, M. Heinze, and U. Hugentobler (2011), GPS-derived orbits for the GOCE satellite, *J. Geodesy* **85**, 11, 807-818, DOI: 10.1007/s00190-011-0484-9.

Bouman, J., S. Fiorot, M. Fuchs, T. Gruber, E. Schrama, Ch. Tscherning, M. Veicherts, and P. Visser (2011), GOCE gravitational gradients along the orbit, *J. Geodesy* **85**, 11, 791-805, DOI: 10.1007/s00190-011-0464-0.

Drożyner, A. (1995), Determination of orbits with Toruń Orbit Processor system, *Adv. Space Res*. **16**, 12, 93-95, DOI: 10.1016/0273-1177(95)98788-P.

ESA (1999), Gravity field and steady-state ocean circulation mission, Report for mission selection of the four candidate Earth Explorer missions, ESA SP-1233(1), European Space Agency, Noordwijk.

ESA (2008), GOCE L1B products user handbook, SERCO/DATAMAT Consortium, ESA Tech. Note GOCE-GSEG-EOPG-TN-06-0137, European Space Agency, Noordwijk.

ESA (2010), GOCE Level 2 Product Data Handbook, European GOCE Gravity Consortium; ESA Tech. Note GO-MA-HPF-GS-0110, European Space Agency, Noordwijk.

Eshagh, M. (2014), From tensor to vector of gravitation, *Artif. Sat.* **49**, 2, 63-80, DOI: 10.2478/arsa-2014-0006.

Eshagh, M., and M. Najafi-Alamdari (2007), Perturbations in orbital elements of a low Earth orbiting satellite, *J. Earth Space Phys.* **33**, 1, 1-12.

Heiskanen, W., and H. Moritz (1967), *Physical Geodesy*, WH Freeman and Co., San Francisco.

Johannessen, J.A., G. Balmino, C. le Provost, R. Rummel, R. Sabadini, H. Sünkel, C.C. Tscherning, P. Visser, P. Woodworth, C.W. Hughes, P. Legrand,

N. Sneeuw,   F.   Perosanz, M.   Aguirre-Martinez,   H.   Rebhan,   and M.R. Drinkwater (2003), The european gravity field and steady-state ocean circulation explorer satellite mission: its impact on geophysics, *Surv. Geophys.* **24**, 4, 339-386, DOI: 10.1023/B:GEOP.0000004264.04667.5e.

Mayer-Gürr, T., A. Eicker, and J. Schall (2010), Regional high resolution geoid and mean sea surface topography determination by a combination of GOCE, GRACE and altimetry data. **In:** *ESA Living Planet Symposium, 28 June – 2 July 2010, Bergen, Norway.*

Métris, G., J. Xu, and I. Wytrzyszczak (1999), Derivatives of the gravity potential with respect to rectangular coordinates, *Celest. Mech. Dyn. Astr.* **71**, 2, 137-151, DOI: 10.1023/A:1008361202235.

Palacios, M., A. Abad, and A. Elipe (1992), An efficient numerical method for orbit computations. **In:** *Proc. AAS/AIAA Astrodynamics Conference, 19-22 August 1991, Durango, USA, Part 1 (A92-43251 18-13)*, Univelt Inc., San Diego, 265-274.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (1986-1992), *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge University Press, New York.

Rebhan, H., M. Aguirre, and J. Johannessen (2000), The Gravity field and steady-state ocean circulation explorer mission – GOCE, *ESA Earth Observ. Quart.* **66**, 6-11.

Rummel, R., and O.L. Colombo (1985), Gravity field determination from satellite gradiometry, *Bull. Géod.* **59**, 3, 233-246, DOI: 10.1007/BF02520329.

Rummel, R.F., D. Muzi, M.R. Drinkwater, R. Floberghagen, and M. Fehringer (2009), GOCE: mission overview and early results. **In:** *American Geophysical Union, Fall Meeting, 14-18 December 2009, San Francisco*.

Rummel, R., W. Yi, and C. Stummer (2011), GOCE gravitational gradiometry, *J. Geodesy* **85**, 11, 777-790, DOI: 10.1007/s00190-011-0500-0.

Yi, W., R. Rummel, and T. Gruber (2013), Gravity field contribution analysis of GOCE gravitational gradient components, *Studia Geophys. Geod.* **57**, 2, 174-202, DOI: 10.1007/s11200-011-1178-8.

Mid-Latitude Single Station F region Storm Morphology and Forecast

Ljiljana R. CANDER

Rutherford Appleton Laboratory, Harwell Oxford, United Kingdom;
e-mail: ljiljana.cander@stfc.ac.uk

A b s t r a c t

This paper describes certain aspects of the F region storm morphology based on vertical incidence measurements at single ionosonde station Chilton (51°.60′N, 358°.70′E). The topics discussed include requirements for better understanding of the ionospheric F region morphology and its forecasting under geomagnetically quiet and disturbed conditions. A few common storms during the years of low (1996 and 1997) and high (2000 and 2001) solar activity are considered as well as the Short-Term Ionospheric Forecasting (STIF) method by using two representative examples. The merits are stressed of near-real-time use of data to provide more accurate specification of the geomagnetically disturbed ionosphere and forecast its structure few hours in advance.

Key words: ionosphere (mid-latitude F region, ionospheric modelling and forecasting), geomagnetic storms, space weather.

1. INTRODUCTION

The ionospheric F region parameters are highly variable on timescales ranging from decades to seconds with the occurrence of ionospheric storms in the F region associated with geomagnetic storms. Even in its quietest moments, the Sun produces the electromagnetic radiation and the solar wind, which simultaneously affect a variety of geomagnetic and ionospheric phenomena. Many modelling techniques successfully describe long-term ionospheric

variations, such as that directly related to the sunspot cycle (Ivanov-Kholodny and Mikhailov 1986, Bilitza 2001, Zolesi and Cander 2014 and references therein). Their results of modelling basic ionospheric parameters are angled towards predicting monthly median values for a given place and time of day. Very often it is possible to predict effectively these values with sufficient accuracy by using only the sunspot number as required input parameter (*e.g.*, Fox and McNamara 1988, Zolesi *et al.* 1993, Bradley 1995, Hanbaba 1999). Although these long-term predictions give a good estimate of the expected average conditions, often defined as the climatological behaviour of the background ionosphere, there are appreciable day-to-day and hour-to-hour variations of different origins with significant scientific and technological consequences.

The conditions and the mechanisms that lead to the ionospheric variability remain the subject of profound investigations. Over the years, particular attempts have been made to study the results of the internationally coordinated investigations of space environmental disturbances, frequently referred to as space weather disturbances (Cander 2008, Hapgood 2011). In the ionospheric domain of these studies, the focus is on forecasting the geomagnetic storm effect on main F region parameters to enable extreme conditions to be quantified so that, particularly for telecommunications planning purposes, likely variability boundaries can be defined (Mannucci *et al.* 2015 and references therein). It is very well known that forecasts are made mostly on the basic of persistence and recurrence that are not always strong enough and/or they are heavily built on an extrapolation of past and prevailing ionospheric conditions (Cander 2003 and references therein). In practical applications this requires an automation of the solar-terrestrial data gathering and processing and on-line forecasting message distribution. There has been ample evidence from different national as well as international research projects that true forecasts of ionospheric disturbances are needed with lead times of up to 24-hours of the present (Hanbaba 1999 and references therein). Very recently, sufficient real-time interplanetary data (*e.g.*, solar-wind parameters) and real-time ionospheric conditions have been obtained to do this effectively (Kamide 2000, Galkin *et al.* 2012 and references therein). However, the F region storm morphology has such a complex spatial and temporal structure that requires not only continuous monitoring of high resolution but morphological studies with theoretical and numerical modeling as well (East-wood 2008, Mikhailov *et al.* 2012, Cander 2015).

Studies of ionospheric parameter variations during geomagnetic storms have been made by a large number of scientists world-wide and summarized in a number of excellent review papers by Matsushita (1959), Rishbeth (1991), Prölss (1995), Buonsanto (1999), Fuller-Rowell *et al.* (1997), Namgaladze *et al.* (2000), Mendillo (2006), and Vijaya Lekshmi *et al.* (2011).

The quantitative modeling of the ionospheric response to geomagnetic activity was also discussed by Fuller-Rowell *et al.* (2000a), Pietrella and Perrone (2008), McNamara *et al.* (2011), and Mukhtarov *et al.* (2013). There is a long list of papers dealing with ionospheric storm case studies during 23th solar cycle. For example, Musman *et al.* (1998) and Jakowski *et al.* (1999) studied the ionospheric total electron content (TEC) during the geomagnetic storm on 10 January 1997 while Cander and Mihajlovic (2005) considered ionospheric spatial and temporal variations during October 2003 storm, among many others. The response of the equatorial ionosphere in the South Atlantic Region to the storm of July 2000 has been described by Basu *et al.* (2001).

In this paper the solar-terrestrial conditions and ionospheric F region response surrounding a few geomagnetic storms in 23th solar cycle are examined to illustrate what knowledge is needed to develop a successful forecasting algorithm. The series did not include the strongest storms of the considered period, but just ordinary ones during the low (1996 and 1997) and the high (2000 and 2001) solar activity. Section 2 gives details concerning the storm periods and use of data. In Section 3 it is shown how day-to-day patterns of the F2 layer critical frequency *fo*F2 (corresponding to the ionospheric peak electron density *Nm*F2) can be used to search outstanding science issues in ionospheric F region storm morphology relevant for its forecasting. Section 4 presents necessary details concerning the current status of F region prediction and covers two examples of F region forecasting at Chilton ionosonde station. A summary of the findings and future work appear in Section 5.

2. STORM PERIODS

Since the beginning of ionospheric measurements a large number of studies have been conducted on the F region storm morphology based on electron density data provided by world-wide network of ionosondes. To illustrate some of more recent findings, the behaviour of the daily hourly critical frequency *fo*F2 measured at Chilton (51°.60′N, 358°.70′E), an ionosonde station that has been operating for decades starting at the Slough (51°.51′N, 359°.40′E) site in the UK, with data available from 1932 onwards, is analyzed. Results shown in this paper are for August, February, and October representing ionospheric summer, winter, and equinox conditions during low (1996/1997) and high solar activity (2000/2001) with emphasize on the raising phase of solar cycle 23 starting from October 1996 as equinox month of absolute solar minimum. Development of a major storm on the last day of March 2001 is analysed in detail making use of *fo*F2 data from Chilton. This ionospheric characteristic was chosen as best representative of the F region behaviour.

Storm periods considered

Month	Monthly mean *Ri*	Storm period with Q days and D days ranks	SSC (UT)	Max *Ap* index
October 1996	0.9*	20, 21, 22 (D2), 23 (D3), 24	None	38 on 22 Oct
October 2000	100.1	2, 3, 4 (D2), 5 (D1), 6 (Q5)	4 Oct at 07:00	116 on 5 Oct
		11, 12 (Q9), 13 (D5), 14 (D3), 15	12 Oct at 22:28	45 on 14 Oct
February 1997	7.6	7 (Q5), 8, 9 (D5), 10 (D3), 11 (D4)	8 Feb at 09:54 9 Feb at 13:22 11 Feb at 04:58	22 on 8 and 10 Feb
		24, 25 (Q9), 26, 27(D2), 28 (D1)	None	37 on 28 Feb
February 2000	112.9	11, 12 (D1), 13, 14 (D3), 15	11 Feb at 02:58 and at 23:52	60 on 12 Feb
		22 (Q6), 23, 24 (D5), 25, 26	23 Feb at 11:00	30 on 24 Feb
August 1997	24.4	2 (Q9), 3 (D1), 4, 5 (Q3), 6 (Q2)	3 Aug at 10:42	21 on 3 Aug
August 2000	130.5	10, 11 (D2), 12 (D1), 13, 14	11 Aug at 18:45	123 on 12 Aug
March 2001	114.2	29, 30, 31 (D1)	31 Mar at 00:52	155 on 31 Mar

*) The consensus minimum value that occurred in October 1996.

The selected 10 storm events are listed in Table 1. The selection of the geomagnetic storms in Table 1 is based on data published in the Solar-Geophysical Data Prompt Report, NOAA, Boulder, USA (http://www.ngdc. noaa.gov), and obtained from the WDC-C2 for geomagnetism, Kyoto University, Japan (http://wdc.kugi.kyoto-u.ac.jp). In Table 1 *Ri* is the monthly mean of the international relative sunspot number and *Ap* is an averaged planetary geomagnetic index based on data from a set of specific geomagnetic observatories. The daily *Ap* index serves to classify the storms as: minor geomagnetic storm when $29 < Ap < 50$; major geomagnetic storm when $49 < Ap < 100$; and severe geomagnetic storm when $Ap \geq 100$ (Coffey and Erwin 2001). In addition, international geomagnetically disturbed days (D-days) and the quietest days (Q-days) are also given in Table 1 together with the UT of the storm sudden commencement (SSC). Criteria used in the rankings the 10 most geomagnetically quiet days of the month from most (Q1) to least quiet (Q10) and the 5 most geomagnetically active days from the most

(D1) to least disturbed (D5) include the sum of the eight planetary 3-hour range Kp index values, the sum of the squares of the eight Kp values, and the greatest Kp value (http://www.iugg.org/IAGA). Thus, the five disturbed and the ten quiet days of each month could define the extremes of day-to-day ionospheric variability if the geomagnetic activity were the only source of this variability (Mendillo 2006). Days without numbers in brackets (Table 1), that are neither geomagnetically quiet nor disturbed, point also to the role of other solar-terrestrial processes involved in the complex ionosphere structure and dynamic.

3. F REGION STORM MORPHOLOGY AT CHILTON IONOSONDE STATION

Realistic simulations of both positive and negative disturbance deviations from the average monthly median behaviour following a storm commencement require detailed descriptions of the storm morphology. In this paper a proper examination how the spatio-temporal storm patterns develop and how they are related to solar and geomagnetic activity (Table 1) for the characteristic descriptive of F region ionisation, foF2, has been done by looking at different individual storms.

The hourly daily critical frequency foF2 measurements at Chilton taken during October 1996 are shown grouped together in Fig. 1 indicating 22 and 23 October as days of the significant foF2 departure from the overall narrow monthly scattering pattern. During this month, some foF2 values were not observable from the ground indicating possible ionospheric G conditions (Piggott and Rawer 1972). These can be clearly seen on the last two days in Fig. 2. The median values, foF2med, shown in Fig. 2 as black line represent average diurnal behaviour at Chilton ionosonde station and enable the storm developments to be studied. They are frequently used to identify the negative and positive storm phase as foF2 decrease (–25%) below and increase (+25%) above median values during disturbed conditions in terms of the relative deviation $\Delta fo\text{F2 (\%)} = 100 \times (fo\text{F2} - fo\text{F2med}) / fo\text{F2med}$. Although Table 1 indicates no storm sudden commencement (SSC) on geomagnetically disturbed day 22 October 1996 (D2), there is an ionospheric storm with short positive phase followed by a negative phase on 23 October 1996 (D3) lasting more than 1 day (Fig. 2). Here ionospheric G conditions perceptible as missing foF2 data are associated with the negative storm phase (Smithtro and Sojka 2005). During this storm interval, the foF2 varied from 7.5 to 3.0 MHz at noontime while the sunspot number was around 0.9. It is reasonable to assume that the solar EUV flux does not change much during the 5 days period when solar activity remained low. Therefore, it is evident that even at the extreme solar minimum, hourly daily foF2 changes during minor geomagnetic storm (max $A_p = 38$) are significant relative to the monthly
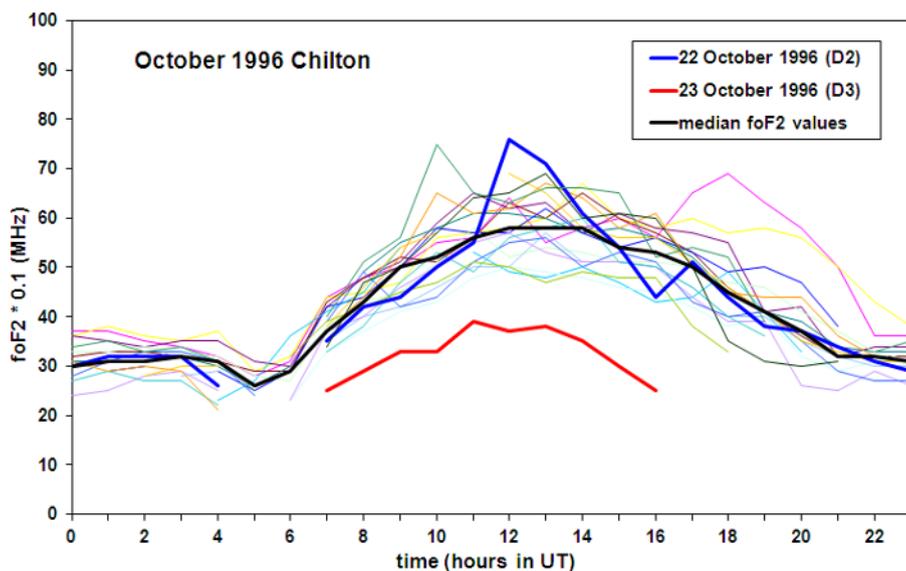
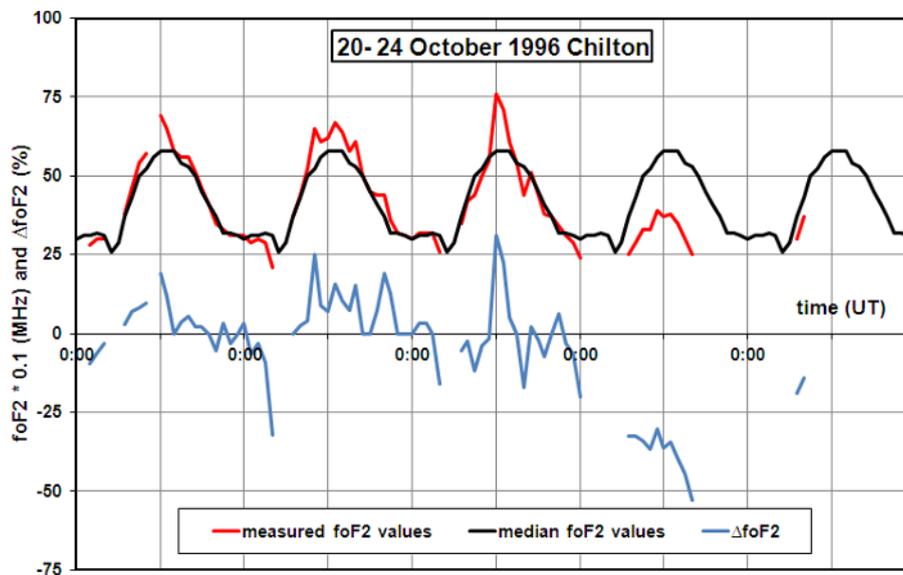Fig. 1. Time variations of *fo*F2 at Chilton ionosonde station for each day in October 1996.

Fig. 2. *fo*F2 measured values with Δ*fo*F2 between 20 and 24 October 1996 at Chilton ionosonde station.
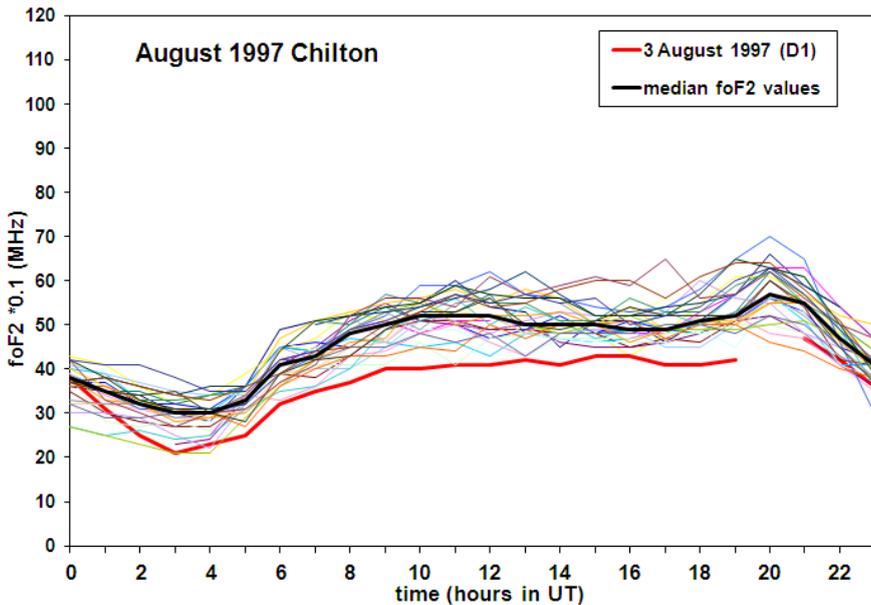
Fig. 3. Time variations of *fo*F2 at Chilton ionosonde station for each day in August 1997.

monthly average behaviour, ΔfoF2 $< -50\%$. Such behaviour is particularly important in high frequency (HF) radio communication applications (Zolesi and Cander 2014 and references therein).

Results of 31 days of *fo*F2 measurements at Chilton taken during August 1997 and August 2000 are shown grouped together in Fig. 3 and Fig. 4, respectively. They present several interesting phenomena apart from an apparent contrast of solar minimum *versus* solar maximum diurnal *fo*F2 behavior in the mid-latitude F region. The diurnal values of *fo*F2 shown in the mass plot of Fig. 3 are significantly lower than those shown in Fig. 4. This is due to the lower solar activity in 1997. Diurnal *fo*F2 values measured in the summer of 2000 were consistently higher than those in the summer of 1997 by a factor of approximately 1.8. This ratio is by far less than the ratio of the mean sunspot number observed during the two summers as 5.3. However, the effect of decreasing solar activity is clear: there are decreases in both the maximum values and the maximum-to-minimum ratio. The late afternoon maximum or the so-called evening anomaly is seen at both figures, though less pronounced at high solar activity. Indeed, monthly median values clearly show an increase in *fo*F2 up to approximately 20:00 UT which indicates that the upper F region moves downward in the summer afternoon with ionization piling up at the peak of the F2 layer.
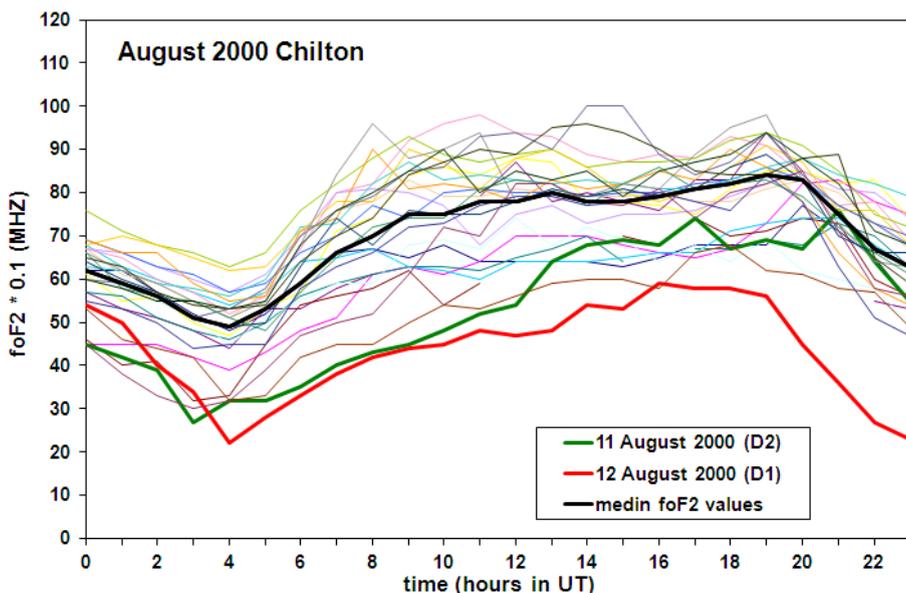
Fig. 4. Time variations of *fo*F2 at Chilton ionosonde station for each day in August 2000.

From Fig. 5 a strong relation can be seen between the late afternoon maximum value in *fo*F2 and the geomagnetic conditions. Figure 5 shows data from 2 to 6 August 1997. As given in Table 1, three of the ten geomagnetically quietest days (*e.g.*, Q9, Q3, and Q2) and one of the five geomagnetically disturbed days (*e.g.*, D1) of the month occurred during the period. Although it appears that the afternoon maximum in *fo*F2 was higher when the geomagnetic activity was low than when activity was high, the late afternoon maximum value of *fo*F2 was not well related to geomagnetic conditions during the 10-14 August 2000 disturbed period (Fig. 6).

However, the most dominant feature of the mass plots in Figs. 1, 3, and 4 is a large day-to-day scatter in the *fo*F2 values, confirming striking F region sensitivity to rapidly changing solar-terrestrial activity. There are many scatters due to the geomagnetic storm that is found to be a permanent feature of all events here examined. It is also evident that all storms in 1996 and 1997 are small-scale geomagnetic disturbances. The 3 August 1997 minor storm affected the upper atmosphere in such a manner that the *fo*F2 variation appeared initially (early afternoon) as a positive-value phase, then was followed by a negative level for the next 36 hours, and finally returned to the median level as shown in Fig. 5. The *fo*F2 median value is here in excellent agreement with the measured *fo*F2 on the geomagnetically quiet (Q2) 6 Au-
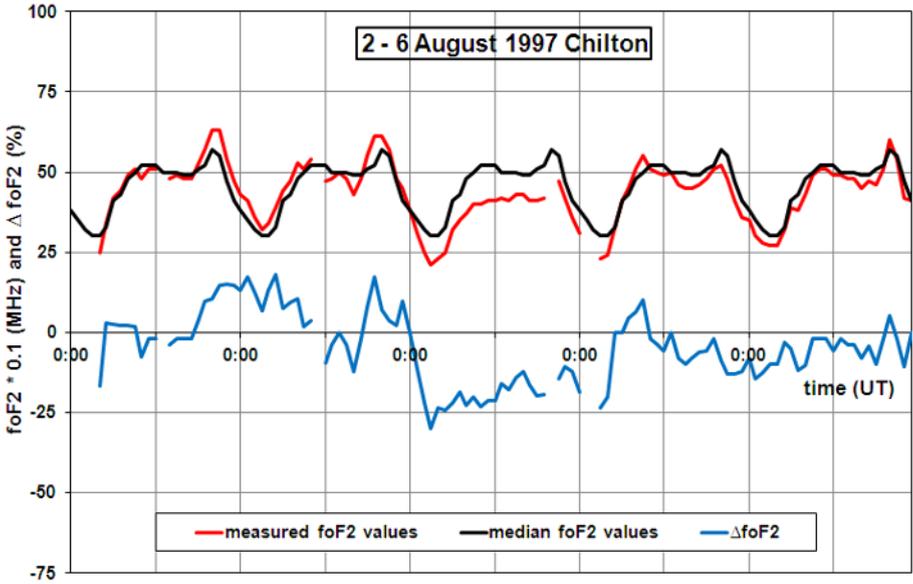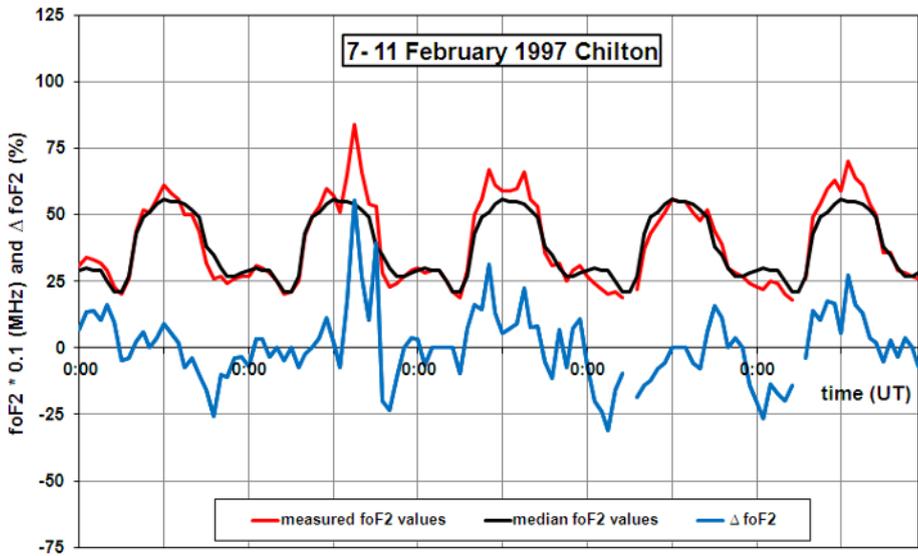
Fig. 5. *fo*F2 measured values with ΔfoF2 between 2 and 6 August 1997 at Chilton ionosonde station.
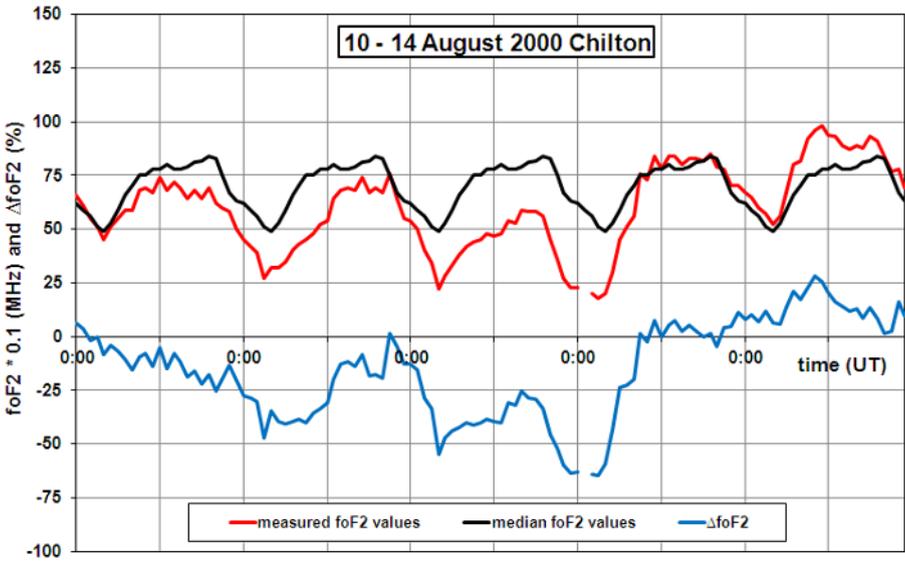
Fig. 6. *fo*F2 measured values with ΔfoF2 between 10 and 14 August 2000 at Chilton ionosonde station.

gust. The measured *fo*F2 values in Fig. 5 illustrate the capability of the
F region to recover quickly after a minor storm, where the magnitudes of
Δ*fo*F2 scarcely crossed the storm threshold level of ± 25%. This pattern was
not repeated during the severe storm of 11 August 2000 illustrated in Fig. 6.
In contrast to Fig. 5, the F region storm pattern exhibits only negative phase
in *fo*F2 with the large depletions in measured *fo*F2 values during both day-
time (around –40%) and night-time (around –60%) ionosphere and the long
lasting disturbed ionospheric structure. Such behaviour puts great limitation
not only on proper understanding of the underlying sources of ionospheric
storm during summer season but its forecasting too. This difference is em-
blematic of the difficulty of bringing the forecast *fo*F2 values into agreement
with the measurements.

   The beginning and end of February 1997 was a period of low solar activ-
ity (*Ri* = 7.6) and moderate geomagnetic activity (max *Ap* = 37). The severi-
ty of the F region disturbances was indicated in Fig. 7 by positive *fo*F2
deviations from median on 8 and 26 February. Accompanying with the mi-
nor geomagnetic disturbance on 8 February 1997 at 09:54 UT, the *fo*F2 val-
ues strongly and suddenly increased with about 55% (Fig. 8a). The levels of
the positive perturbed *fo*F2 values also grow longer at daytime on 9 and
11 February, which was caused by additional SSCs at 13:22 and 04:58 UT,
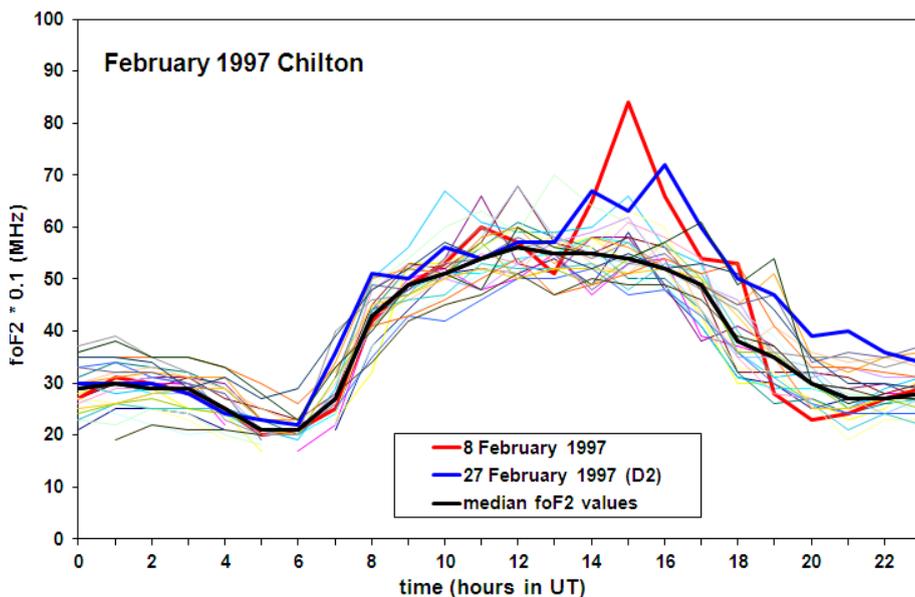
Fig. 7. Time variations of *fo*F2 at Chilton ionosonde station for each day in February
1997.

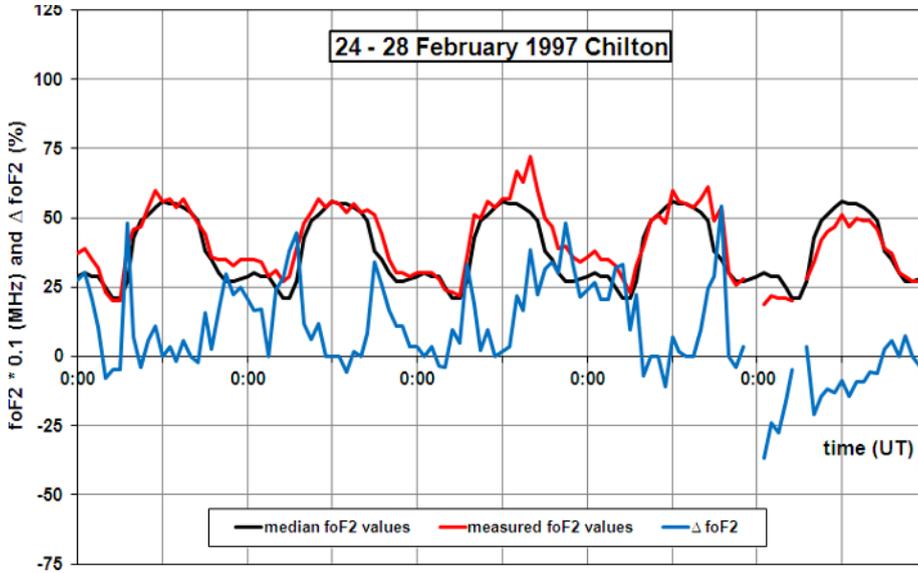Fig. 8a. *fo*F2 measured values with Δ*fo*F2 between 7 and 11 February 1997 at Chilton ionosonde station.

Fig. 8b. *fo*F2 measured values with Δ*fo*F2 between 24 and 28 February 1997 at Chilton ionosonde station.
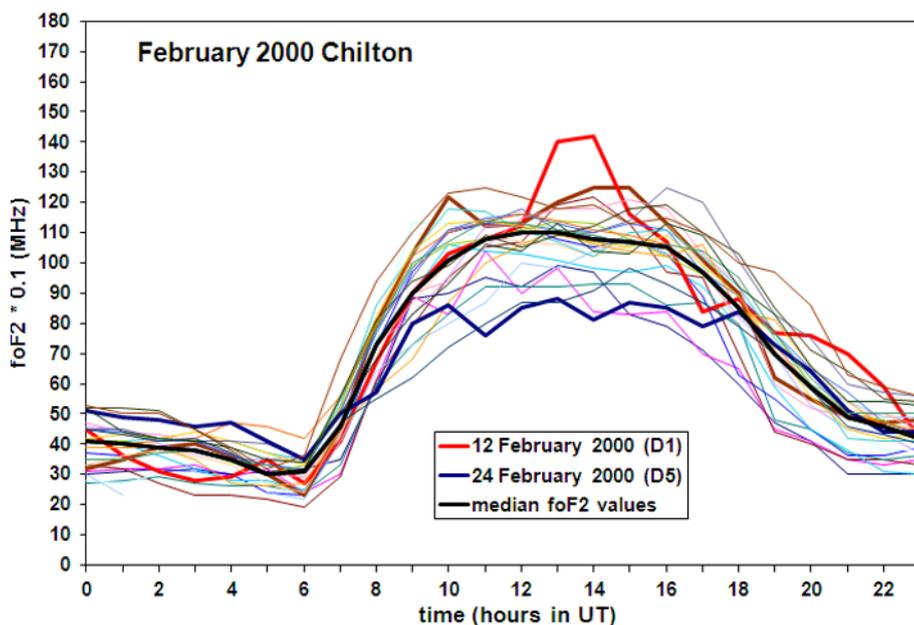
Fig. 9. Time variations of *fo*F2 at Chilton ionosonde station for each day in February 2000.

respectively (Table 1). As is also seen in Fig. 8b, the level of the *fo*F2 variations tends to increase during the first day of the minor storm of 26 February 1997 and then oscillate to the usual day level in a few days after the beginning of the storm. In addition, an interesting influence of the Travelling Ionospheric Disturbances (TIDs) is visible in *fo*F2 oscillations on 26 and 27 February 1997 (Dominici *et al.* 1988, Hocke and Schlegel 1996, Afraimovich *et al.* 2001).

The scatter in data plotted at Fig. 9 for each day in February 2000 shows that the measured *fo*F2 values were far from close to average, as represented by the monthly median, on a number of days following two storm periods. The major storm of 11 February 2000 (Fig. 10a) affected the *fo*F2 variation with daytime enhancements on the first and second days, negative phase during the night-time and speedy recovery later on. During the second storm in February 2000, the enhancement starts after the storm onset and produces clear but relatively small positive phase (Fig. 10b). A negative phase occurs on the second day and lasts more than 24 hours. However, having in mind the high solar ($Ri = 112.9$) as well as geomagnetic activity in February 2000 (max $Ap = 60$) it should be noted that the *fo*F2 modification was comparatively small during disturbed times. Consequently, it seems that minor storms
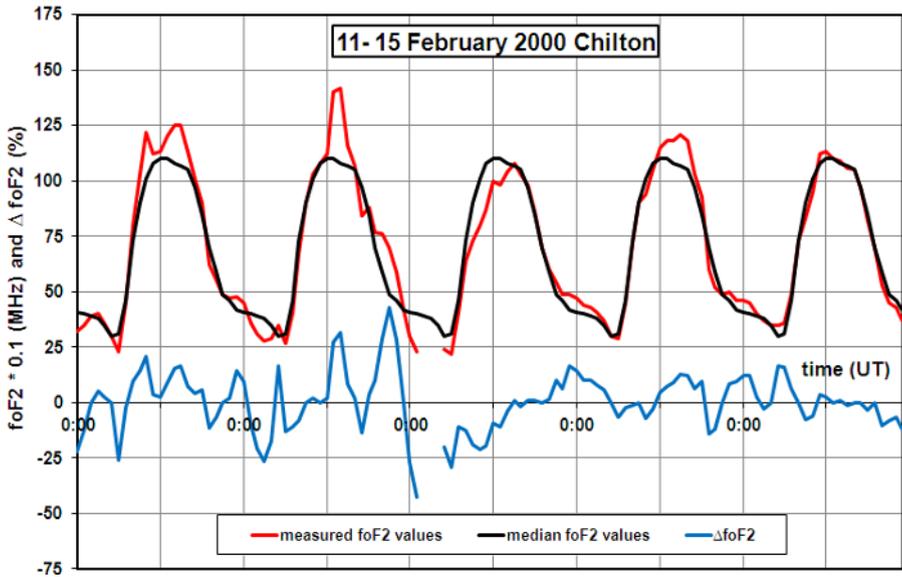
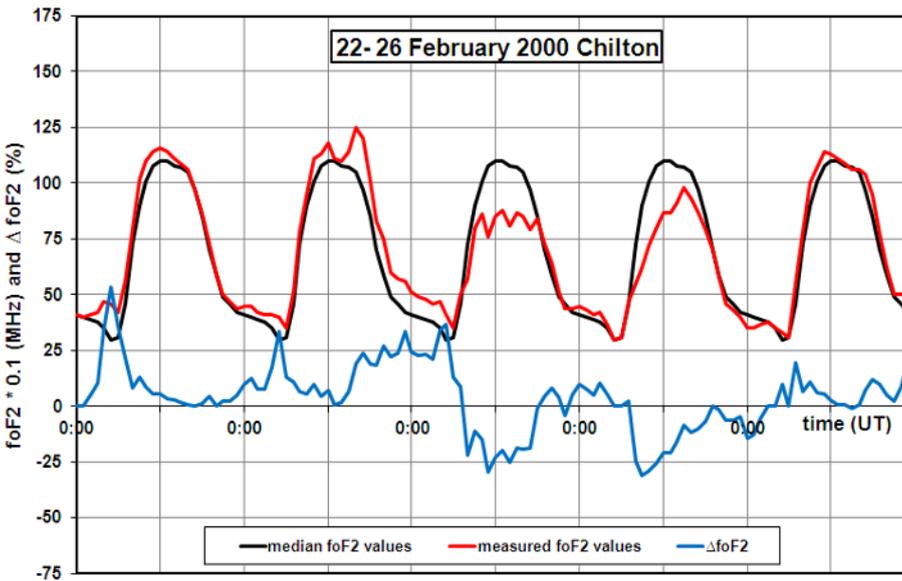Fig. 10a. *fo*F2 measured values with Δ*fo*F2 between 11 and 15 February 2000 at Chilton ionosonde station.

Fig. 10b. *fo*F2 measured values with Δ*fo*F2 between 22 and 26 February 2000 at Chilton ionosonde station.
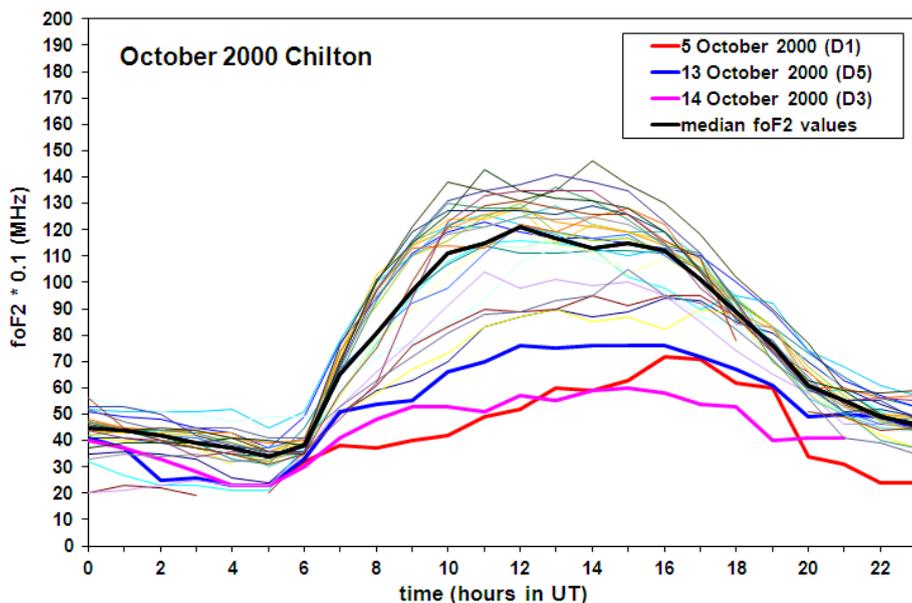
Fig. 11. Time variations of *fo*F2 at Chilton ionosonde station for each day in October 2000.

in February at low level of solar activity (Figs. 8a and b) produced similar if not stronger effect than major storms in February at high level of solar activity (Figs. 10a and b). This important finding neither can be easily explained nor accurately forecast with existing knowledge.

All this certainly was not the case in October 2000, where the scatter *fo*F2 measurements for 31 days at Chilton ionosonde station are shown grouped together in Fig. 11. The large scatter at any fixed UT again emphasises significant day-to-day variability. Although there is considerable scatter above the monthly median, the figure shows a clear tendency for negative storm effects. This is well illustrated in Figs. 12a and b where details concerning two storms from Table 1 are given. Both ionospheric storms show no positive phase at any stage of development. The most severe depletions (about –40 to –60%) occur in the daytime sectors but persist into the night-side too. One of the most interesting findings that has profound consequence on storm forecasting is the changes in *fo*F2 from quiet to disturbed times on 2 and 12 October when measured *fo*F2 decreases more quickly than increases during the long lasting recovery phase on 6 and 15 October, respectively.

Although most results confirm the very well known fact that a seasonal variation of the positive and negative phases in *fo*F2 exists when storm effects are compared for winter (Figs. 7 and 9) and summer (Figs. 3 and 4) it is
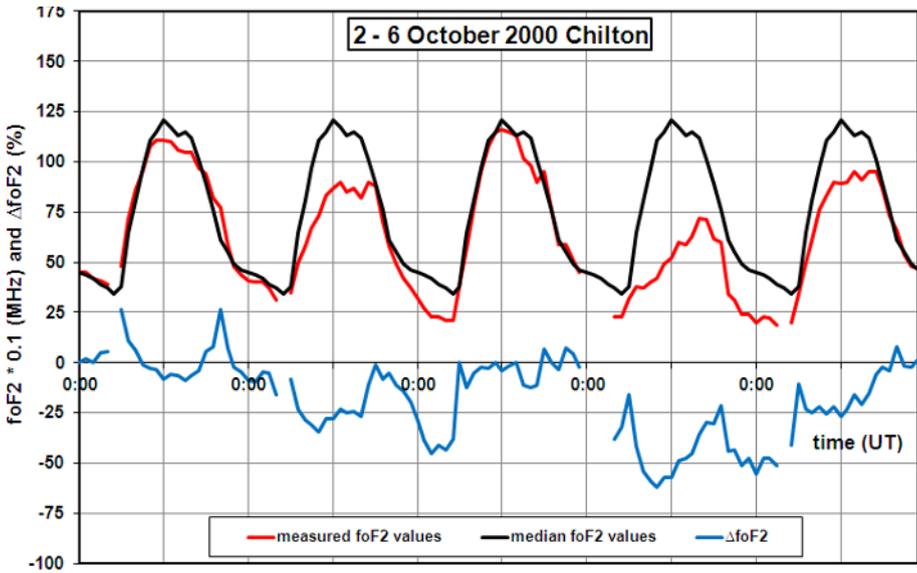
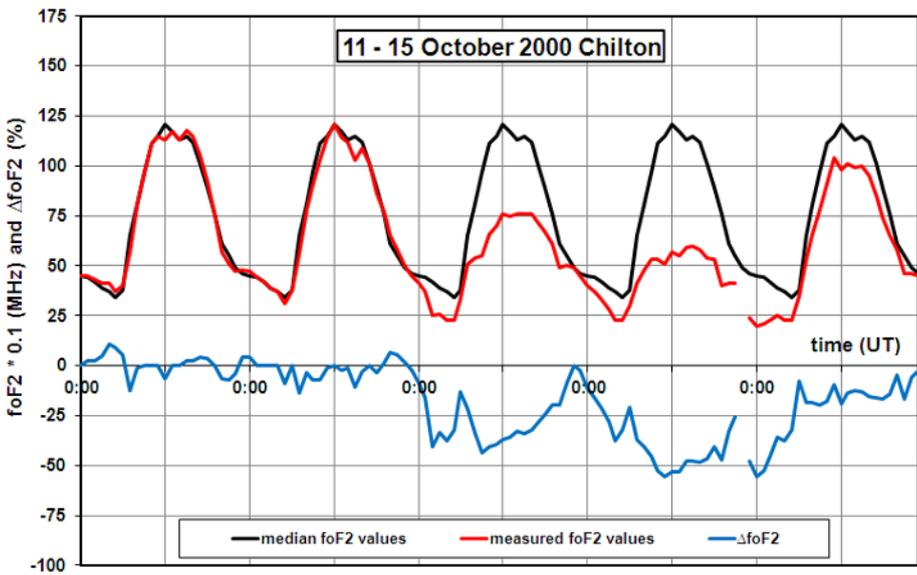Fig. 12a. *fo*F2 measured values with Δ*fo*F2 between 2 and 6 October 2000 at Chilton ionosonde station.

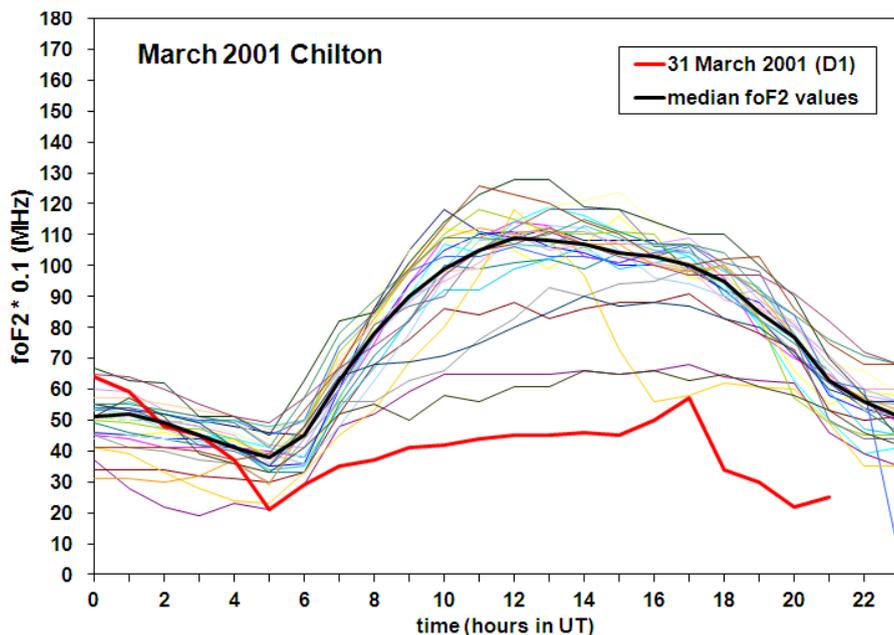Fig. 12b. *fo*F2 measured values with Δ*fo*F2 between 11 and 15 October 2000 at Chilton ionosonde station.

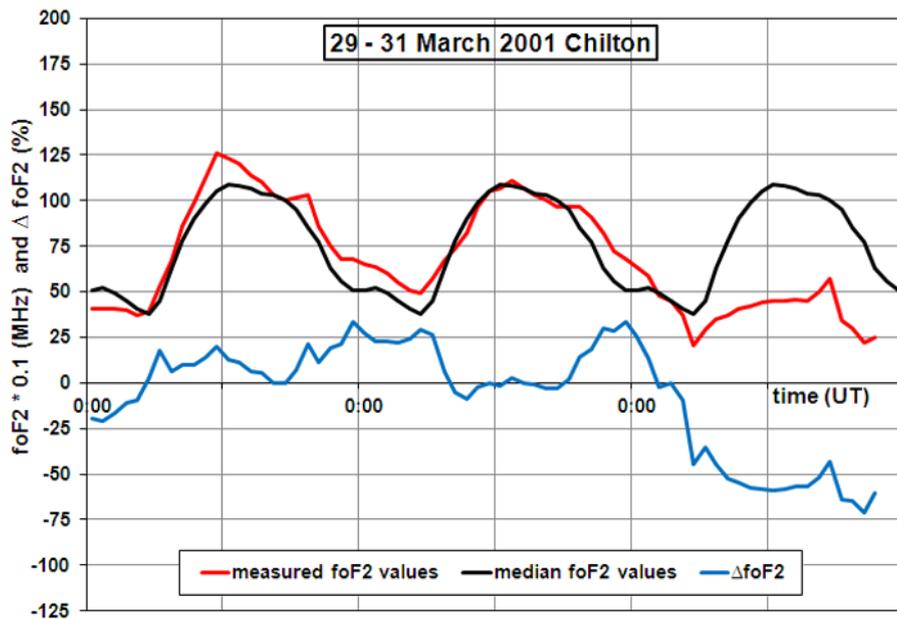Fig. 13. Time variations of *fo*F2 at Chilton ionosonde station for each day in March 2001.

Fig. 14. *fo*F2 measured values with Δ*fo*F2 between 29 and 31 March 2001 at Chilton ionosonde station.

evident that the storm pattern is much more complex. We argued that foF2 enhancements at Chilton winter ionospheric storm patterns are not to be strongly dependent neither on geomagnetic nor solar activity levels. While varying in magnitude, depletions in foF2 are generally more severe in summer and equinox storm patterns. Furthermore, all the results show that daytime foF2 is greater in winter (Fig. 9) than in summer by day (Fig. 4) which is a clear demonstration of the winter or seasonal anomaly (Rishbeth *et al.* 2000). The semi-annual anomaly can be also seen in Figs. 11 and 4 when foF2 is greater at equinox than at summer. It has been emphasised by Millward *et al.* (1996) that the changes in chemical composition, specifically [O/N$_2$] ratio, taken in conjunction with the seasonal changes of solar zenith angle are responsible for the existence of both seasonal and semi-annual anomaly behaviour of foF2.

Changes in daily hourly foF2 values for all the March 2001 data are shown in Fig. 13. Evidently there is a part of scatter in Fig. 13 that can be attributed to the high level of geomagnetic activity (Table 1). As in previous figures, black and red solid lines represent median and storm day foF2 variations, respectively. Once more during daytime the variability in foF2 is found to be considerably larger than by night-time. Both Figs. 13 and 14 displays the last day of March 2001 being the most disturbed period of the month with very well define negative phase in foF2. The negative phase of the ionospheric storm is detected in early morning hours (after 05:00 UT) of 31 March with the severe depletions of more than –60% relative to the median foF2 values. Accordingly, it is clear that disturbed ionisation structure during geomagnetic storms creates a large number of different technical, scientific and numerical problems in matching the rapid foF2 variations in measured data with corresponding forecasting model's outputs as storms develop. The factors that lead to such a complex F region behaviour even at a single ionosonde station of Chilton are beyond the scope of this paper.

4. F REGION STORM FORECASTING

There are number of ionospheric global and regional prediction models based on an empirical fit to hourly monthly median values extracted from an archived ionosonde data base for past solar cycles (Bradley 1995, Hanbaba 1999 and references therein). These climatological models are often used, among other purposes, to provide very important quiet-day reference for the world-wide ionospheric conditions. Critical to this type of modelling is the formulation of maps giving the geographical variations of individual model parameters. Reference global maps in current international use are those recommended by International Telecommunication Union (ITU). Important element in these long-term models is a statistical indication of the ionospheric day-to-day variability that is parameterised by the decile factors in

ITU-R approach (ITU-R 1997). Currently the best known ionospheric model is the IRI-International Reference Ionosphere (Bilitza *et al.* 2011). It provides monthly averages of ionospheric electron density, plasma temperature, and ion composition as a function of altitude for any location and time. More recently the ionospheric day-to-day variability is incorporated into IRI model (Fuller-Rowell *et al.* 2000b, Huang *et al.* 2001).

The same is the case with European regional ionospheric monthly median models developed under the four ionospheric COST (European Co-operation in the field of Scientific and Technical Research) Actions over the period 1999-2008 and currently available for implementation in different applications (Zolesi and Cander 2014 and references therein). It is important to note that within these Actions related to the effects of the upper atmosphere on terrestrial and Earth-space communications, studies have been conducted aiming to produce statistically derived expressions indicative of departures from ionospheric median values as a function of different conditions including time of day, season, and extreme solar-terrestrial activity (Strangeways *et al.* 2009). Considerable progress has been made in developing techniques which permit the updating of F region monthly median prediction using either single-station ionospheric data or an effective sunspot number or both (Reilly *et al.* 1991, Wilkinson *et al.* 2001, Pezzopane *et al.* 2011, 2013). An empirical storm-term correction model has been developed by Fuller-Rowell *et al.* (2000b) to predict the departure of *fo*F2 from the monthly median or a quiet-time reference model during a geomagnetic storm. These techniques provide a link and a reference base between the climatological modelling and short-term forecasting. Thus it would be useful to make these techniques internationally available for testing and verification.

Although the F region forecasting might be applied in terms of deviations from median or average conditions upcoming from the prediction models, the morphological characteristics of individual ionospheric storms, as shown in Section 3, clearly demonstrate that storm forecasting requires different approach. It requires a specification of storm-time ionospheric variability on generally less predictable time-scale such as few hours in advance. This is because the largest contribution to the ionospheric day-to-day variability comes from storm effects and their pattern has to be known. One of the first steps towards forecasting and mapping ionospheric space weather over Europe has been made at the Rutherford Appleton Laboratory, UK, by providing an interactive tool for the Short-Term Ionospheric Forecasting (STIF) over Europe. The forecasts of *fo*F2, MUF(3000)F2 (Maximum Usable Frequency for 3000 km distance), TEC and FOT (Frequency of Optimum Traffic) up to 24-hours ahead were available on-line (Dick *et al.* 1999). Figure 15 shows the measured *fo*F2 at Chilton compared with the 24-hours ahead corresponding STIF values for this particular single station
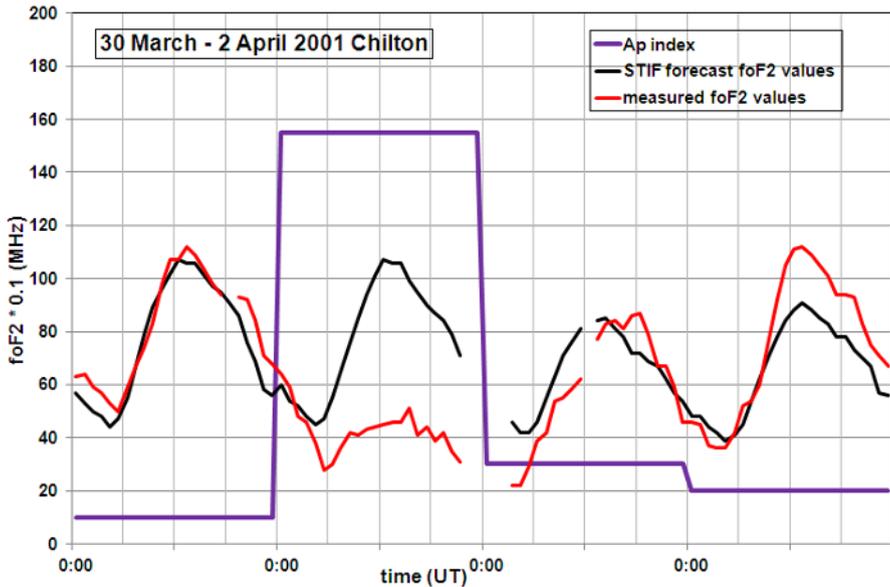
Fig. 15. The STIF 24-hours ahead forecast and measured *fo*F2 values between 30 March and 2 April 2001 at Chilton ionosonde station with *Ap* values as an indicator of high geomagnetic activity level.

over the period 30 March – 2 April 2001. Related geomagnetic conditions are described by daily *Ap* index at the same figure. While *fo*F2 forecast values indicate a little departure from the measured data on 30 March, they show a substantial increase above (31 March) and a slight decrease below (2 April) the measured values during the larger event (*Ap* = 155) started on early hours of 31 March 2001. This disagreement between the forecast and measured F region densities at the beginning of the storm period, particularly in the first 24 hours, is a well-known and long-standing ionospheric forecasting problem. It is also closely related to the problem of the geomagnetic storm forecasting evidently present in sudden change of the *Ap* index from *Ap* = 10 on 30 March to *Ap* = 155 on 31 March. However, reasonably good agreement between forecast and measured *fo*F2 values just before the storm beginning and on the second storm day seen at Fig. 15 is satisfactory under the conditions of the STIF method's generation and its purpose.

Following this approach, the subsequent STIF forecasting example is shown in Fig. 16. It demonstrates how well the STIF 24-hours ahead forecast agrees with measured *fo*F2 values during the ten quietest days in March 2001 at Chilton ionosonde station. Overall comparisons with observations are encouraging because of the small errors (RMSE = 0.87 MHz and NRMSE = 0.35 over 239 measured hourly *fo*F2 values) and a high consis-
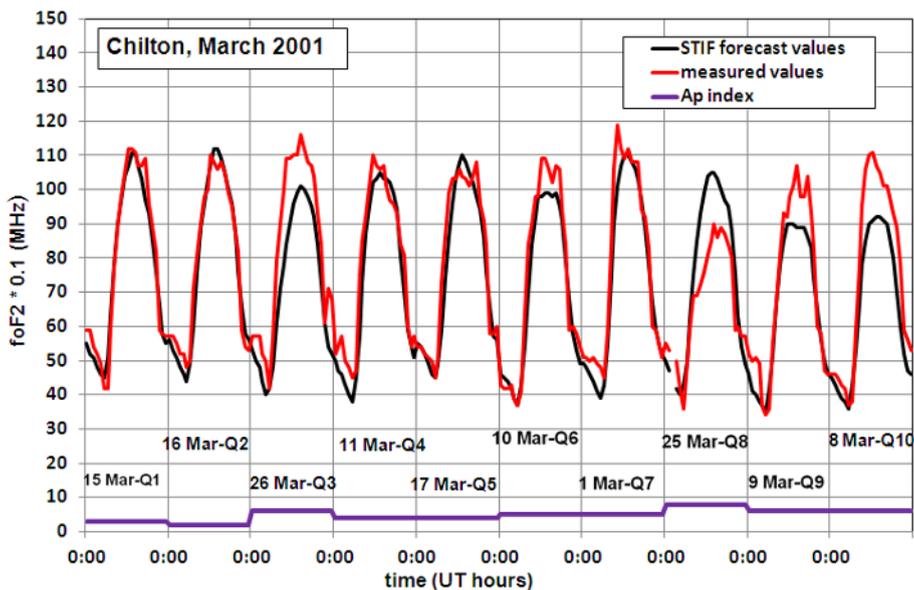
Fig. 16. The STIF 24-hours ahead forecast and measured *fo*F2 values during the ten quietest days (from Q1 to Q10) in March 2001 at Chilton ionosonde station conditions with *Ap* values as an indicator of low geomagnetic activity level.

tence between diurnal morphologies exhibited by the STIF method and the observed trends (Cander *et al.* 2003). The STIF method does not require neither predicted nor past values of any solar and geomagnetic indices to forecast 24-hour ahead the *fo*F2 values even during geomagnetically very disturbed periods as those in March 2001 at Chilton ionosonde station (Fig. 15). In addition, the STIF method does not require *fo*F2 monthly median values, so that the newly developed sliding median approach (Mukhtarov *et al.* 2013), although very useful in ionospheric modelling, is not applicable. However, it has to be emphasized that the STIF technique's success is based on a continuos and rapid measurements over as largest area as possible because it requires near real-time data availability in conjunction with mathematical algorithms to extrapolate the ionospheric conditions some time ahead. Thus, the world and/or regional centres for data collection and processing in near-real time should be available making the final products, such as the F region storm forecasting, useful and promptly available for the different users.

5. CONCLUSIONS

Modern ground-based and satellite communications networks and systems rely on innovations in the radio frequency (RF) domain and demand a full

understanding of the underlying details within the different propagation RF aspects under expected conditions (Bilitza 2002). Hence one of the requirements of these cutting-edge technologies and processes partly addressed by this paper is F region ionospheric storm morphology and its forecasting. The literature on F region storm morphology is now so extensive that a comprehensive review is practically impossible. Instead, only few special topics closely related to its forecasting have been selected for discussion. There is generally accepted theory that explains the F region morphology during geomagnetic storms in terms of the thermospheric changes in both neutral winds and chemical composition which ultimately result in changes of the rates of ionisation production and loss. In addition, changes of the electric field and current systems make ionospheric storm morphology yet more complicated. See Prölss (1995), Rishbeth *et al.* (2000), and Namgaladze *et al.* (2000) as well as references mentioned therein for reviews of these approaches. In spite of that consensus over physical processes involved in main causes of the F region storm morphology, the forecasting ability in case of an individual storm is rather limited at the very best (Pietrella 2012). Established by careful examination of Fig. 15 and Fig. 16, it is clear that the STIF method provides acceptable forecast a few hours in advance just before the beginning of the storm, unsatisfactory forecast during the main phase and the right choices of *fo*F2 values 24-hours in advance during the recovery phase as well as during geomagnetically quiet days.

As studies continue to search for the actual physical mechanisms that explain the spatial and temporal morphologies over various space weather conditions and to test forecasting algorithms for practical applications, it would be useful to have well defined patterns of mid-latitude storm F region and a technique to forecast it. A contribution in that direction is given by findings in Section 3 of this paper. It demonstrates that the whole issue of the forecasting techniques of space weather related events in the Earth's ionosphere cannot be a replacement for the solar-terrestrial physic studies where the highly sophisticated equipment is employed to investigate space and Earth phenomena and where models and methods of prediction and forecasting require powerful computers and long-lasting calculations of the high accuracy. To the contrary, ionospheric space weather is a dynamic system dealing with data in near-real time and predicting tools to alert users on major effects which can damage their technological or financial systems. However, many existing ionosondes are not good enough for these purposes. They are not homogeneously distributed around the world and most of them do not provide data in near-real time. In fact most of them are not technical capable to do that. Fortunately, for quite some time, the ionospheric community has been aware of the fact that the International Global Positioning System (GPS) Service for Geodynamics (IGS) with a global ground-based station

network offers a new multi-point opportunity to observe directly the terrestrial ionosphere and use that information in prediction and forecasting purposes. The total electron content (TEC) evaluated from Global Navigation Satellite System (GNSS) is convenient because: (a) RINEX files for TEC evaluations are available through IGS properly organised at the international level, which means a readily accessible data; (b) RINEX files allow the TEC evaluation with high resolution (1 s, 30 s, 1 min, 10 min, and so on), which means a continuous coverage in time with high temporal resolution; and (c) the number of GPS receiver sites is growing every day, which means the simultaneous global coverage. So what is the problem, then? The problem is the technique for TEC data evaluation from GPS measurements. In case of the ionosonde data there is an internationally agreed procedure for data evaluation from ionograms by Piggott and Rawer (1972). On the other hand, in spite of a lot of work, it is not the case with the GNSS TEC values. Does it mean that there is no way anyone can evaluate TEC from GPS or any other GNSS in an exact way? Still whatever technique is used for the TEC evaluation, a large amount of data available in near real-time is a unique opportunity to be used in space weather issues and statistical studies (Fuller-Rowell *et al.* 2006, Jakowski *et al.* 2011).

With the service for solar monitoring (*e.g.*, http://sidc.oma.be) already available, the multi-point observing capabilities of GNSS, and shared international real-time ionosonde data resources as these *via*, *e.g.*, GIRO (Galkin *et al.* 2012) and DIAS (Cander 2006), the appearance of an effective operational Short-Term Ionospheric Storm Forecasting (STISF) tool to replace existing STIF technique for the global and/or regional interactive use on the World Wide Web is only a matter of time. However, the F region storm morphology and its patterns should still be subject to further investigations.

References

Afraimovich, E.L., E.A. Kosogorov, O.S. Lesyuta, I.I. Ushakov, and A.F. Yakovets (2001), Geomagnetic control of the spectrum of travelling ionospheric disturbances based on data from a global GPS network, *Ann. Geophys.* **19**, 7, 723-731, DOI: 10.5194/angeo-19-723-2001.

Basu, S., S. Basu, K.M. Groves, H.-C. Yeh, S.-Y. Su, F.J. Rich, P.J. Sultan, and M.J. Keskinen (2001), Response of the equatorial ionosphere in the South

Atlantic region to the great magnetic storm of July 15, 2000, *Geophys. Res. Lett.* **28**, 18, 3577-3580, DOI: 10.1029/2001GL013259.

Bilitza, D. (2001), International Reference Ionosphere 2000, *Radio Sci.* **36**, 2, 261-275, DOI: 10.1029/2000RS002432.

Bilitza, D. (2002), Ionospheric models for radio propagation studies. **In:** W.R. Stone (ed.), *The Review of Radio Science 1999-2002: Advances in 3G Mobile Communications, Cryptography and Computer Security, EMC for Integrated Circuits, Remote Sensing, Radio Astronomy and More*, IEEE Press, Piscataway, 625-680.

Bilitza, D., L.A. McKinnell, B. Reinisch, and T. Fuller-Rowell (2011), The international reference ionosphere today and in the future, *J. Geod.* **85**, 12, 909-920, DOI: 10.1007/s00190-010-0427-x.

Bradley, P.A. (1995), PRIME (Prediction Regional Ionospheric Modelling over Europe), COST Action, Vol. 238, Final Report, Commission of the European Communities, Brussels.

Buonsanto, M.J. (1999), Ionospheric storms – A review, *Space Sci. Rev.* **88**, 3, 563-601, DOI: 10.1023/A:1005107532631.

Cander, Lj.R. (2003), Toward forecasting and mapping ionospheric space weather under the COST actions, *Adv. Space Res.* **31**, 4, 957-964, DOI: 10.1016/S0273-1177(02)00793-7.

Cander, Lj.R. (2006), Why bother with ionospheric prediction and forecasting? **In:** *Manual. European Digital Upper Atmosphere Server, DIAS*, 40-42.

Cander, Lj.R. (2008), Ionospheric research and space weather services, *J. Atmos. Sol.-Terr. Phys.* **70**, 15, 1870-1878, DOI: 10.1016/j.jastp.2008.05.010.

Cander, Lj.R. (2015), Forecasting foF2 and MUF(3000)F2 ionospheric characteristics – A challenging space weather frontier, *Adv. Space Res.* **56**, 9,1973-1981, DOI: 10.1016/j.asr.2015.06.013.

Cander, Lj.R., and S.J. Mihajlovic (2005), Ionospheric spatial and temporal variations during the 29-31 October 2003 storm, *J. Atmos. Sol.-Terr. Phys.* **67**, 12, 1118-1128, DOI: 10.1016/j.jastp.2005.02.020.

Cander, Lj.R., R. Bamford, and J.G. Hickford (2003), Nowcasting and forecasting the foF2, MUF(3000)F2 and TEC based on empirical models and real-time data. **In:** *Proc. 12th Int. Conf. on Antennas and Propagation, 31 March – 3 April 2003,* Conf. Publ. No. 491, Vol. 1, 139-142, DOI: 10.1049/cp:20030035.

Coffey, H.E., and E.H. Erwin (2001), When do the geomagnetic aa and Ap indices disagree? *J. Atmos. Sol.-Terr. Phys.* **63**, 5, 551-556, DOI: 10.1016/S1364-6826(00)00171-1.

Dick, M.I., M.F. Levy, Lj.R. Cander, I. Kutiev, and P. Muhtarov (1999), Short-term ionospheric forecasting over Europe. **In:** *IEE Nat. Conf. on Antennas and Progation, 31 March – 1 April 1999, York, UK*, 105-107, DOI: 10.1049/cp:19990025.

Dominici, P., B. Zolesi, and Lj.R. Cander (1988), Preliminary results concerning atmospheric gravity waves deduced from f_oF_2 large-scale oscillations, *Phys. Scripta* **37**, 3, 516-522, DOI: 10.1088/0031-8949/37/3/041.

Eastwood, J.P. (2008), The science of space weather, *Philos. Trans. Roy. Soc. A* **366**, 1884, 4489-4500, DOI: 10.1098/rsta.2008.0161.

Fox, M.W., and L.F. McNamara (1988), Improved world-wide maps of monthly median foF2, *J. Atmos. Terr. Phys.* **50**, 12, 1077-1086, DOI: 10.1016/0021-9169(88)90096-7.

Fuller-Rowell, T.J., M.V. Codrescu, R.G. Roble, and A.D. Richmond (1997), How does the thermosphere and ionosphere react to a geomagnetic storm? **In:** B.T. Tsurutani, W.D. Gonzales, Y. Kamide, and J.K. Arballo (eds.), *Magnetic Storms*, Geophysical Monograph, Vol. 98, AGU, Washington, 203-225, DOI: 10.1029/GM098p0203.

Fuller-Rowell, T.J., M.C. Codrescu, and P. Wilkinson (2000a), Quantitative modeling of the ionospheric response to geomagnetic activity, *Ann. Geophys.* **18**, 7, 766-781, DOI: 10.1007/s00585-000-0766-7.

Fuller-Rowell, T.J., E. Araujo-Pradere, and M.V. Codrescu (2000b), An empirical ionospheric storm-time correction model, *Adv. Space Res.* **25**, 1, 139-146, DOI: 10.1016/S0273-1177(99)00911-4.

Fuller-Rowell, T., E. Araujo-Pradere, C. Minter, M. Codrescu, P. Spencer, D. Robertson, and A.R. Jacobsen (2006), US-TEC: A new data assimilation product from the Space Environment Center characterizing the ionospheric total electron content using real-time GPS data, *Radio Sci.* **41**, 6, RS6003, DOI: 10.1029/2005RS003393.

Galkin, I.A., B.W. Reinisch, X. Huang, and D. Bilitza (2012), Assimilation of GIRO data into a real-time IRI, *Radio Sci.* **47**, 4, RS0L07, DOI: 10.1029/2011RS004952.

Hanbaba, R. (1999), Improved quality of service in ionospheric telecommunication systems planning and operation, COST Action 251 Final Report, Space Research Centre, Warsaw, Poland.

Hapgood, M.A. (2011), Towards a scientific understanding of the risk from extreme space weather, *Adv. Space. Res.* **47**, 12, 2059-2072, DOI: 10.1016/j.asr.2010.02.007.

Hocke, K., and K. Schlegel (1996), A review of atmospheric gravity waves and travelling ionospheric disturbances: 1982-1995, *Ann. Geophys.* **14**, 9, 917-940, DOI: 10.1007/s00585-996-0917-6.

Huang, X., B.W. Reinisch, and D. Bilitza (2001), IRI in Windows environment, *Adv. Space Res.* **27**, 1, 127-131, DOI: 10.1016/S0273-1177(00)00148-4.

ITU-R (1997), ITU-R Recommendations, Vol. 1997, P Series – Part 1, International Telecommunication Union, Geneva, Switzerland.

Ivanov-Kholodny, G.S., and A.V. Mikhailov (1986), *The Prediction of Ionospheric Conditions*, Geophysics and Astrophysics Monographs, D. Reidel Publ. Co., Dordrecht.

Jakowski, N., S. Schlüter, and E. Sardon (1999), Total electron content of the ionosphere during the geomagnetic storm on 10 January 1997, *J. Atmos. Sol.-Terr. Phys.* **61**, 3-4, 299-307, DOI: 10.1016/S1364-6826(98)00130-8.

Jakowski, N., C. Mayer, M.M. Hoque, and V. Wilken (2011), Total electron content models and their use in ionosphere monitoring, *Radio Sci.* **46**, 6, RS0D18, DOI: 10.1029/2010RS004620.

Kamide, Y. (2000), From discovery to prediction of magnetospheric processes, *J. Atmos. Sol.-Terr. Phys.* **62**, 17-18, 1659-1668, DOI: 10.1016/S1364-6826(00)00118-8.

Mannucci, A.J., O.P. Verkhoglyadova, B.T. Tsurutani, X. Meng, X. Pi, C. Wang, G. Rosen, E. Lynch, S. Sharma, A. Ridley, W. Manchester, B. Van Der Holst, E. Echer, and R. Hajra (2015), Medium-range thermosphere-ionosphere storm forecasts, *Space Weather* **13**, 3, 125-129, DOI: 10.1002/2014SW001125.

Matsushita, S. (1959), A study of the morphology of ionospheric storms, *J. Geophys. Res.* **64**, 3, 305-321, DOI: 10.1029/JZ064i003p00305.

McNamara, L.F., G.J. Bishop, and J.A. Welsh (2011), Analog ionospheric forecasts: Space weather forecasts by analogy with previous events, *Radio Sci.* **46**, 1, RS1002, DOI: 10.1029/2010RS004399.

Mendillo, M. (2006), Storms in the ionosphere: Patterns and processes for total electron content, *Rev. Geophys.* **44**, 4, RG4001, DOI: 10.1029/2005RG000193.

Mikhailov, A.V., L. Perrone, and N.V. Smirnova (2012), Two types of positive disturbances in the daytime mid-latitude F2-layer: Morphology and formation mechanisms, *J. Atmos. Sol.-Terr. Phys.* **81-82**, 59-75, DOI: 10.1016/j.jastp.2012.04.003.

Millward, G.H., H. Rishbeth, T.J. Fuller-Rowell, A.D. Aylward, S. Quegan, and R.J. Moffett (1996), Ionospheric F2 layer seasonal and semiannual variations, *J. Geophys. Res.* **101**, A3, 5149-5156, DOI: 10.1029/95JA03343.

Mukhtarov, P., B. Andonov, and D. Pancheva (2013), Global empirical model of TEC response to geomagnetic activity, *J. Geophys. Res.* **118**, 10, 6666-6685, DOI: 10.1002/jgra.50576.

Musman, S., G. Mader, and C. Everett Dutton (1998), Total electron content changes in the ionosphere during the January 10, 1997 disturbance, *Geophys. Res. Lett.* **25**, 15, 3055-3058, DOI: 10.1029/98GL51785.

Namgaladze, A.A., M. Förster, and R.Y. Yurik (2000), Analysis of the positive ionospheric response to a moderate geomagnetic storm using a global numerical model, *Ann. Geophys.* **18**, 4, 461-477, DOI: 10.1007/s00585-000-0461-8.

Pezzopane, M., M. Pietrella, A. Pignatelli, B. Zolesi, and L.R. Cander (2011), Assimilation of autoscaled data and regional and local ionospheric models as input sources for real-time 3-D International Reference Ionosphere modeling, *Radio Sci.* **46**, 5, RS5009, DOI: 10.1029/2011RS004697.

Pezzopane, M., M. Pietrella, A. Pignatelli, B. Zolesi, and Lj.R. Cander (2013), Testing the three-dimensional IRI-SIRMUP-P mapping of the ionosphere for disturbed periods, *Adv. Space Res.* **52**, 10, 1726-1736, DOI: 10.1016/j.asr.2012.11.028.

Pietrella, M. (2012), A short-term ionospheric forecasting empirical regional model (IFERM) to predict the critical frequency of the F2 layer during moderate, disturbed, and very disturbed geomagnetic conditions over the European area, *Ann. Geophys.* **30**, 2, 343-355, DOI: 10.5194/angeo-30-343-2012.

Pietrella, M., and L. Perrone (2008), A local ionospheric model for forecasting the critical frequency of the F2 layer during disturbed geomagnetic and iono-

spheric conditions, *Ann. Geophys.* **26**, 2, 323-334, DOI: 10.5194/angeo-26-323-2008.

Piggott, W.R., and K. Rawer (1972), U.R.S.I. handbook of ionogram interpretation and reduction, Report UAG-23, National Oceanic and Atmospheric Administration, Boulder, Colorado.

Prölss, G.W. (1995), Ionospheric F-region storms. **In:** H. Volland (ed.), *Handbook of Atmospheric Electrodynamics*, Vol. 2, CRC Press, Boca Raton, 195-248.

Reilly, M.H., F.J. Rhoads, J.M. Goodman, and M. Singh (1991), Updated climatological model predictions of ionospheric and HF propagation parameters, *Radio Sci.* **26**, 4, 1017-1024, DOI: 10.1029/91RS00583.

Rishbeth, H. (1991), F-region storms and thermospheric dynamics, *J. Geomagn. Geoelectr.* **43**, Suppl. 1, 513-524, DOI: 10.5636/jgg.43.Supplement1_513.

Rishbeth, H., I.C.F. Müller-Wodarg, L. Zou, T.J. Fuller-Rowell, G.H. Millward, R.J. Moffett, D.W. Idenden, and A.D. Aylward (2000), Annual and semi-annual variations in the ionospheric F2-layer: II. Physical discussion, *Ann. Geophys.* **18**, 8, 945-956, DOI: 10.1007/s00585-000-0945-6.

Smithtro, C.G., and J.J. Sojka (2005), Behavior of the ionosphere and thermosphere subject to extreme solar cycle conditions, *J. Geophys. Res.* **110**, A8, A08306, DOI: 10.1029/2004JA010782.

Strangeways, H.J., I. Kutiev, Lj.R. Cander, S. Kouris, V. Gherm, D. Marin, B. De La Morena, S. Eleri Pryse, L. Perrone, M. Pietrella, S. Stankov, L. Tomasik, E. Tulunay, Y. Tulunay, N. Zernov, and B. Zolesi (2009), Near-Earth space plasma modelling and forecasting, *Ann. Geophys.* **52**, 3-4, 255-271, DOI: 10.4401/ag-4579.

Vijaya Lekshmi, D., N. Balan, S. Tulasi Ram, and J.Y. Liu (2011), Statistics of geomagnetic storms and ionospheric storms at low and mid latitudes in two solar cycles, *J. Geophys. Res.* **116**, A11, A11328, DOI: 10.1029/2011JA017042.

Wilkinson, P., J. Wu, J. Du, and Y.-J. Wang (2001), Real-time total electron content estimates using the International Reference Ionosphere, *Adv. Space Res.* **27**, 1, 123-126, DOI: 10.1016/S0273-1177(00)00147-2.

Zolesi, B., and L.R. Cander (2014), *Ionospheric Prediction and Forecasting*, Springer Geophysics, Springer, Berlin Heildelberg, DOI 10.1007/978-3-642-38430-1.

Zolesi, B., Lj.R. Cander, and G. De Franceschi (1993), Simplified ionospheric regional model for telecommunication applications, *Radio Sci.* **28**, 4, 603-612, DOI: 10.1029/93RS00276.