



Properties of a Finite Stochastic Cellular Automaton Toy Model of Earthquakes

Mariusz BIAŁECKI

Institute of Geophysics, Polish Academy of Sciences, Warszawa, Poland
e-mail: bialecki@igf.edu.pl

Abstract

Finite version of Random Domino Automaton – a recently proposed toy model of earthquakes – is investigated in detail. Respective set of equations describing stationary state of the FRDA is derived and compared with infinite case. It is shown that for a system of large size, these equations are coincident with RDA equations. We demonstrate a non-existence of exact equations for size $N \geq 5$ and propose appropriate approximations, the quality of which is studied in examples obtained within the framework of Markov chains.

We derive several exact formulas describing properties of the automaton, including time aspects. In particular, a way to achieve a quasi-periodic like behaviour of RDA is presented. Thus, based on the same microscopic rule – which produces exponential and inverse-power like distributions – we extend applicability of the model to quasi-periodic phenomena.

Key words: stochastic cellular automaton, avalanches, cellular automata – exact solutions, toy model of earthquakes, Markov chains.

1. INTRODUCTION

The Random Domino Automaton (Białecki and Czechowski 2013, 2014) is a stochastic cellular automaton with avalanches. It was introduced as a toy model

of earthquakes, but can be also regarded as an substantial extension of 1-D forest-fire model proposed by Drossel and Schwabl (1992), see also Drossel *et al.* (1993) and Malamud *et al.* (1998).

A remarkable feature of the RDA is the explicit one-to-one relation between details of the dynamical rules of the automaton (represented by rebound parameters μ_i/ν defined in cited article and also below) and the produced stationary distribution n_i of clusters of size i , which implies a distribution of avalanches. It has already been shown how to reconstruct details of the “microscopic” dynamics from the observed “macroscopic” behaviour of the system (Białecky and Czechowski 2013, Białecky 2013).

As a field of application of RDA we studied a possibility of constructing the Ito equation from a given time series and – in a broader sense – applicability of Ito equation as a model of natural phenomena. For RDA – which plays a role of a fully controlled stochastic natural phenomenon – the relevant Ito equation can be constructed in two ways: derived directly from equations of RDA and by a histogram method from time series generated by RDA. Then these two results are compared and investigated in Czechowski and Białecky (2012a,b).

Note that the set of equations of the RDA in a special limit case reduces to the recurrence, which leads to known integer sequence – the Motzkin numbers, which establishes a new, remarkable link between the combinatorial object and the stochastic cellular automaton (Białecky 2012).

In the present paper a finite version of Random Domino Automaton is investigated. The mathematical formulation in finite case is precise and the presented results clarify which formulas are exact and allow to estimate approximations we impose in infinite case presented in Białecky and Czechowski (2013). We also show, that equations of finite RDA can reproduce results of Białecky and Czechowski (2013), when the size N of the system is increasing and distributions satisfy an additional assumption ($n_i \rightarrow 0$ for large i).

On the other hand, a time evolution of Finite RDA can exhibit a periodic-like behaviour (the assumption $n_i \rightarrow 0$ for large i is violated), which is a novel property. Thus, based on the same microscopic rules, depending on a choice of parameters of the model, a wide range of properties is possible to obtain. In particular, such behaviour is interesting in the context of recurrence parameters of earthquakes, see, *e.g.*, Weatherley (2006) and Parsons (2008). For other simple periodic-like models, see Tejedor *et al.* (2008) and Vazquez-Prada *et al.* (2002).

The finite case makes an opportunity to employ Markov chains techniques to analyse RDA. Investigating the automaton in Markov chains framework we arrive at several novel conclusions, in particular related to expected waiting times for some specified behaviour.

This article completes and substantially extends previous studies of RDA on the level of mathematical structure. We analyse properties of the automaton, including those related to time evolution, as a preparation for further prospective comparisons with natural phenomena, including earthquakes. An application of the model to the universal distribution of recurrence time for earthquakes is the topic of the forthcoming paper.

The plan of the article is as follows. In Section 2 we define the finite RDA and in Section 3 we present respective equations for finite RDA. In Section 4 we will specify them for some chosen cases. In Section 5 we will shortly describe Markov chains setting and describe time aspects of FRDA. Several examples are presented in Section 6. The last Section 7 contains conclusions and remarks. In the Appendix we show non existence of exact equations for RDA of size $N \geq 5$.

2. FINITE RDA

The rules for Finite Random Domino Automaton (Białecki and Czechowski 2013) are as follows. We assume:

- space is 1-dimensional and discrete – consists of N cells;
- periodic boundary conditions (the last cell is adjacent to the first one);
- cell may be in one of two states: empty or occupied by a single ball;
- time is discrete and in each time step an incoming ball hits one arbitrarily chosen cell (each cell is equally probable).

The state of the automaton changes according to the following rule:

- if the chosen cell is empty, it becomes occupied with probability ν ; with probability $(1 - \nu)$ the incoming ball is rebounded and the state remains unchanged;
- if the chosen cell is occupied, the incoming ball provokes an avalanche with probability μ (it removes balls from hit cell and from all adjacent cells); with probability $(1 - \mu)$ the incoming ball is rebounded and the state remains unchanged.

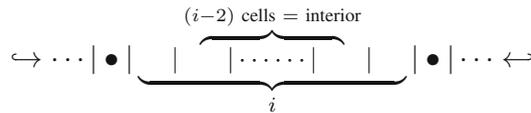
The parameter ν is assumed to be a constant but the parameter μ is allowed to be a function of size of the hit cluster. The way in which the probability of removing a cluster depends on its size strongly influences evolution of the system and leads to various interesting properties, as presented in the following sections. We note in advance that in fact there is only one effective parameter μ/ν which affects properties of the automaton. Changing of μ and ν proportionally in a sense corresponds to a rescaling of time unit.

A diagram shown below presents an automaton of size $N = 12$, with three clusters (of size 1, 2, and 4) in time t . An incoming ball provokes an relaxation of the size 2 – the size of an avalanche is equal to the number of cells changing their state. Thus in time $t+1$ there are two clusters (of size 1 and 4).

We emphasise, the above result is exact – no correlations were neglected. Its form is directly analogous to the respective formula in Białeck and Czechowski (2013).

3.2 Balance of the total number of clusters

Gain. A new cluster (can be of size 1 only) can be created in the interior of empty cluster of size ≥ 3 .



If the empty cluster is of size N , then each cell is in interior. Summing up contributions for all empty clusters, the probability is

$$\sim \sum_{i=3}^{N-1} \nu \left(\frac{i-2}{i} \right) \frac{n_i^0 i}{N} + \nu n_N^0, \tag{5}$$

which can take a form (for $N \geq 3$)

$$\sim \nu(1 - \rho) - 2\nu \frac{n_R}{N} + \nu \frac{n_1^0}{N}. \tag{6}$$

Loss. Two ways contribute: joining a cluster with another one and removing a cluster due to avalanche.

Joining of two clusters can occur if there exists an empty cluster of length 1 between them. The exception is when the empty 1-cluster is the only one empty cluster, and the system consists of a single cluster of length $N - 1$. Hence, the probability of joining two clusters is

$$\sim \nu \left(\frac{n_1^0}{N} - n_{N-1} \right). \tag{7}$$

The probability of avalanche is just

$$\sim \sum_{i=1}^N \mu_i \frac{n_i i}{N}. \tag{8}$$

By gathering these terms one obtains equation for the balance of the total number of clusters n

$$N(1 - \rho) - \sum_{i=1}^N \frac{\mu_i}{\nu} n_i i + n_{N-1} = 2n_R. \tag{9}$$

Again we emphasise that the above result is exact – no correlations were neglected. The finite size of the system is reflected by the appearance of $(2n_R - n_{N-1})$ instead of $2n$ in the respective formula in Białeck and Czechowski (2013).

3.3 Balance of n_i s

Loss. There are two modes.

(a) Enlarging – an empty cluster on the edge of an i -cluster becomes occupied. There are two such empty clusters except for the case when system contains a single cluster of length $N - 1$. Hence, the respective rates are

$$\sim 2\nu \frac{n_i}{N} \quad i = 1, \dots, N - 2, \tag{10}$$

$$\sim \nu \frac{n_{N-1}}{N} \quad i = N - 1. \tag{11}$$

(b) Relaxation rate for any $i = 1, \dots, N$ is given by

$$\sim \mu_i \frac{in_i}{N}. \tag{12}$$

Gain. Again, there are two modes.

(a) Enlarging. For $N \geq 3$, there are the following rates depending on the size i of the cluster

$$\sim \nu(1 - \rho) - 2\nu \frac{n_R}{N} + \nu \frac{n_1^0}{N}, \quad i = 1, \tag{13}$$

$$\sim 2\nu \frac{n_{i-1}}{N} \alpha_{i-1}^E, \quad 2 \leq i \leq N - 1, \tag{14}$$

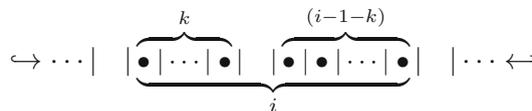
$$\sim \nu \frac{n_{N-1}}{N}, \quad i = N, \tag{15}$$

where $\alpha_E(i)$ is a probability that the size of empty cluster adjacent to the i -cluster is bigger than 1. It is clear that

$$\alpha_{N-2}^E = 1 \quad \text{and} \quad \alpha_{N-1}^E = 0. \tag{16}$$

Formula 15 does not have a factor 2, because there is only one empty cluster (of size 1).

(b) Merger of two clusters up to the cluster of size i . Two clusters: one of size $k \in \{1, 2, \dots, (i - 2)\}$ and the other of size $((i - 1) - k)$ will be combined if the ball fills an empty cell between them.



The probability is proportional to the number of empty 1-clusters between k -cluster and $(i - 1 - k)$ -cluster,

$$\sim \nu \frac{n_1^0}{N} \gamma_i^E \quad 3 \leq i \leq N - 1, \tag{17}$$

where γ_i^E is a probability of such merger. For $i = N$ there is a single cluster in the lattice (there are no two clusters to merge) – filling the gap between ends of $(N - 1)$ -cluster is already considered in (a).

By gathering the terms, one obtains

$$n_1 = \frac{1}{\frac{\mu_1}{\nu} + 2} (N(1 - \rho) - 2n_R + n_1^0), \tag{18}$$

$$n_2 = \frac{1}{2\frac{\mu_2}{\nu} + 2} 2n_1\alpha_1^E, \tag{19}$$

$$n_i = \frac{1}{\frac{\mu_i}{\nu}i + 2} (2n_{i-1}\alpha_{i-1}^E + n_1^0\gamma_i^E), \tag{20}$$

$$n_{N-1} = \frac{1}{\frac{\mu_{N-1}}{\nu}(N-1) + 1} (2n_{N-2} + n_1^0\gamma_{N-1}^E), \tag{21}$$

$$n_N = \frac{1}{\frac{\mu_N}{\nu}N} n_{N-1}, \tag{22}$$

where $3 \leq i \leq (N - 2)$.

The last Eq. 22 has a simple explanation. The state with all cells being occupied (corresponding to n_N) can be achieved only from the state with a single empty cell (corresponding to n_{N-1}) with probability $\nu(1/N)$. On the other hand, the automaton leaves the state with all cells being occupied with probability μ_N .

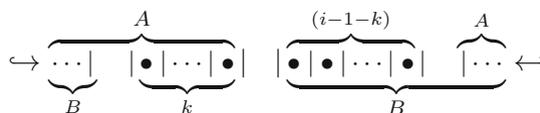
Note that Eqs. 18 and 22 are exact. Correlations in the systems reflect an appearance of multipliers α_i^E and γ_i^E . Their values depend on possible configurations of states of the automaton. As shown in the Appendix, for $N \geq 5$ exact formulas for α_i^E and γ_i^E as functions of n_i s do not exist. Hence, it is necessary to propose approximated formulas.

A mean field type approximation for α_i^E is

$$\alpha_i^E \approx \alpha_i^A = \left(1 - \frac{n_1^0}{\sum_{k=1}^{N-i} n_k^0} \right). \tag{23}$$

For a given cluster of size i , the probability of the appearance of an empty cluster of size 1 is calculated as proportional to the number of empty 1-clusters divided by the sum of the numbers of all empty clusters with size not exceeding $N - 1$, because there is no room for larger.

When merger of two clusters up to a cluster of size i is considered, the room denoted by A is of size $(N - 2 - (i - 1 - k))$ and the room denoted by B is of size $(N - 2 - k)$ – see a diagram below.



Hence a mean field type approximation for γ_i^E is of the form

$$\gamma_i^E \approx \gamma_i^A = \sum_{k=1}^{i-2} \left(\frac{n_k}{\sum_{j=1}^{N-(i-1-k+2)} n_j} \cdot \frac{n_{i-1-k}}{\sum_{j=1}^{N-(k+2)} n_j} \right). \tag{24}$$

It is also instructive to consider another approximation

$$\gamma_i^E \approx \gamma_i^{AR} = \sum_{k=1}^{i-2} \left(\frac{n_k}{n_R} \cdot \frac{n_{i-1-k}}{n_R} \right). \tag{25}$$

Section 6 contains quantitative estimation of proposed approximations. Comparison of this approximation with exact results for small sizes N is discussed in Section 7.

3.4 Correspondence to a limit $N \rightarrow \infty$

In the paper Białecki and Czechowski (2013) an assumption of independence of clusters was considered. To have it adequate, it is required that there are no limitations in space, like those encountered when formulas 23 and 24 were considered. For systems that are large enough, *i.e.*, when $N \rightarrow \infty$, an empty cluster adjacent to a given i -cluster can be of any size, and thus

$$\alpha_i^E \approx \alpha = \left(1 - \frac{n_1^0}{\sum_{k=1}^{\infty} n_k^0} \right) = \left(1 - \frac{n_1^0}{n} \right). \tag{26}$$

This is consistent with the requirement that $n_i \rightarrow 0$ when $i \rightarrow \infty$, which is required to have moments of the n_i s convergent. Similarly,

$$\gamma_i^E \approx \gamma(i) = \sum_{k=1}^{i-2} \left(\frac{n_k}{n} \cdot \frac{n_{i-1-k}}{n} \right). \tag{27}$$

These formulas substituted into 18-20 give the respective set of equations considered in Białecki and Czechowski (2013). The same reasoning can be applied to balance equations. The form of Eq. 4 is left unchanged under the limit. For Eq. 9, $(2n_R - n_{N-1}) \rightarrow 2n$, and it becomes of the form presented in Białecki and Czechowski (2013).

4. SPECIAL CASES

For a fixed form of the rebound parameters the equations describing the automaton can be written in more specific form, in particular the balance Eqs. 4 and 9, as well as formulas for average cluster size

$$\langle i \rangle = \frac{\sum_{i=1}^N n_i i}{\sum_{i=1}^N n_i} = \frac{N\rho}{n_R + n_N} \tag{28}$$

and average avalanche size

$$\langle w \rangle = \frac{\sum_{i=1}^N \mu_i n_i i^2}{\sum_{i=1}^N \mu_i n_i i}. \tag{29}$$

We emphasize, these formulas are exact – correlations are encountered. We consider three special cases investigated in detail and illustrated by examples below.

■ $\mu = \text{const.}$

For $\mu = \text{const.}$ and $\nu = \text{const.}$ Eq. 4 is of the form

$$(1 - \rho) = \frac{1}{N} \frac{\mu}{\nu} \sum_{i=1}^N n_i i^2. \tag{30}$$

and Eq. 9

$$N(1 - \rho(1 + \frac{\mu}{\nu})) + n_{N-1} = 2n_R. \tag{31}$$

Also formulas for $\langle i \rangle$ and $\langle w \rangle$ are simplified only a little.

■ $\mu(i) = \delta/i$ where $\theta = \delta/\nu = \text{const.}$

Equation 4 is of the form

$$(1 - \rho) = \theta\rho, \tag{32}$$

hence the density is given by remarkably neat (end exact) formula

$$\rho = \frac{1}{1 + \theta}. \tag{33}$$

Note that there is no dependence on the size of the system N ; for $N \rightarrow \infty$ it remains the same.

Equation 9 can be written as

$$N \frac{\theta}{1 + \theta} = (2 + \theta)n_R, \tag{34}$$

where we use Eqs. 22 and 33. Hence the formula for n_R is of the form

$$n_R = N \frac{\theta}{(\theta + 1)(\theta + 2)} \tag{35}$$

in direct analogy with n in $N \rightarrow \infty$ case (Bialecki and Czechowski 2013). Thus, n_R plays the role of n , as indicated also in balance of n_1 Eq. 18. The

formula for n is

$$n = n_R + n_N = N \frac{\theta(1 + \varepsilon)}{(\theta + 1)(\theta + 2)} \quad \text{where} \quad \varepsilon = \frac{n_N}{n_R}. \tag{36}$$

The average cluster size is given by

$$\langle i \rangle = \frac{1}{1 + \varepsilon} \left(1 + \frac{2}{\theta} \right). \tag{37}$$

The average avalanche size is equal to the average cluster size

$$\langle w \rangle = \langle i \rangle, \tag{38}$$

because each cluster has the same probability to be removed from the lattice.

The above formulas are exact (include correlations) and possess well defined limit when $N \rightarrow \infty$ (which imply $\varepsilon \rightarrow 0$). Note also that variables ρ and n_R depend on single parameter θ . Formulas with dependence on θ can be rewritten as functions of density ρ .

■ $\mu(i) = \eta/i^2$ and $\chi = \sigma/\nu = \text{const.}$

Equation 4 is of the form

$$N(1 - \rho) = \chi(n_R + n_N). \tag{39}$$

Equation 9 can be written as

$$N(1 - \rho) = 2n_R - \chi \frac{1}{N} n_N + \chi \sum_{i=1}^N \frac{n_i}{i} \tag{40}$$

where Eq. 22 is used, namely $n_{N-1} = \chi \frac{1}{N} n_N$.

The average cluster size is

$$\langle i \rangle = \chi \frac{\rho}{1 - \rho}, \tag{41}$$

and the average avalanche size is

$$\langle w \rangle = \frac{1}{\left(1 + \frac{\sigma}{N}\right) - (1 - \sigma) \frac{2}{\chi}}, \tag{42}$$

where

$$\sigma = \frac{n_N}{n_R + n_N} = \frac{\varepsilon}{\varepsilon + 1}. \tag{43}$$

Note that also these formulas are exact.

5. FINITE RDA AS A MARKOV CHAIN

5.1 General settings

Finite Random Domino Automaton is a Markov chain, hence we use the respective framework to compute examples in an exact way and derive several important formulas for time aspects of the evolution of the system.

In general, for the lattice of size N there are 2^N states, because each of N cells may be empty or occupied. For $N = 4$, an exemplary state is

$$\hookrightarrow | \quad | \bullet | \quad | \bullet | \hookleftarrow$$

where the assumed periodic boundary conditions are depicted by hook-arrows.

For periodic boundary conditions it is irrelevant to distinguish between states which differ by a translation only. Hence, for example, we consider the following states equivalent:

$$\hookrightarrow | \quad | \bullet | \quad | \bullet | \bullet | \hookleftarrow \quad \equiv \quad \hookrightarrow | \bullet | \quad | \bullet | \quad | \bullet | \hookleftarrow$$

Thus the states a_i are defined up to translational equivalence (see Tables 1 and 2). The label numbers are assigned to the states, as shown in tables. Further reduction of the number of states using reflections can be done, but it is not very efficient procedure. We do not perform it, keeping symmetrical states separate. They deliver a simple computational check – their probabilities are necessarily equal.

Such space of states for the finite random domino automaton is irreducible, aperiodic and recurrent. The transition matrix \mathbf{P} is defined by

$$[\mathbf{P}]_{ij} = \text{probability of transition } a_i \longrightarrow a_j \tag{44}$$

For $N = 3$ the transition matrix is of the form

$$\mathbf{P} = \begin{pmatrix} 1 - \nu & \nu & 0 & 0 \\ \frac{\mu_1}{3} & 1 - \frac{\mu_1}{3} - \frac{2\nu}{3} & \frac{2\nu}{3} & 0 \\ \frac{2\mu_2}{3} & 0 & 1 - \frac{2\mu_2}{3} - \frac{\nu}{3} & \frac{\nu}{3} \\ \mu_3 & 0 & 0 & 1 - \mu_3 \end{pmatrix} \tag{45}$$

Table 1

States for the size of the lattice $N = 3$

State number	Example	Multiplicity	Contributing to
1	$\hookrightarrow \quad \quad \hookleftarrow$	1	n_3^0
2	$\hookrightarrow \quad \quad \bullet \hookleftarrow$	3	n_1, n_2^0
3	$\hookrightarrow \quad \bullet \bullet \hookleftarrow$	3	n_2, n_1^0
4	$\hookrightarrow \bullet \bullet \bullet \hookleftarrow$	1	n_3

Table 2

States for the size of the lattice $N = 5$

State number	Example	Multiplicity	Contributing to
1	$\leftrightarrow \leftarrow$	1	n_5^0
2	$\leftrightarrow \bullet \leftarrow$	5	n_1, n_4^0
3	$\leftrightarrow \bullet \bullet \leftarrow$	5	n_2, n_3^0
4	$\leftrightarrow \bullet \bullet \leftarrow$	5	n_1, n_1^0, n_2^0
5	$\leftrightarrow \bullet \bullet \bullet \leftarrow$	5	n_2, n_2^0
6	$\leftrightarrow \bullet \bullet \bullet \leftarrow$	5	n_1, n_2, n_1^0
7	$\leftrightarrow \bullet \bullet \bullet \bullet \leftarrow$	5	n_4, n_1^0
8	$\leftrightarrow \bullet \bullet \bullet \bullet \bullet \leftarrow$	1	n_5

where entries are found from analysis of transition probability of all possible states a_i (see Table 1).

For $N = 5$ the transition matrix \mathbf{P} is

$$\frac{1}{5} \begin{pmatrix} 5-5\nu & 5\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mu_1 & 5-\mu_1-4\nu & 2\nu & 2\nu & 0 & 0 & 0 & 0 & 0 \\ 2\mu_2 & 0 & 5-2\mu_2-3\nu & 0 & 2\nu & \nu & 0 & 0 & 0 \\ 0 & 2\mu_1 & 0 & 5-2\mu_1-3\nu & \nu & 2\nu & 0 & 0 & 0 \\ 3\mu_3 & 0 & 0 & 0 & 5-3\mu_3-2\nu & 0 & 2\nu & 0 & 0 \\ 0 & 2\mu_2 & \mu_1 & 0 & 0 & 5-2\mu_2-\mu_1-2\nu & 2\nu & 0 & 0 \\ 4\mu_4 & 0 & 0 & 0 & 0 & 0 & 0 & 5-4\mu_4-\nu & \nu \\ 5\mu_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5-5\mu_5 \end{pmatrix} \quad (46)$$

The stationary distribution is given by

$$v \cdot P = v. \quad (47)$$

The number of states increase rapidly with N : for $N = 6$ there are 14 states, for $N = 7$ there are 20 states, and for $N = 10$ there are 108 states. The number of states for any N is bigger than $2^N/N$, because translational symmetry of states is at most N , but there are always states with smaller symmetry, like empty state and fully occupied state. Thus practical usage of Markov chain settings for calculations is rather limited without a systematic procedure of obtaining a transition matrix \mathbf{P} , which is generally nontrivial. This is one of the reasons for developing more handy framework presented in Białecki and Czechowski (2013) and here. On the other hand, the framework of Markov chains is very handy in order to investigate various properties of the system, as presented below.

5.2 Expected time of return

As the system evolves, it hits a given state many times. Here we consider expected value of the time of return from state with density $\rho = 0$ to itself and next from the state with $\rho = 1$ to itself. Starting from state 1 (state with $\rho = 0$) the next state (different from state 1) contains a single 1-cluster only. This state – denoted by label 2 – has density $\rho = 1/N$. Expected time for this change is $1/\nu$.

Let τ_i be the expected time to hit state 1 starting in state i . Then $\tau_1 = 0$ and for $i \neq 1$

$$\begin{aligned} \tau_i &= \mathbb{E}(\text{time to hit 1} \mid \text{start in } i) \\ &= 1 + \sum_k p_{ik} \mathbb{E}(1|k) = 1 + \sum_k p_{ik} \tau_k, \end{aligned} \tag{48}$$

where $\mathbb{E}(1|k) = \mathbb{E}(\text{time to hit 1} \mid \text{start in } k)$. After solving this system of equations, the return time is

$$t_{1 \rightarrow 1} = 1/\nu + \tau_2. \tag{49}$$

Similarly, for state with $\rho = 1$ (state L) the next state (different from state L) is the empty state (with $\rho = 0$) and

$$t_{L \rightarrow L} = 1/\mu_N + \hat{\tau}_1, \tag{50}$$

where $\hat{\tau}_1$ is the expected time to hit state L starting in state 1. The respective equation to determine $\hat{\tau}_i$ for $i \neq L$ reads

$$\hat{\tau}_i = 1 + \sum_k p_{ik} \hat{\tau}_k, \tag{51}$$

and obviously $\hat{\tau}_L = 0$. Note that the expected time $t_{L \rightarrow L}$ is equal to expected time of return from state 1 to state 1 through state L :

$$t_{L \rightarrow L} = t_{1 \rightarrow L \rightarrow 1}. \tag{52}$$

The expected time between two consecutive avalanches is

$$t_{\text{av}} = \frac{\langle w \rangle + 1}{1 - P_r}, \tag{53}$$

where P_r is the probability that the incoming ball is rebounded both form empty or occupied cell (see Table 3):

$$P_r = (1 - \rho)(1 - \nu) + \frac{1}{N} \sum_{i=1}^N n_i i (1 - \mu_i). \tag{54}$$

Note that $(1 - P_r)$ is equal to the sum of the probabilities of triggering an avalanche and the probability that an empty cell becomes occupied, hence

$$P_r + \frac{1}{N} \sum_{i=1}^N \mu_i n_i i + (1 - \rho)\nu = 1. \tag{55}$$

Formula 53 can be derived as follows. In time between two consecutive avalanches, on average, $(t_{av}(1 - P_r) - 1)$ cells become occupied in the system – it receives one ball per a time step, part of them are rebounded and one ball triggers the avalanche. An avalanche is reducing the number of occupied cells by $\langle w \rangle$. These two quantities compensate for each other, giving 53.

On the other hand, the expected time between two consecutive avalanches is equal to the inverse of the probability of triggering an avalanche

$$t_{av} = \left(\frac{1}{N} \sum_{i=1}^N \mu_i n_i i \right)^{-1}. \tag{56}$$

Both expressions given in 53 and 56 are equal to each other.

5.3 Frequency distribution of avalanches

The probability of states obtained from condition 47 allows to determine the distribution of frequency of avalanches. The frequency f_i of the avalanche of size i is given by the sum of products of probabilities v_k of state k and respective transition probability p_{kj} to the appropriate states j for all states that transition $k \rightarrow j$ produce the avalanche of size i .

For example, for $N = 5$, as can be seen in Table 2, transitions $2 \rightarrow 1$, $4 \rightarrow 2$, and $6 \rightarrow 3$ result in an avalanche of size 1, transitions $3 \rightarrow 1$ and $6 \rightarrow 2$ give an avalanche of size 2, transition $5 \rightarrow 1$ of size 3, $7 \rightarrow 1$

Table 3
Probabilities of all four possibilities occurring in a single time step during evolution of the automaton

Probability of	Value
rebound – occupied cell	$\frac{1}{N} \sum_{i=1}^N n_i i (1 - \mu_i)$
rebound – empty cell	$(1 - \rho)(1 - \nu)$
occupation of empty cell	$(1 - \rho)\nu$
triggering an avalanche	$\frac{1}{N} \sum_{i=1}^N \mu_i n_i i$

of size 4, and 8 \rightarrow 1 of size 5. Hence

$$f_1 = v_2\mu_1/5 + v_42\mu_1/5 + v_6\mu_1/5, \tag{57}$$

$$f_2 = v_32\mu_2/5 + v_62\mu_2/5, \tag{58}$$

$$f_3 = v_53\mu_3/5, \tag{59}$$

$$f_4 = v_74\mu_4/5, \tag{60}$$

$$f_5 = v_8\mu_5, \tag{61}$$

where respective p_{kj} are taken from transition matrix 46.

The average time t_i between two avalanches of size i is given by

$$t_i = 1/f_i, \tag{62}$$

in particular, for a maximum size N

$$t_{L \rightarrow L} = t_N. \tag{63}$$

The average time between (any) consecutive avalanches given by formula 56 may be also calculated as

$$t_{av} = \left(\sum_{i=1}^N t_i^{-1} \right)^{-1}, \tag{64}$$

because the probability of an avalanche of any size is just a sum of probabilities of all possible avalanches. In this way one can calculate also average time between any two consecutive avalanches of prescribed size – for example, size 4 and 5 (or any other subset of possible sizes).

6. EXAMPLES

Below we present several examples to illustrate properties of finite RDA as well as to demonstrate application of the schemes outlined above.

■ $N = 3$

This is the simplest non-trivial, warm-up example. For $N = 3$ the general results – *i.e.*, for arbitrary μ_1, μ_2, μ_3 , and ν – can be calculated explicitly. Usage of equations 18-22 leads to exact results as presented below (see Appendix). The same can be also obtained from Markov chains framework. Equations 4, 9, and 22 give

$$n_1 = 3 \left(\frac{\mu_2}{\nu} + \frac{1}{2} \right) / D, \tag{65}$$

$$n_2 = 3/D, \tag{66}$$

$$n_3 = \left(\frac{\nu}{\mu_3} \right) / D, \tag{67}$$

where

$$D = \frac{11}{2} + \frac{\mu_1}{2\nu} + 5\frac{\mu_2}{\nu} + \frac{\mu_1\mu_2}{\nu^2} + \frac{\nu}{\mu_3}.$$

From inspecting of Table 1 it is evident that $n_1^0 = n_2$, $n_2^0 = n_1$, and $n_3^0 = 1 - n_1 - n_2 - n_3$ (all possibilities sum up to 1), hence

$$n_3^0 = \left(1 + \frac{\mu_1}{2\nu} + 2\frac{\mu_2}{\nu} + \frac{\mu_1\mu_2}{\nu^2}\right) / D. \tag{68}$$

General formulas for expected times of return are

$$t_{1 \rightarrow 1} = \frac{1}{\nu} \left(1 + \frac{2\nu^2 + 9\mu_3\nu + 6\mu_2\mu_3}{\mu_3(\mu_1 + 2\nu)(2\mu_2 + \nu)}\right), \tag{69}$$

$$t_{L \rightarrow L} = \frac{1}{\nu} \left(\frac{\nu}{\mu_3} + \frac{11}{2} + \frac{\mu_1}{2\nu} + 5\frac{\mu_2}{\nu} + \frac{\mu_1\mu_2}{\nu^2}\right). \tag{70}$$

The ratio $t_{L \rightarrow L} / t_{1 \rightarrow 1}$ is

$$t_{L \rightarrow L} / t_{1 \rightarrow 1} = \frac{1}{2} \left(\frac{\mu_1}{\nu} + 2\right) \left(\frac{2\mu_2}{\nu} + 1\right). \tag{71}$$

Note that it does not depend on μ_3 . If the probability of triggering an avalanche of size 1 and 2 is small comparing to the probability of occupation of an empty cell (*i.e.*, $\mu_1/\nu \approx 0$ and $\mu_2/\nu \approx 0$) then $t_{L \rightarrow L} \approx t_{1 \rightarrow 1}$. The next stage after the lattice is fully occupied is the empty state; hence, if these two average waiting times are comparable, then they occur with comparable frequency. That means quasi-periodic like behaviour of the system: within average time $11/2\nu$ the lattice become fully occupied, then the triggering of an avalanche of maximal size N occurs with average waiting time $1/\mu_3$. The same can be observed for bigger sizes N .

Figure 1 and Table 4 present examples of three types of dependence of rebound parameters on size i of clusters considered in Section 4, each having the same density $\rho = 1/2$ (with 8 digits accuracy). To obtain this density we put for these three cases $\mu/\nu = 0.444118$ ($\mu = 0.444118$, $\nu = 1$), $\theta = 1$ ($\delta = 1$, $\nu = 1$), and $\chi = 2.113440690$ ($\eta = 1$, $\nu = 1/2.113440690$), respectively. As seen from Fig. 1 it is possible to obtain flat distribution for $\mu_i = \delta/i$ – on that background, differences between the cases are clearly visible: $\mu_i = \text{const.}$ discriminate the existence of large clusters fostering large avalanches; the opposite is for $\mu_i = \sigma/i^2$. Average cluster size and avalanche size data presented in Table 4 confirms this conclusion.

■ $N = 5$

For $N = 5$ it is impossible to write down exact equations 18-22 depending on values of n_i s only – see Appendix for details. The case can be solved as a Markov process, but obtained general formulas are relatively complicated.

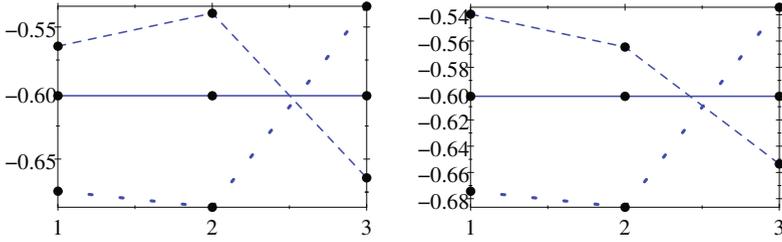


Fig. 1. Plot of the Log_{10} of n_i s (left) and n_i^0 s (right) versus i for $N = 3$ in three cases: $\mu_i = \text{const.}$ (dashed line), $\mu_i = \delta/i$ (solid line), and $\mu_i = \sigma/i^2$ (dotted line). Rebound parameters are chosen to have density $\rho = 1/2$ in all cases (see main text for respective values).

Table 4

Average cluster size $\langle i \rangle$ and average avalanche size $\langle w \rangle$ for three different rebound parameters

	$\mu_i = \text{const.}$	$\mu_i = \delta/i$	$\mu_i = \sigma/i^2$
$\langle i \rangle$	1.9281668	2	2.1134407
$\langle w \rangle$	2.2516538	2	1.7226121

Note: density $\rho = 1/2$, the size of the lattice $N = 3$.

In this example we investigate properties of the system with density $\rho = 1/4$. Figure 2 and Table 5 compare results in three cases: $\mu/\nu = 16257/10000$ the density $\rho = 0.2500003184$, for $\theta = 3$ the density $\rho = 0.25$ exactly, and $\chi = 5.95682$ gives the density $\rho = 0.2500004527$.

General expressions for return times $t_{1 \rightarrow 1}$ and $t_{L \rightarrow L}$ as well as their ratio are relatively complex. Note that the return times – except of the dependence on

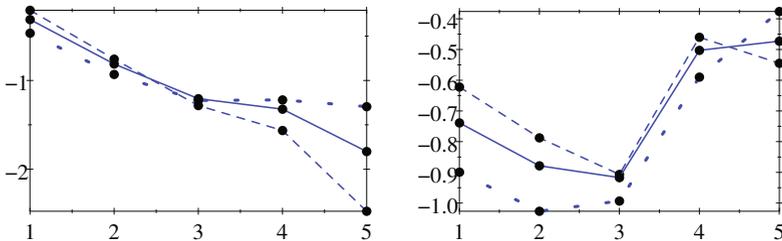


Fig. 2. Plot of the Log_{10} of n_i s (left) and n_i^0 s (right) versus i for $N = 5$ in three cases: $\mu_i = \text{const.}$ (dashed line), $\mu_i = \delta/i$ (solid line), and $\mu_i = \sigma/i^2$ (dotted line). Rebound parameters are chosen to have density $\rho = 1/4$ in all cases (see main text for respective values).

Table 5

Average cluster size $\langle i \rangle$ and average avalanche size $\langle w \rangle$ for three different rebound parameters

	$\mu_i = \text{const.}$	$\mu_i = \delta/i$	$\mu_i = \sigma/i^2$
$\langle i \rangle$	1.427017126	1.632218845	1.985611461
$\langle w \rangle$	1.845355789	1.632218845	1.41360643

Note: density $\rho = 1/4$, the size of the lattice $N = 5$.

t – are proportional to $1/\nu$. Below we specify the ratio $t_{L \rightarrow L}/t_{1 \rightarrow 1}$ in three cases: for $\mu_i = \text{const.}$, where $t = \mu/\nu$, it is equal to

$$\frac{24t^6 + 154t^5 + 413t^4 + 586t^3 + 467t^2 + 182t + 24}{24t^2 + 54t + 24}, \tag{72}$$

for $\mu_i = \delta/i$, where $\delta = \text{const.}$ and $t = \delta/\nu$, it is equal to

$$\frac{4t^6 + 40t^5 + 169t^4 + 395t^3 + 550t^2 + 432t + 144}{56t^2 + 192t + 144}, \tag{73}$$

and for $\mu_i = \sigma/i^2$, where $\sigma = \text{const.}$ and $t = \sigma/\nu$, is

$$\frac{2t^6 + 39t^5 + 304t^4 + 1232t^3 + 2840t^2 + 3744t + 2304}{496t^2 + 2208t + 2304}. \tag{74}$$

In each case the ratio is a rational function of t , which is equal to 1 for $t = 0$ and asymptotically $\sim t^4$ for $t \rightarrow \infty$. A generalisation of this observation is a conjecture formulated in Section 7. A comparison of these ratios is presented in left part of Fig. 3.

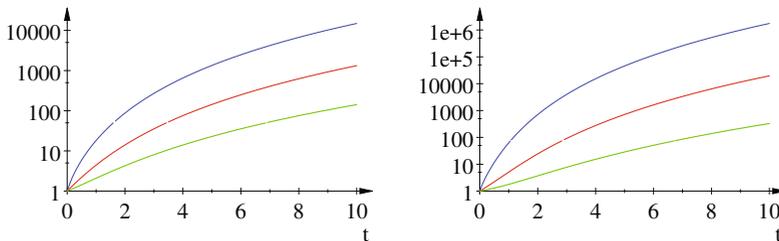


Fig. 3. Ratio of return times $t_{L \rightarrow L}/t_{1 \rightarrow 1}$ for $N = 5$ (left) and $N = 7$ (right) for three cases: $\mu_i = \text{const.}$ (blue line on the top), $\mu_i = \delta/i$ (red line in the middle), and $\mu_i = \eta/i^2$ (green line on the bottom). Parameter t is equal to μ/ν , δ/ν , and η/ν , respectively.

Table 6

Coefficient $R = t_{L \rightarrow L} / t_{1 \rightarrow 1}$
for three different rebound parameters (see main text for details)

	$\mu_i = \text{const.}$	$\mu_i = \delta/i$	$\mu_i = \sigma/i^2$
$R = t_{L \rightarrow L} / t_{1 \rightarrow 1}$	≈ 52.212	≈ 35.441	≈ 34.801

Note: density for all cases $\rho = 1/4$, the size of the lattice $N = 5$.

Table 6 shows that for the cases discussed above with average density $\rho = 1/4$ the highest value of R is for $\mu_i = \text{const.}$ and the smallest for $\mu_i = \sigma/i^2$ (not much different from the value for $\mu_i = \delta/i$).

Average waiting times t_i for an avalanche of size i can be also found, but are long. The average time between any two consecutive avalanches is

$$t_{\text{av}} = \frac{4t^5 + 48t^4 + 237t^3 + 603t^2 + 762t + 360}{\nu t(4t^4 + 36t^3 + 121t^2 + 168t + 72)}, \quad (75)$$

where $t = \delta/\nu$. All these quantities are proportional to $1/\nu$. Figure 4 in the left panel presents waiting times t_i in for fixed density $\rho = 1/4$ in three cases mentioned above. There are no big differences both in character of dependence of t_i on i and also values of t_{av} do not differ much: for $\mu_i = \text{const.}$ average time is $t_{\text{av}} \approx 24.60$, for $\mu_i = \delta/i$ it is ≈ 21.76 , and for $\mu_i = \sigma/i^2$ it is ≈ 18.85 . (Choosing parameters to have density $\rho = 1/4$ we put $\nu = 1/10$ for all cases.)

Average waiting times $t_i, i = 1, \dots, 5$ in the case $\mu_i = \delta/i$ for various densities are shown in the right panel of Fig. 4. For small densities the maximal waiting time t_i is for $i = 5$, while for larger densities the maximum is for

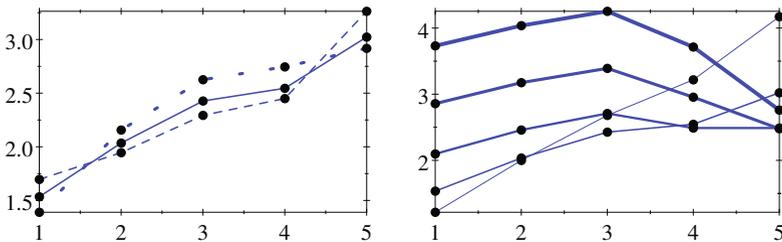


Fig. 4: (left) Plot of Log_{10} of t_i s versus i for three rebound parameters for fixed density $\rho = 1/4$ for $N = 5$. Three cases: $\mu_i = \text{const.}$ (dashed line), $\mu_i = \delta/i$ (solid line), and $\mu_i = \sigma/i^2$ (dotted line). Rebound parameters are chosen to have density $\rho = 1/4$ in all cases (see main text for respective values); (right) plot of Log_{10} of t_i s versus i for various densities for rebound parameter of the form $\mu_i = \delta/i$ for $N = 5$. Densities are chosen as $\frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{9}{10}$; thinner line corresponds to smaller density.

$i = 3$. Average waiting times range from ≈ 13.57 for $\rho = 1/10$ through ≈ 21.76 , ≈ 50.22 , ≈ 145.01 for densities $1/4$, $1/2$, $3/4$, respectively, up to ≈ 441.60 for density $\rho = 9/10$. (Again $\nu = 1/10$ for all cases.)

■ $N = 7$

For $N = 7$ we investigate properties of the system with the density $\rho = 3/4$. Parameters are chosen as follows: $\mu = 1$, $\nu = 173024/10000$ gives the density $\rho = 0.7500001621$, $\theta = 1/3$ gives $\rho = 3/4$ exactly, and $\mu = 1$, $\nu := 1000000/1578886$ gives $\rho = 0.7500002817$. Distributions of clusters are presented in Fig. 5 and average cluster and avalanche sizes in Table 7. Again differences in distributions n_i are not big, but average avalanche size differs significantly between considered cases.

The novel property visible in the figure is that the highest probability is for the cluster of maximal size $i = N$. Thus, the system prefers merging clusters for high density.

A comparison of the ratios of return times $R = t_{L \rightarrow L} / t_{1 \rightarrow 1}$ is presented in the right panel of Fig. 3, while formulas are presented in the Appendix. In each case the ratio is a rational function of t , which is equal to 1 for $t = 0$ and

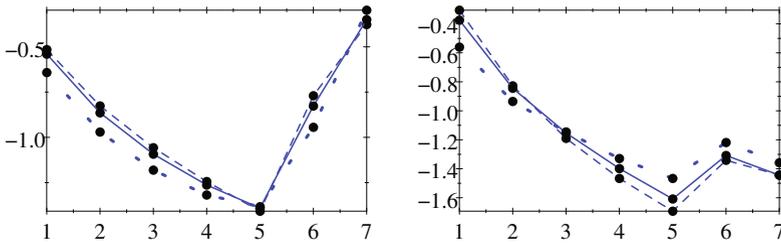


Fig. 5. Plot of the Log_{10} of n_i s (left) and n_i^0 s (right) versus i for $N = 7$ in three cases: $\mu_i = \text{const.}$ (dashed line), $\mu_i = \delta/i$ (solid line), and $\mu_i = \sigma/i^2$ (dotted line). Rebound parameters are chosen to have density $\rho = 3/4$ in all cases (see main text for respective values).

Table 7

Average cluster size $\langle i \rangle$ and average avalanche size $\langle w \rangle$ for three different rebound parameters

	$\mu_i = \text{const.}$	$\mu_i = \delta/i$	$\mu_i = \sigma/i^2$
$\langle i \rangle$	4.274328495	4.385371765	4.736665115
$\langle w \rangle$	5.767461682	4.385371765	2.671314107

Note: density $\rho = 3/4$, the size of the lattice $N = 7$.

Table 8

Coefficient $R = t_{L \rightarrow L} / t_{1 \rightarrow 1}$
 for three different rebound parameters (see main text for details)

	$\mu_i = \text{const.}$	$\mu_i = \delta/i$	$\mu_i = \sigma/i^2$
$R = t_{L \rightarrow L} / t_{1 \rightarrow 1}$	≈ 1.4844	≈ 1.6887	≈ 2.7001

Note: density for all cases $\rho = 3/4$, the size of the lattice $N = 7$.

asymptotically $\sim t^6$ for $t \rightarrow \infty$, which supports a conjecture formulated in Section 7. Table 8 shows that for the cases discussed above, with average density $\rho = 3/4$, the highest value of the ratio R is for $\mu_i = \delta/i$ and the smallest for $\mu_i = \text{const.}$ (which does not differ much from the value for $\mu_i = \delta/i$). This is an opposite order comparing to the case with $\rho = 1/5$ for $N = 5$ considered above. Thus, for higher densities the automaton prefers more periodic-like behaviour when it is relatively easier to trigger large avalanches.

The size $N = 7$ is big enough to notice how the actual density of the system (possible values are $0, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}, 1$) is distributed for various average densities. Results are shown in Fig. 6. For small densities, like $\rho = 0.2$, the maximum is for small i , that means that large densities and large avalanches are rare. Then, when the density increases, the bell-like shape distribution appears and its maximum is shifted to the bigger values. Next, for densities like

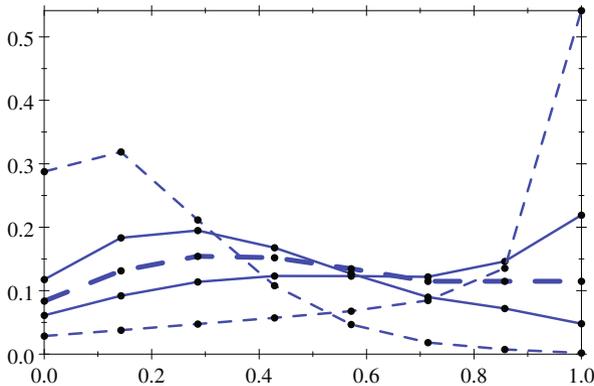


Fig. 6. Probability distributions of actual density of the system for the case $\mu_i/\nu = \theta/i$ for various average densities of 0.2 (dashed line), 0.4 (solid line), 0.5 (thick dashed line), 0.6 (solid line), 0.8 (dashed line) – the respective parameters are $\theta = 4, \frac{3}{2}, 1, \frac{2}{3}, \frac{1}{4}$. Smaller average density corresponds to the higher probability rate for density equal to 0, *i.e.*, starts from the top on the left side (and to the lower probability rate for density equal to 1). The size of the system is $N = 7$.

Cases with constants equal to 1

As a first set we consider three cases with the minimal possible rebounds factors, *i.e.*, we put all constants equal to 1. Cases with $\mu_i = 1$, $\mu_i = 1/i$ and $\mu_i = 1/i^2$ with $\nu = 1$ are presented in Fig. 7 and Table 9.

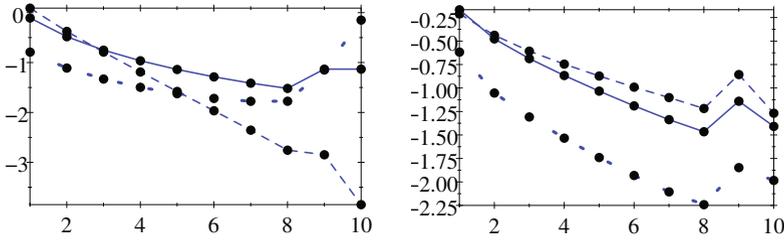


Fig. 7. Plot of the Log_{10} of n_i s (left) and n_i^0 (right) for $N = 10$ in three cases: $\mu_i = \text{const.}$ (dashed line), $\mu_i = \delta/i$ (solid line), and $\mu_i = \sigma/i^2$ (dotted line).

Table 9

Three cases: $\mu = 1$, $\delta = 1$, and $\sigma = 1$ (and always $\nu = 1$)
for the size of the lattice $N = 10$

	$\mu_i = 1$	$\mu_i = 1/i$	$\mu_i = 1/i^2$
ρ	0.3076370614	0.5	0.8822697788
$\langle i \rangle$	1.5985438	2.872872532	7.493995763
$\langle w \rangle$	2.250583644	2.872872532	3.725820785
$p_{88} - 2p_{89}$	0.00865	0.01899	0.01868
$\frac{p_{88}}{\alpha_1^E - \alpha_1^A}$	0.00909	0.066784	0.19468
$\frac{\alpha_4^E - \alpha_4^A}{\alpha_4^E}$	0.08795	0.07428	0.09239
$\frac{\gamma_5^E - \gamma_5^A}{\gamma_5^E}$	-0.01141	-0.28676	-0.89662
$\frac{\gamma_5^E - \gamma_5^{AR}}{\gamma_5^E}$	-0.00110	-0.06842	-0.13292
$\frac{\gamma_9^E - \gamma_9^A}{\gamma_9^E}$	0.35717	0.29878	0.17045
$\frac{\gamma_9^E - \gamma_9^{AR}}{\gamma_9^E}$	0.48970	0.62573	0.71428
$t_{1 \rightarrow 1}$	18.51	25.59	96.28
$t_{L \rightarrow L} / t_{1 \rightarrow 1}$	379.61	5.2988	1.4606

Three different rebound parameter types result in various average density values, and hence different distributions. In all cases, the assumption of independence of clusters is well satisfied; the respective error Δ does not exceed 2%. An approximation for α_4^E is less than 10% for all cases, but α_1^E strongly depends on the case (in fact it depends on density, as will be seen below). Approximation formulas for γ_E perform in diversified way – γ_{AR} is better for mid i terms, while γ_A is better for big i terms. Nevertheless, both cases provide rather roughly appropriate values. These examples also suggest that for higher densities the system exhibits a periodic-like evolution.

Large densities

In order to investigate evolution of the system with high average density (and strong deviations in actual density) we consider case $\mu = \text{const.}$ with $\mu_1 = 1/100$ and $\nu = 1$, which gives the density $\rho \approx 0.91$, and case δ/i with $\mu_1 = 4673077001/5 * 10^{10} \approx 0.093$ and $\nu = 1$ to obtain the same density (with 10 digits accuracy) for comparison. Also we consider case of σ/i^2 with $\mu_1 = 1/10$ and $\nu = 1$ which gives the density $\rho \approx 0.99$. The results are presented in Fig. 8 and Table 10.

Plots of respective distributions for $\mu_i \sim 1/i$ and $\mu_i \sim 1$ are overlapping each other. For relatively small size $N = 10$, fixing the average density of the system strongly determines distributions, making the dependence on re-

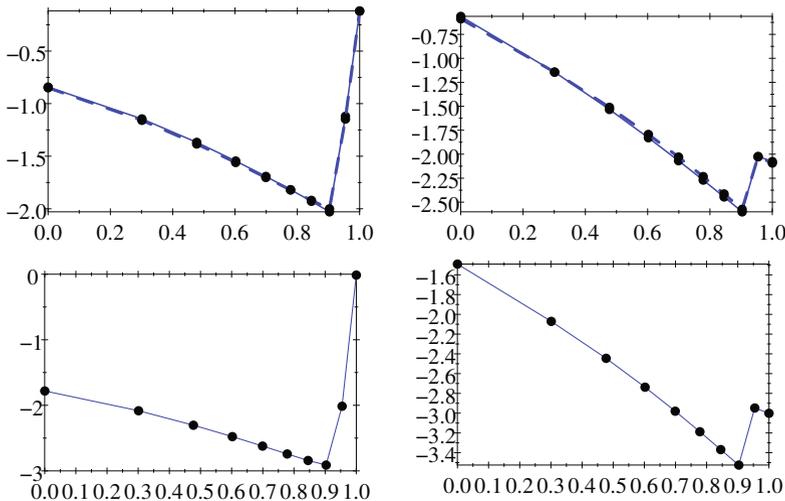


Fig. 8. Plot of the Log_{10} of n_i s and n_i^0 versus $\text{Log}_{10}(i)$ for $N = 10$ in two cases: $\mu_i = \text{const.}$ (solid line) and $\mu_i = \delta/i$ (dashed line) – upper panels, and in case $\mu_i = \delta/i^2$ – lower panels.

Table 10

Three cases with “large” ρ for the size of the lattice $N = 10$
(see main text for details)

	$\mu_i \sim 1$	$\mu_i \sim 1/i$	$\mu_i \sim 1/i^2$
ρ	0.9145269069	0.9145269069	0.9897960692
$\langle i \rangle$	7.764600567	7.805017612	9.700144892
$\langle w \rangle$	9.346154002	7.805017612	8.33021261
$p_{88} - 2p_{89}$	0.00161	0.00464	0.00191
$\frac{p_{88}}{\alpha_1^E - \alpha_1^A}$	0.30898	0.28421	0.31377
$\frac{\alpha_1^E}{\alpha_4^E - \alpha_4^A}$	-0.01511	0.01520	0.00330
$\frac{\alpha_4^E}{\gamma_5^E - \gamma_5^A}$	-0.96763	-0.93825	-1.05131
$\frac{\gamma_5^E}{\gamma_5^E - \gamma_5^{AR}}$	-0.15298	-0.14354	-0.14789
$\frac{\gamma_5^E}{\gamma_9^E - \gamma_9^A}$	0.42749	0.39218	0.38965
$\frac{\gamma_9^E}{\gamma_9^E - \gamma_9^{AR}}$	0.80211	0.78888	0.80019
$\frac{\gamma_9^E}{t_{1 \rightarrow 1}}$	119.18	123.60 ^a	1002.44
$t_{L \rightarrow L} / t_{1 \rightarrow 1}$	1.1066	1.1339	1.0277

^a) The system stays in fully occupied state $1/\mu_{10} \approx 107, 5$, which is longer than 100 as in $\mu = \text{const.}$ case (previous column). The respective average times for filling up the lattice are ≈ 19 and ≈ 16 .

bound parameters not essential. Their influence becomes more visible for larger sizes N of the lattice. In the case of high densities, the system just spend much of it's time being fully occupied.

For high densities, the assumption of independence of clusters is well satisfied; the respective error Δ does not exceed 0.5%. An approximation for α_4^E is fairly good ($\approx 1.5\%$ or less), but α_1^E has only accuracy $\approx 30\%$. Approximation formula for γ_{AR} is much better for mid i terms (though giving only $\approx 15\%$ accuracy), while γ_A is better for large i terms ($\approx 40\%$). Thus, for high density cases the proposed set of equations for n_i s does not reproduce actual distribution. Note, however, that there are other exact equations valid for any density.

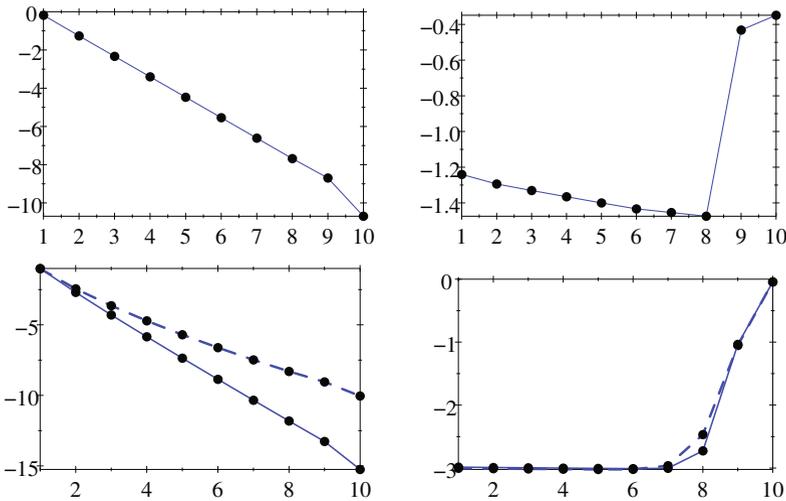


Fig. 9. Plot of the Log_{10} of n_i s and n_i^0 for $N = 10$ in case $\mu_i = \text{const.}$ – upper line, and in two cases: $\mu_i = \delta/i$ (solid line), and $\mu_i = \delta/i^2$ (dashed line) – lower line.

The parameter $t_{L \rightarrow L}/t_{1 \rightarrow 1}$ for $\mu_i \sim 1/i$ case is bigger than for $\mu_i \sim 1$ case (both cases have the same “big” density), which agrees with the results for $N = 7$ with $\rho = 3/4$ presented in Table 8.

Small densities

To present system behaviour in small average density we choose $\mu_1 = 1$ and $\nu = 1/10$ for case $\mu = \text{const.}$ – it gives density $\rho \approx 0.08$. Then for the remaining two cases we have the same density $\rho \approx 0.01$ (with 10 digits accuracy), with the following parameters: $\mu_1 = 1$ and $\nu = 50000000/4798952601 \approx 0.01$ – for case δ/i , and $\mu_1 = 1$ and $\nu = 1/100$ for case σ/i^2 . The results are presented in Fig. 9 and Table 11.

For small densities assumption of independence of clusters is well satisfied. In general, all proposed approximations are fairly good. An approximation for α_4^E is the worst; its accuracy is only $\approx 15\%$. As previously, approximation formula for γ_A is better than γ_{AR} for large i terms, but it appears that for mid i terms both formulas give almost the same values (because n_i s decrease rapidly). Thus, for small densities the set of equations for n_i s can be used to reproduce the actual distribution.

It is very improbable to find the lattice fully occupied for small average densities, which is reflected in high values of the parameter $R = t_{L \rightarrow L}/t_{1 \rightarrow 1}$. The parameter R for $\mu_i \sim 1/i$ case is bigger than its for $\mu_i \sim 1/i^2$ case (both cases have the same “small” density), which agrees with the results for $N = 5$ with $\rho = 1/4$ presented in Table 6.

Table 11

Three cases with “small” ρ for the size of the lattice $N = 10$
 (see main text for details)

	$\mu_i \sim 1$	$\mu_i \sim 1/i$	$\mu_i \sim 1/i^2$
ρ	0.0779280356	0.01031150521	0.01031150521
$\langle i \rangle$	1.09149321	1.02083788	1.041894016
$\langle w \rangle$	1.183235221	1.02083788	1.020408163
$p_{88} - 2p_{89}$	$6.9231 \cdot 10^{-5}$	$6.4609 \cdot 10^{-4}$	$3.5663 \cdot 10^{-3}$
$\frac{p_{88}}{\alpha_1^E - \alpha_1^A}$	$1.7483 \cdot 10^{-5}$	$2.7693 \cdot 10^{-7}$	$1.0585 \cdot 10^{-6}$
$\frac{\alpha_4^E}{\alpha_4^E - \alpha_4^A}$	0.14229	0.16313	0.16274
$\frac{\gamma_5^E - \gamma_5^A}{\gamma_5^E}$	0.0009069	0.0020519158	0.0040100
$\frac{\gamma_5^E - \gamma_5^{AR}}{\gamma_5^E}$	0.0009101	0.0020519262	0.0040124
$\frac{\gamma_9^E - \gamma_9^A}{\gamma_9^E}$	0.09615	0.01473	0.00606
$\frac{\gamma_9^E - \gamma_9^{AR}}{\gamma_9^E}$	0.12138	0.02356	0.03305
$t_{1 \rightarrow 1}$	22.32	106.35	110.57
$t_{L \rightarrow L} / t_{1 \rightarrow 1}$	2220903488.0 $\sim 2 \cdot 10^{10}$	$1.666292752 \cdot 10^{14}$ $\sim 2 \cdot 10^{14}$	9971770329.0 $\sim 1 \cdot 10^{11}$

7. CONCLUSIONS

In this article we investigated in detail a finite version of one-dimensional non-equilibrium dynamical system – Random Domino Automaton. It is a simple, slowly driven system with avalanches. The advantage of RDA (comparing to Drossel-Schwabl model) is the dependence of rebound parameters on the size of a cluster. This crucial extension allows for producing a wider class of distributions by the automaton, as well as leads to several exact formulas. Exponential type and inverse-power type distributions of clusters were studied in Bialecki and Czechowski (2013); the present work examines also V-shape distributions and quasi-periodic like behavior.

Detailed analysis of finite RDA, including finite size effects, extends and explains the previously obtained results for RDA. Moreover, we also analyzed approximations made when deriving equations for the stationary state of the automaton. This allows for the following conclusions.

The balance of ρ Eq. 4 and the balance of N Eq. 9 are exact – their forms incorporate all correlations present in the system. The first one has a form independent of the size of the lattice N , thus it is exactly the same as for RDA. The second one contains correction for finite size effect, namely a term $(2n_R - n_{N-1})$, which replaces the term $2n$ for RDA. When n_{N-1} and n_N are negligible, these two terms coincide. For finite RDA, balance of n_i s Eqs. 18-22 contains two extra equations, for $i = N - 1$ and $i = N$, comparing to the those for RDA. The first (for n_1) and the last (for n_N) are exact. Note that all those equations are written for rebound parameter $\mu = \mu(i)$ being a function of cluster size and ν being a constant.

The most remarkable special case is when $\mu = \delta/i$, when any cluster has the same probability to be removed as an avalanche independently of its size i . It appears that the system depends on a single parameter $\theta = \delta/\nu$, or equivalently, due to neat exact formula (Eq. 33)

$$\rho = \frac{1}{1 + \theta},$$

the properties of the system may be characterized by the value of the average density. Note that the above expression does not depends on the size N , and is the same as for RDA. This specialization leads to more neat formulas, like the equation for n_R (Eq. 35)

$$n_R = N \frac{\theta}{(\theta + 1)(\theta + 2)}.$$

Note again that it has the same form as for RDA, except that n is replaced by n_R ($n = n_R + n_N$). Summarizing, the model allows to derive a number of explicit dependencies, as shown in Sections 3 and 4.

The Random Domino Automaton defines a discrete time Markov process of order 1 and, in principle, may be solved exactly. However, it turns out that computations are fairly complex and exact formulas are long. Also, the exact numerical values are in the form of large numbers – in every considered example ($N = 3, 4, 5, 6, 7, 10$) significantly large prime numbers were encountered. For example, for the simplest possible rebound parameters ($\mu = 1$ and $\nu = 1$) the exact value of denominators of probabilities of states for $N = 10$ is a 65-digit integer. Its prime factorization contains a 56-digit integer, which cannot be simplified with numerators.

Nevertheless, Markov chains framework leads to interesting results concerning analysis of times of recurrence for specific states. A return time to the state with density $\rho = 1$ (Eq. 50)

$$t_{L \rightarrow L} = 1/\mu_N + \hat{\tau}_1,$$

consists of two parts: waiting of fully occupied lattice for triggering a maximal avalanche and “loading” time, when the lattice is filled up, respectively. If the average density of the system is small, the second time is very long. The formula is more interesting for systems with relatively large average density, when the “loading” time is comparable to waiting time for triggering the largest avalanche. Such a system exhibits a periodic like behavior. Dividing the waiting time $t_{L \rightarrow L}$ by the waiting time $t_{1 \rightarrow 1}$ (given by Eq. 49) one has the following measure of quasi-periodicity

$$R = t_{L \rightarrow L} / t_{1 \rightarrow 1} = t_{1 \rightarrow L \rightarrow 1} / t_{1 \rightarrow 1} .$$

If $R = 1$ then the system is periodic.

Several considered examples lead to the following conjecture concerning the coefficient R .

Conjecture. The ratio of return times $t_{L \rightarrow L} / t_{1 \rightarrow 1}$ as a function of t being the ratio of constants from rebound parameters (μ/ν , δ/ν , σ/ν) are rational functions of t , $f(t) = t_{L \rightarrow L} / t_{1 \rightarrow 1}$ with the following properties for any size N of the system

$$f(t = 0) = 1 , \tag{76}$$

$$\lim_{t \rightarrow \infty} \frac{f(t)}{t^{N-1}} = \text{const} . \tag{77}$$

The conjecture relates the size of the system N with asymptotic behavior of ratio of waiting times.

There are large fluctuations (variations of actual density) during the evolution of systems with relatively large average densities. If the system is likely to achieve a fully occupied state, the next state is an empty state, and the variations in density are maximal. Nevertheless, some parameters of stationary state (more precisely, statistically stationary state) satisfy exact equations, as shown above. For large average densities, the system fluctuates within the whole possible range, and cannot be thought of as having approximately stationary values during the evolution. This aspect is easy to be overlooked (see Paczuski and Bak 1993).

It is argued in the Appendix that no exact equations for n_i s exist for the size $N \geq 5$. Thus, to have compact equations for n_i s, some approximation formulas are proposed. The first general conclusion from the examples is that the approximations are acceptable for small densities, but for large densities the errors are substantial. The main reason is that for large densities correlations become more important and fluctuations makes actual values of the parameters substantially different from their stationary values, which are present in the formulas. These properties are particularly severe for small sizes of the system, where every avalanche changes the actual density considerably.

Table 12

Comparison of α_1^E with α_1^A for $N = 3, 4, 5, 6, 7$ and 10
for parameters $\mu = 1$ and $\nu = 1$

N	$\alpha_1^E = \frac{2n_2}{n_1}$	$\alpha_1^A = \left(1 - \frac{n_1^0}{n_R}\right)$	$\frac{\alpha_1^E - \alpha_1^A}{\alpha_1^E}$	ρ
3	1.33(3)	0.6	0.55	≈ 0.3462
4	0.66(6)	≈ 0.6316	≈ 0.053	0.32(32)
5	≈ 0.6829	≈ 0.6565	≈ 0.039	≈ 0.3139
6	≈ 0.685296	≈ 0.669232	≈ 0.023	≈ 0.3102
7	≈ 0.685523	≈ 0.675066	≈ 0.015	≈ 0.3086
10	≈ 0.685436	$\approx 0.679205^a$	≈ 0.0091	≈ 0.3076
4000	≈ 0.677	≈ 0.677	–	≈ 0.3076

^{a)} $(1 - n_1^0/n) \approx 0.679229$

Note: The last line presents value of α_1^A for $N = 4000$ obtained from simulations and equations (respectively) in Białecki and Czechowski (2013).

Table 12 presents a dependence of a relative error of α_1^E with respect to α_1^A on size N of the system. For larger N the accuracy of approximation is growing, which corresponds well with the remark in the last paragraph.

It can be noticed from the distributions of n_i s of examples presented above that all n_i s except of the last two (namely n_{N-1} and n_N) are placed on one “regular” curve, while the last two deviate from it. It may be regarded as a (correction of) finite size effect. Also in the respective set of Eqs. 18-22, the last two (for $i = N - 1$ and $i = N$) have a form different from the previous ones. Thus, neglecting the size restriction, which in fact ignores the last two equations, is justified when the deviations of the last two n_i s from the “regular” curve are not big. That happens for small densities.

It appears also that for index i in his middle range of values an approximation formula γ_{AR} works better than γ_A , in spite of the fact that it looks to be more rough approximation. For distribution of n_i s vanishing rapidly (*i.e.*, for small densities) both give comparable results.

All this justifies the form of equations for n_i s presented in Białecki and Czechowski (2013) as valid for small densities. A detailed examination of the RDA for large densities requires further investigations.

This article explores properties of FRDA in order prepare to modeling of real data. In this context, among others, formulas for waiting times can be used. We ephasize also a formula 53

$$t_{av} = \frac{\langle w \rangle + 1}{1 - P_r},$$

which relates the measure of scattering (dissipation) of balls P_r with the average size of avalanche $\langle w \rangle$ and average time between any two consecutive avalanches t_{av} , which are *a priori* measurable quantities.

The Random Domino Automaton proved to be a stochastic dynamical system with interesting mathematical structure. It may be viewed as extension of Drossel-Schwabl model, and we showed that this is a substantial generalization with a wide range of novel properties. We expect it can also be applied to natural phenomena, including earthquakes and forest-fires. This is our aim for the future work.

Appendix

Exact equations for $N = 3, 4$ and their non-existence for $N \geq 5$

For arbitrary size N , there are four exact equations: balance of ρ – Eq. 4, balance of n – Eq. 9, for n_1 – Eq. 18, and for n_N – Eq. 22.

Size $N = 3$. Equation for n_2 is of the form 21, namely

$$n_2 = \frac{1}{2\frac{\mu_1}{\nu} + 1} (2n_1\alpha_1^E + n_1^0\gamma_2^E).$$

In this case the only companion to single one-cluster is an empty two-cluster (see state 2 in Table 1), hence

$$\alpha_1^E = 1 \quad \text{and} \quad \gamma_2^E = 0.$$

Thus, we arrived at the exact form of the equation for n_2 .

Size $N = 4$. All states of the automaton (other states differ by shifts are identified) and their labels are presented in Table 13. Equation for n_2 is of the form 19

$$n_2 = \frac{2}{2\frac{\mu_2}{\nu} + 2} n_1\alpha_1^E,$$

where to α^E contributes only state 2, not state 4. Hence

$$\alpha_1^E = \frac{p_2}{p_2 + 2p_4} = \left(1 - \frac{2p_4}{n_1}\right) = \left(1 - \frac{n_1^0 - n_3}{n_1}\right),$$

Table 13

States for the size of the lattice $N = 4$

State number	Example	Multiplicity	Contributing to
1	$\hookrightarrow \leftarrow$	1	n_4^0
2	$\hookrightarrow \bullet \leftarrow$	4	n_1, n_3^0
3	$\hookrightarrow \bullet \bullet \leftarrow$	4	n_3, n_2^0
4	$\hookrightarrow \bullet \bullet \leftarrow$	2	n_1, n_1^0
5	$\hookrightarrow \bullet \bullet \bullet \leftarrow$	4	n_3, n_1^0
6	$\hookrightarrow \bullet \bullet \bullet \bullet \leftarrow$	1	n_4

where p_i is probability of state i . Thus α_1^E is expressed as function of n_i s and n_1^0 . The equation for n_3 is of the form 21

$$n_3 = \frac{1}{3\frac{\mu_3}{\nu} + 1} (2n_2\alpha_2^E + n_1^0\gamma_3^E).$$

In this case

$$\alpha_2^E = 1,$$

because only state 3 contributes. The state 4 (and not state 5) contributes to γ_3^E , therefore

$$\gamma_3^E = \frac{2p_4}{2p_4 + p_5} = \left(1 - \frac{p_5}{n_1^0}\right) = \left(1 - \frac{n_3}{n_1^0}\right).$$

This completes the task of writing exact equations for $N = 4$.

Size $N = 5$. States and their labels are presented in Table 2. In this case, the coefficients are as follows

$$\alpha_2^E = \frac{p_2 + p_4}{n_1}, \quad \gamma_3^E = \frac{p_4 - p_6 - n_4}{n_1^0},$$

$$\alpha_3^E = \frac{p_3}{n_2}, \quad \gamma_4^E = \frac{p_6 - p_4 - n_4}{n_1^0}.$$

Summing up the probabilities contributing to n_1^0, n_1 and n_2 one obtains

$$n_1^0 = p_4 + 2p_6 + n_4, \tag{78}$$

$$n_1 = p_2 + 2p_4 + p_6, \tag{79}$$

$$n_2 = p_3 + p_6. \tag{80}$$

The set cannot be solved for p_2, p_3, p_4, p_6 . Since there are no more equations for those coefficients, respective α s and γ s cannot be expressed as functions of n_i s only in an exact manner.

Sizes bigger than 5. An argument for non-existence of exact set of Eqs. 18-22, *i.e.*, non-existence of exact formulas for α_i^E s and γ_i^E s as functions of n_i s and n_1^0 is based on the same impossibility of solving equations as presented above.

An increase of size of a grid N by 1 results in an increase of the set of n_1, n_2, \dots by one and much bigger increase of the number of states. An analog of the set of Eqs. 78-80 will contain much more probabilities of states p_1, p_2, \dots , on the right hand side – there will be more states containing 1-clusters, 2-clusters and so on, and contributing to n_1, n_2, \dots , respectively. Thus, it is impossible to express those probabilities of states as functions of n_1^0, n_1, n_2, \dots only. As a consequence, there are no general exact formulas for α_i^E s and γ_i^E .

Acknowledgment. The author would like to express his gratitude to Professors Adam Doliwa and Maciej Wojtkowski for inspiring comments and discussions. This work was partially supported by National Science Centre, project 2012/05/B/ST10/00598 and partially supported within statutory activities No. 3841/E-41/S/2015 of the Ministry of Science and Higher Education of Poland.

References

- Bialecki, M. (2012), Motzkin numbers out of Random Domino Automaton, *Phys. Lett. A* **376**, 45, 3098-3100, DOI: 10.1016/j.physleta.2012.09.022.
- Bialecki, M. (2013), From statistics of avalanches to microscopic dynamics parameters in a toy model of earthquakes, *Acta Geophys.* **61**, 6, 1677-1689, DOI: 10.2478/s11600-013-0111-7.
- Bialecki, M., and Z. Czechowski (2010), On a simple stochastic cellular automaton with avalanches: simulation and analytical results. **In:** V. De Rubeis, Z. Czechowski, and R. Teisseyre (eds.), *Synchronization and Triggering: From Fracture to Earthquake Processes*, GeoPlanet: Earth and Planetary Sciences, Springer, Berlin Heidelberg, 63-75, DOI: 10.1007/978-3-642-12300-9_5.
- Bialecki, M., and Z. Czechowski (2013), On one-to-one dependence of rebound parameters on statistics of clusters: exponential and inverse-power distributions out of Random Domino Automaton, *J. Phys. Soc. Jpn.* **82**, 1, 014003, DOI: 10.7566/JPSJ.82.014003.
- Bialecki, M., and Z. Czechowski (2014), Random Domino Automaton: Modeling macroscopic properties by means of microscopic rules. **In:** R. Bialik, M. Majdański, and M. Moskalik (eds.), *Achievements, History and Challenges in*

- Geophysics*, GeoPlanet: Earth and Planetary Sciences, Springer, Int. Publ. Switzerland, 223-241, DOI: 10.1007/978-3-319-07599-0_13.
- Czechowski, Z., and M. Białecki (2012a), Three-level description of the domino cellular automaton, *J. Phys. A: Math. Theor.* **45**, 15, 155101, DOI: 10.1088/1751-8113/45/15/155101.
- Czechowski, Z., and M. Białecki (2012b), Ito equations out of domino cellular automaton with efficiency parameters, *Acta Geophys.* **60**, 3, 846-857, DOI: 10.2478/s11600-012-0021-0.
- Drossel, B., and F. Schwabl (1992), Self-organized critical forest-fire model, *Phys. Rev. Lett.* **69**, 11, 1629-1632, DOI: 10.1103/PhysRevLett.69.1629.
- Drossel, B., S. Clar, and F. Schwabl (1993), Exact Results for the One-Dimensional Self-Organized Critical Forest-Fire Model, *Phys. Rev. Lett.* **71**, 23, 3739-3742, DOI: 10.1103/PhysRevLett.71.3739.
- Malamud, B.D., G. Morein, and D.L. Turcotte (1998), Forest fires: an example of self-organized critical behavior, *Science* **281**, 5384, 1840-1842, DOI: 10.1126/science.281.5384.1840 .
- Paczuski, M., and P. Bak (1993), Theory of the one-dimensional forest-fire model, *Phys. Rev. E* **48**, 5, R3214-R3216, DOI:10.1103/PhysRevE.48.R3214.
- Parsons, T. (2008), Monte Carlo method for determining earthquake recurrence parameters from short paleoseismic catalogs: Example calculations for California, *J. Geophys. Res.* **113**, B3, B03302, DOI: 10.1029/2007JB004998.
- Tejedor, A., S. Ambroj, J.B. Gomez, and A.F. Pacheco (2008), Predictability of the large relaxations in a cellular automaton model, *J. Phys. A: Math. Theor.* **41**, 37, 375102, DOI: 10.1088/1751-8113/41/37/375102.
- Vazquez-Prada, M., A. Gonzalez, J.B. Gomez, and A.F. Pacheco (2002), A minimalist model of characteristic earthquakes, *Nonlin. Process. Geophys.* **9**, 5-6, 513-519, DOI: 10.5194/npg-9-513-2002.
- Weatherley, D. (2006), Recurrence interval statistics of cellular automaton seismicity models, *Pure Appl. Geophys.* **163**, 9, 1933-1947, DOI: 10.1007/s00024-006-0105-3 .

Received 20 March 2015

Received in revised form 13 May 2015

Accepted 22 May 2015



Surface Deformation due to Loading of a Layered Elastic Half-space: Constructing the Solution for a General Polygonal Load

Michael BEVIS¹, Ernian PAN², Hao ZHOU¹, Feng HAN²,
and Richard ZHU²

¹School of Earth Sciences, Ohio State University, Columbus, Ohio, USA
e-mail: mbevis@osu.edu

²Department of Civil Engineering, University of Akron, Akron, Ohio, USA

Abstract

We describe an algorithm for rapidly computing the surface displacements induced by a general polygonal load on a layered, isotropic, elastic half-space. The arbitrary surface pressure field is discretized using a large number, n , of equally-sized circular loading elements. The problem is to compute the displacement at a large number, m , of points (or stations) distributed over the surface. The essence of our technique is to reorganize all but a computationally insignificant part of this calculation into an equivalent problem: compute the displacements due to a single circular loading element at a total of mn stations (where mn is the product $m \times n$). We solve this “parallel” problem at high computational speed by utilizing the sparse evaluation and massive interpolation (SEMI) method. Because the product mn that arises in our parallel problem is normally very large, we take maximum possible advantage of the acceleration achieved by the SEMI algorithm.

Key words: surface loading, elastic response, isotropic, layered half-space.

1. INTRODUCTION

The elastic response to surface loading has interested mathematicians, scientists and engineers for more than a century (Boussinesq 1885, Lamb 1901, Terazawa 1916, Love 1929). Interest in this topic within the Earth sciences has rapidly expanded in the last decade, following the widespread recognition that geodetic measurements can be used to observe oscillations of the Earth's surface that are driven by seasonal changes in the mass loads imposed on the solid Earth by the atmosphere, the hydrosphere and the cryosphere (Heki 2001, Blewitt *et al.* 2001, Van Dam *et al.* 2001, Dong *et al.* 2002). The elastic response to these environmental loading cycles is developed at global (Blewitt *et al.* 2001), regional (Heki 2001, Davis *et al.* 2004, Bevis *et al.* 2012) and local scales (Bevis *et al.* 2004, 2005). The near-field response to concentrated loads is influenced by shallow elastic structure which is highly variable in continental crust, particularly in the vertical direction (Bevis *et al.* 2004, 2005; Mooney *et al.* 1998). The layered, elastic half-space provides a useful mathematical framework for modeling Earth's instantaneous response to loading cycles when these loads are imposed over apertures which are small compared to the radius of the planet. The emerging use of continuous GPS networks to sense changes in ice mass within active ice sheets (*e.g.*, Hager 1991, Khan *et al.* 2007, Bevis *et al.* 2009, 2012) constitutes an important class of application.

Typically, the Earth's elastic response to changing surface loads of regional extent is computed using spectral methods (*e.g.*, Sasgen *et al.* 2005) and a global Earth structure model such as PREM (Dziewonski and Anderson 1981). The vertical resolution of whole Earth models like PREM is limited, especially near the surface. Indeed, PREM does not distinguish between the oceans and continents, let alone account for the rapid vertical variations in elastic moduli that frequently occur in the upper few kilometers of continental crust (Mooney *et al.* 1998). Bevis *et al.* (2004, 2005) argued that the near-field response to surface load changes is sensitive to the details of shallow elastic structure. This sensitivity could complicate (at least locally) the agenda of using crustal motion geodesy to gauge ice mass changes. We will illustrate this possibility by considering the elastic rebound associated with ice mass changes in Greenland. It is easier to include thin crustal layers in the Cartesian framework of Pan *et al.* (2007) (and this paper) than it is to do so in a spectral model implemented for a spherical Earth, since the latter approach would require harmonic expansions of very high degree to achieve the necessary radial resolution.

This paper is the sequel to the paper by Pan *et al.* (2007), in which we developed a precise numerical solution for the surface displacements produced by a uniform circular load imposed on the surface of a layered, elastic

half-space. This solution is computationally expensive, and so we also developed the sparse evaluation and massive interpolation (SEMI) method, which provides an approximate solution for the displacement field at very large numbers of points with vastly less computation per point (Pan *et al.* 2007). This approach utilizes the symmetry of the problem and the fact that the radial and vertical components of surface displacement are functions only of the radial distance, r , from the center of the load. We use our high accuracy but computationally expensive method to compute the displacement vectors at a limited number of r values (called control points or knots), and then use a variety of fast interpolation methods to determine the displacements at much larger numbers of intervening points. We can trade off the computational leverage achieved with the SEMI method and the magnitude of the errors associated with its approximations by choosing to use fewer or greater numbers of knots. The computational advantage of the SEMI method increases with the ratio of the number of surface stations to the number of knots.

In this paper we show how a circular loading element can be used to compute the surface displacement fields for an arbitrary surface load. Suppose we wish to compute the displacement at n stations due to a load we will approximate with m circular loading elements. The essence of our technique is to reorganize all but a computationally insignificant part of this calculation into an equivalent problem: compute the displacements due to a single circular loading element at a total of mn stations (where mn is the product $m \times n$). This maximizes the ratio of the number of stations to the number of knots, and so takes maximum possible advantage of the SEMI algorithm.

After describing this method in detail, we present some basic numerical tests of our code. In particular we address the problem of determining an appropriate number of loading elements for approximating a given load. Here the goal is to control the amplitude of artifacts or errors associated with discretization of the load. Lastly we consider some example loading problems and show how depth controlled variations in the elasticity constants can cause interesting and diagnostic features in surface displacement fields. These examples indicate that measuring the spatial development of the elastic response to known patterns of surface loading will enable us to infer information about subsurface structure. Of course, inversions of this kind require the forward problem to be evaluated many times, and it is this requirement which prompted us to develop a computationally efficient means for solving the forward problem for non-trivial loading geometries.

Before launching into this agenda, we note the following elementary points: (i) that the great advantage of using circular loads rather than point loads is that point loads produce displacement singularities and circular

loads do not; (ii) we use circular loading cells rather than finite loading cells of some other shape because circles are axi-symmetric, as is the layered material that we invoke; and (iii) all linear systems allow for superposition, and since we invoke a linear elastic material, then we can compute the response to multiple loading cells by superposing the solutions obtained for each individual loading cell. It is the axi-symmetric nature of the loading element problem, plus our ability to invoke superposition, that allows us to reframe the original problem into the parallel problem, and thereby achieve a large computational acceleration.

2. PROBLEM STATEMENT

Following Pan *et al.* (2007), we consider a layered half-space made up of p parallel, elastic, isotropic layers lying over an elastic, isotropic half-space. We adopt a Cartesian coordinate system in which the x and y axes lie in the surface plane ($z = 0$), and the z axis is positive downwards into the half-space. We assume that the surface of this half-space is subject to an imposed pressure field within a polygonal boundary B . The pressure P is zero everywhere outside of B , and is a known function $P(x, y)$ within and on B . The pressure P can be identified with the normal stress component σ_{zz} , and P is taken to be positive if the associated force is directed in the positive z direction. We assume that no shear tractions are imposed on the surface.

We wish to represent the pressure field within B using a suite of simple loading elements or cells, with the surface pressure applied within a single element being laterally uniform. The pressure field $P(x, y)$ is most easily approximated as piecewise constant by dividing the polygon into a regular grid of square loading cells (Fig. 1a), and assuming that the pressure everywhere within the i -th square cell is $\bar{P}_i = P(x_i, y_i)$, where (x_i, y_i) are the coordinates of the point in the center of that cell. There are two weaknesses to this approach: (i) the outer edges of the suite of square loading cells do not exactly correspond to the geometry of the polygon (Fig. 1a), and (ii) the actual pressure field $P(x, y)$ is not piecewise constant, and so the net force imposed by the pressure field on any square may deviate from that implied by the piecewise constant representation described above. However, by refining the grid so as to reduce the size of the individual cells, the errors associated with these problems can be reduced until they are negligible.

We cannot, in fact, use square loading cells, because square loads lack the symmetry which is essential to the SEMI method. We must use circular loading cells instead. However, it is useful to consider the circular loading cells as representing the square cells discussed above. Let us assume that the square cells in Fig. 1a had a width of $2a$, in which case the corresponding circular cells have radius a , as seen in Fig. 1b. Let us suppose that the i -th

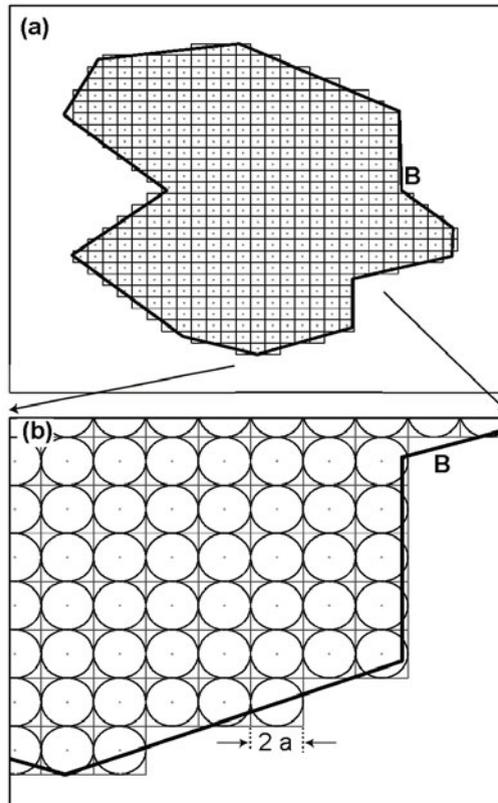


Fig. 1. Discretization of a surface pressure field $P(x,y)$ applied within a polygonal boundary B: (a) a square grid is developed for the loading area, and the average pressure in each square cell is approximated by the pressure at the center of that cell; (b) the load imposed within each square loading cell will actually be represented by a uniform circular load of diameter $2a$, which nominally exerts the same net force as the square load.

circle lies within the i -th square, so both are centered at point (x_i, y_i) . The area of the square is $A_s = 4a^2$, and that of the circle is $A_c = \pi a^2$. Clearly the use of a circular loading element is problematic because it cannot properly tile or cover the entire polygon – there are gaps between adjacent circles (Fig. 1b). However, we can largely overcome this problem by appropriate choice of the pressure we will assign to each circular element. If the constant pressure applied in the i -th circular cell is Q_i , and this load is to produce the same total force on the surface as the constant pressure \bar{P}_i applied within the square cell, then we require $A_c Q_i = A_s \bar{P}_i$. So we should set

$$Q_i = \frac{4}{\pi} P(x_i, y_i) . \quad (1)$$

We have scaled the actual pressure at the center of each circular loading element by an amount that accounts for the gaps between the circles. There are still some minor problems associated with this piecewise constant but discontinuous representation of the original pressure field, as we will discuss in Section 4, but by making the circular elements sufficiently small, we can reduce the magnitude of these problems to any level that we desire.

We can now state our problem, assuming that the decomposition of the load into circular loading elements has already been achieved: given a multi-layered elastic half-space described using the notation of Pan *et al.* (2007), and given a set of n circular loads with the same radius (a) but different pressures (Q_i , for $i = 1, 2, \dots, n$), compute the displacements at m stations located on the surface of the half-space.

3. DESCRIPTION OF THE ALGORITHM

In this section we describe the algorithm used to solve the problem just stated. Before developing this algorithm in a form suitable for efficient coding of the general problem, we explain the essence of our approach by considering an extremely simple example involving just two circular loads, 1 and 2, and a single station, S (Fig. 2). The pressure applied in the first circle is Q_1 , and that in the second circle is Q_2 . By linear superposition we can

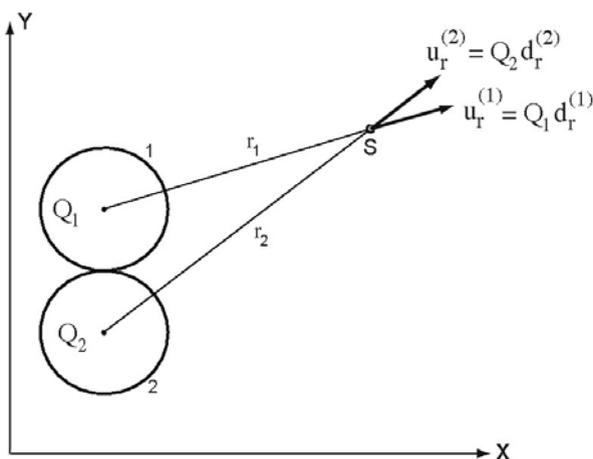


Fig. 2. The horizontal displacements induced at station S by uniform circular pressure loads Q_1 and Q_2 . This problem geometry including the displacement field can be expressed in radial coordinate systems attached to the center of each load, and in a global Cartesian coordinate system $\{X, Y\}$.

state that the displacement vector \mathbf{u} at S is the vector sum of the displacements $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ due to the first and second loads, respectively. Given the symmetry of a single circular load, the displacement it causes at any station is most easily described and computed in a cylindrical coordinate system whose origin lies at the center of the load. In this coordinate system, the displacement vector at any point has only two non-zero components – the vertical component u_z and the radial component u_r . Because each load has its own cylindrical coordinate system, we must transform vectors $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ into Cartesian coordinates in order to perform the vector summation and deliver the solution \mathbf{u} in the same coordinate system in which the load and station geometry are described. Let us represent this coordinate transformation in standard matrix form $\mathbf{u}_{\text{cart}} = \mathbf{T} \mathbf{u}_{\text{cyl}}$. (Because the matrix \mathbf{T} is sparse in this particular context, this is not the most efficient way in which to implement the transformation. We ignore this minor detail for now). We choose not to compute $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ explicitly, but instead compute $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$ which are the displacement vectors at S produced by unit pressure loading of our circular loading domains. Let us assume that $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$ are computed and expressed in cylindrical (or local) coordinate systems attached to each load, and that we wish to express the net displacement \mathbf{u} at S in the Cartesian (or global) coordinate system. Then

$$\mathbf{u} = Q_1 \mathbf{T}^{(1)} \mathbf{d}^{(1)} + Q_2 \mathbf{T}^{(2)} \mathbf{d}^{(2)} \quad , \quad (2)$$

where $\mathbf{T}^{(1)}$ is the transformation matrix associated with load 1, *etc.* There are two important points about this equation. Firstly, almost all of the computational burden involved in evaluating this equation is incurred in evaluating $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$. The coordinate transformations and the summation are computationally trivial in comparison. Secondly, when considered from the perspective of their local coordinate systems the two unit loads are identical, and since the spatial variability of the surface displacement field \mathbf{d} depends only on r (since the symmetry of the load implies no θ dependence), then $\mathbf{d}^{(1)} = \mathbf{d}(r_1)$ and $\mathbf{d}^{(2)} = \mathbf{d}(r_2)$, and we can view the evaluation of \mathbf{d} as solving the problem of two stations and a single unit load, rather than two unit loads and a single station.

This second insight is the crucial one. If we want to compute the displacement at a single station due to m circular loads, we can do most of the work by solving the parallel problem of determining the displacements caused at m stations by a single unit circular load. In the original problem r_i is the distance between the station and the center of the i -th unit circular load. In the parallel problem, r_i is the distance between the center of the single unit circular load and the i -th station. Having solved this parallel problem we can construct the solution to the original problem by the obvious general-

ization of Eq. 2. We can extend this trick even further: the problem of solving the displacements at m surface stations due to n circular loads (with common radius a) can be very largely transformed into the problem of finding the displacements at $n m$ stations due to a single unit circular load. The great advantage of diverting to the parallel form of the problem is that the computational efficiency of the SEMI method increases in proportion with the total number of stations.

We are ready now to present the algorithm for the problem stated in its most general form. We wish to compute the displacements at m stations due to n circular loads. Each of these loads has identical radius a . The i -th circular load is centered at (x_i^c, y_i^c) and is subject to uniform pressure Q_i , which is considered positive if the associated force is oriented in the z -direction, *i.e.*, into the half-space. The j -th station has surface coordinates (x_j^s, y_j^s) . Consider the k -th combination of the i -th circle and the j -th station (Fig. 3). We can consider k the station number in the parallel problem. The relative position vector describing the position of station j relative to circle i is

$$\mathbf{r}^k = \mathbf{r}^{ij} = (x_j^s - x_i^c) \hat{\mathbf{x}} + (y_j^s - y_i^c) \hat{\mathbf{y}} \tag{3}$$

which has Euclidean length

$$r^k = r^{ij} = \sqrt{(x_j^s - x_i^c)^2 + (y_j^s - y_i^c)^2} \tag{4}$$

which is simply the distance from the center of load i to station j in the original problem, or the distance from the center of the single unit load to the k -th station in the parallel problem. The unit vector which points from the center of circle i to station j is

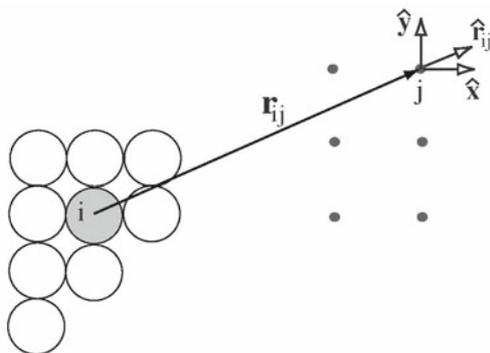


Fig. 3. The coordinate systems used to describe the horizontal displacement at station j due to load i .

$$\hat{\mathbf{r}}^{ij} = \hat{r}_x^{ij} \hat{\mathbf{x}} + \hat{r}_y^{ij} \hat{\mathbf{y}} \quad (5)$$

where

$$\hat{r}_x^{ij} = \frac{(x_j^s - x_i^c)}{r^{ij}}, \quad \hat{r}_y^{ij} = \frac{(y_j^s - y_i^c)}{r^{ij}}. \quad (6)$$

We use the SEMI method (Pan *et al.* 2007) to solve for the parallel problem, finding a total of $m n$ displacement vectors

$$\mathbf{d}^k = d_r^k \hat{\mathbf{r}} + d_z^k \hat{\mathbf{z}} \equiv d_r^{ij} \hat{\mathbf{r}} + d_z^{ij} \hat{\mathbf{z}} \quad (7)$$

produced by the single unit circular load. In order to revert to our original problem we must transform these vectors into the global Cartesian coordinate system, *i.e.*, find

$$\mathbf{d}^{ij} = d_x^{ij} \hat{\mathbf{x}} + d_y^{ij} \hat{\mathbf{y}} + d_z^{ij} \hat{\mathbf{z}} \quad (8)$$

for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$, or, equivalently, for $k = 1, 2, \dots, n m$. The vertical (z) component has the same value in local and global coordinates. We need only to transform the horizontal vector components. This can be done by projecting the radial component of \mathbf{d} onto unit vectors in the X and Y directions (Fig. 3), *i.e.*

$$d_x^{ij} = d_r^{ij} \hat{r}_x^{ij}, \quad d_y^{ij} = d_r^{ij} \hat{r}_y^{ij}. \quad (9)$$

We can now express the displacement at each station due to a unit load at the position of the n non-unit loads in the original problem. To solve the displacement at a given station (j) in response to the original n circular loads, we simply scale the unit load responses (Eq. 8) with the appropriate loads, *i.e.*

$$\mathbf{u}^j = \sum_{i=1, n} Q_i \mathbf{d}^{ij}. \quad (10)$$

4. SOME NUMERICAL TESTS AND EXAMPLES

We begin with a test in which we attempt to reproduce the solution presented by Becker and Bevis (2004) for a uniform rectangular load on a uniform elastic half-space (UHS), which is known as Love's Problem. We do this using our SEMI approach by invoking a layered elastic half-space (LHS), consisting of 3 layers on a half-space, in which all four of these layers have identical elastic properties – the general problem thus degenerates into the special case (UHS), and the solutions should match.

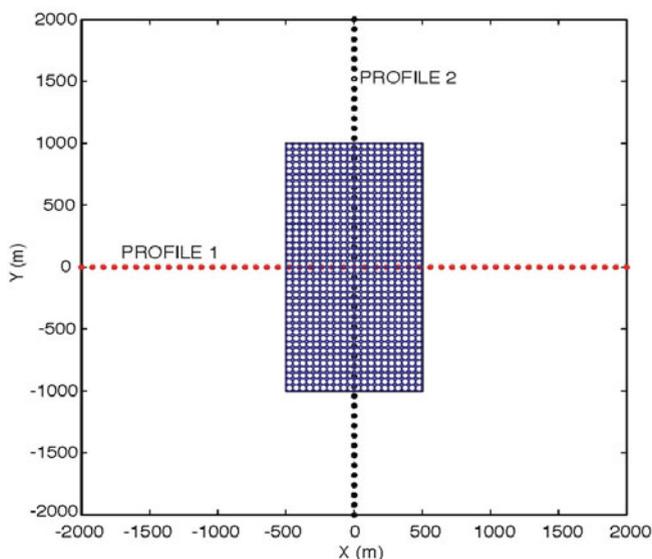


Fig. 4. A uniform rectangular load is approximated by an array of circular loading elements, and the resulting surface displacement field is sampled along profiles 1 and 2.

The geometry associated with this problem is shown in Fig. 4. The $1 \text{ km} \times 2 \text{ km}$ rectangle is an idealized representation of a lake, and the pressure loading corresponds to a water depth of 100 m. We evaluate the surface displacement field along two profiles (1 and 2) each of which bisect the load. We follow the notation of Becker and Bevis (2004) in which the x , y , and z components of displacement are called u , v , and w , respectively. We evaluate the UHS solution assuming that Young's modulus $E = 0.6 \times 10^{11} \text{ Nm}^{-2}$ and Poisson's ratio $\nu = 0.25$ (or, equivalently, that the Lamé parameters λ and μ are given by $\lambda = \mu = 2.4 \times 10^{10} \text{ Nm}^{-2}$). We set surface pressure $P = \rho gh$ where $\rho = 1000 \text{ kg m}^{-3}$, $g = 9.82 \text{ ms}^{-2}$ and $h = 100 \text{ m}$. These exact solutions are shown by the red dots in Fig. 5. In order to test our SEMI code, we approximate the load using $20 \times 40 = 800$ circular loading elements (Fig. 4) and invoke a LHS in which the first three layers have thicknesses of 1500, 3000, and 6000 m, and the fourth layer is semi-infinite. We set $E = 0.6 \times 10^{11} \text{ Nm}^{-2}$ and $\nu = 0.25$ in all four layers. The resulting solution is shown by the blue curves in Fig. 5. We can see that the UHS and LHS correspond very closely for all points (or stations) outside of the rectangle, and closely within the rectangle. A careful examination shows that within the rectangle the SEMI solution for the degenerate LHS oscillates around the exact solution for the UHS.

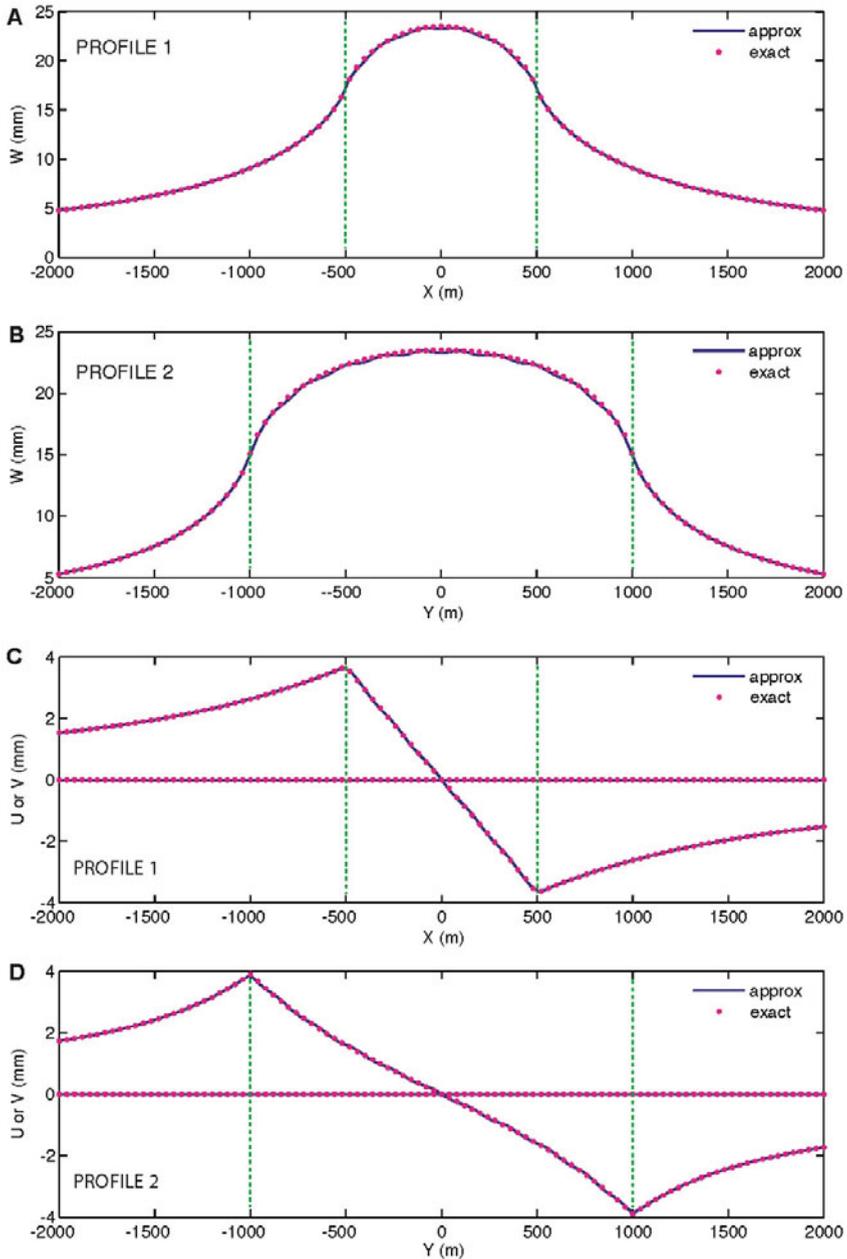


Fig. 5. The horizontal (u and v) and vertical (w) components of displacement along each profile (see Fig. 4) computed exactly using the equations of Becker and Bevis (2004), and using the approximation techniques developed in this paper. The dotted green lines indicate the edges of the rectangular load.

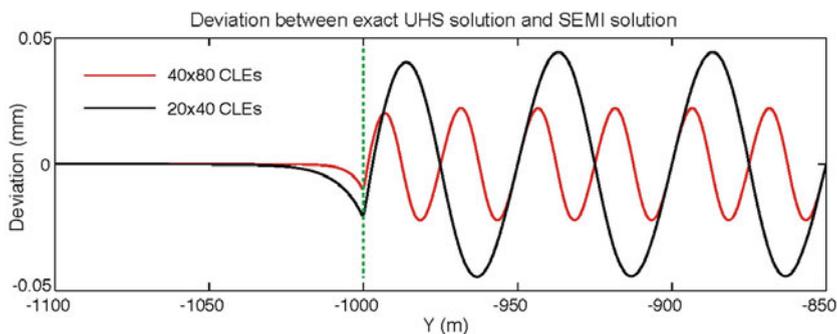


Fig. 6. The deviation between the exact and the SEMI solutions for the vertical displacements near the edge of the rectangle (see Fig. 4) for two cases in which the load is approximated by: (i) a 20×40 array, and (ii) a 40×80 array of circular loading elements.

This oscillatory behavior, which we call ripple, is made more obvious in Fig. 6 in which we difference the two sets of solutions for the horizontal component of displacement, u , along a small segment of profile 2, where it crosses the bottom side of the rectangle. The black curve shows the difference between the solutions when the load is approximated by $20 \times 40 = 800$ circles (corresponding to the SEMI solution shown in Fig. 5). This oscillation has a sinusoidal or “egg crate” form within the rectangle, and wavelength of this sinusoid (in the x and y directions) is the distance ($2a$) between the centers of adjacent circular loading elements (CLEs). Clearly the ripple in the SEMI solution manifests the discretization of the load using CLEs. The red curve shows the SEMI solution error or ripple when $40 \times 80 = 3200$ CLEs are used. Halving the radius of the CLE reduces the wavelength and the amplitude of the ripple by a factor of two (Fig. 6), at the cost of increasing the computational burden by a factor of 4. But since the SEMI algorithm is so fast, it will usually be possible to reduce the magnitude of ripple to an acceptable level at an acceptable computational cost.

We now consider some examples which test the SEMI algorithm in the context of a non-degenerate LHS (*i.e.*, the layers have different elastic constants) by developing certain special or limiting cases in which our intuition provides us with an expected outcome or value. In both of the examples that follow we consider the surface response to a single uniform circular load of radius a . This load is applied to an elastic space consisting of one layer of thickness t overlying a half-space. It is often more instructive to refer to the normalized layer thickness, $T = t/a$. We shall assume that Poisson’s ratio, ν , is 0.25 for both layers, and that Young’s modulus is E_1 in the upper layer, and E_2 in the underlying half-space. Because of the symmetry of the circular

load, the surface displacement vector at each point has only vertical and radial components, and both are purely a function of r , the radial distance from the center of the load. In plotting this dependence we shall use the normalized radial distance, r/a . We contrast the surface response for this 2-layer space with the surface response of a UHS subject to the same load, assuming that this UHS has $\nu = 0.25$ and Young's modulus equal either to E_1 or to E_2 .

We first consider the problem of a thick layer on a half-space, for which $t \gg a$, or $T \gg 1$. This geometry is shown in the left hand part of the diagram (comprising two vertical sections) inset into Fig. 7a. We assume that $E_1 = 0.1$, $E_2 = 1$ and consider two values for the thickness of the upper layer: $T = 10$ and $T = 100$. The vertical and radial displacement profiles are shown in Fig. 7a, b, respectively, and can be compared to the UHS response for the case in which Young's modulus $E_{\text{UHS}} = E_1 = 0.1$. By comparing these response curves we can confirm what we might have guessed intuitively: if the upper layer of the 2-layer space is very much thicker than radius of the load, the surface response in the near and medium field of the load is almost identical to that of a UHS whose properties are those of the upper layer.

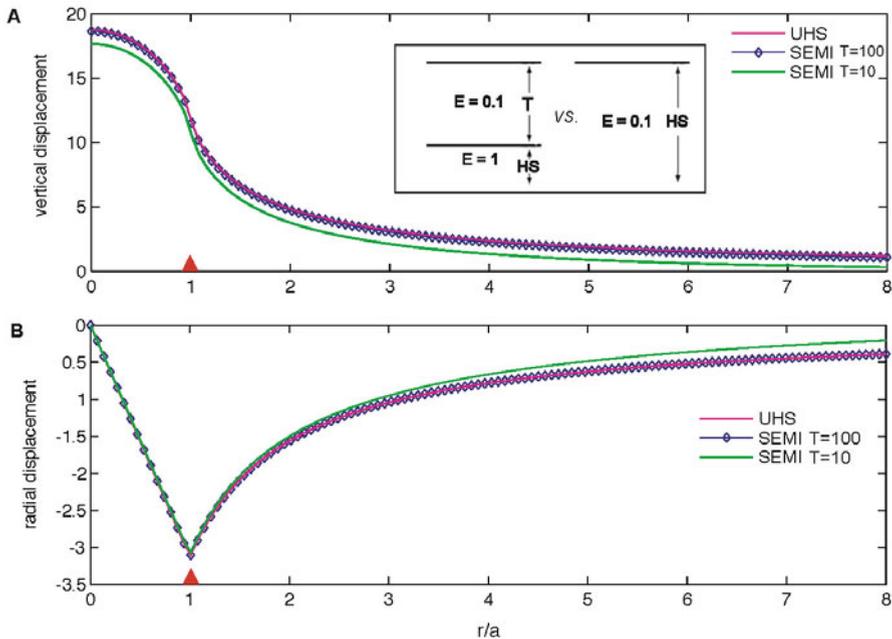


Fig. 7. A comparison between the: (a) vertical, and (b) radial displacements caused by a uniform circular load imposed on: (i) a uniform half-space (UHS) with $E = 0.1$, and (ii) a layer of thickness t and $E = 0.1$ overlying a half-space with $E = 1.0$, as depicted in the inset in sub-plot (a). The comparison between the UHS and 2-layer solutions are presented for $T = t/a = 10$ and $T = 100$.

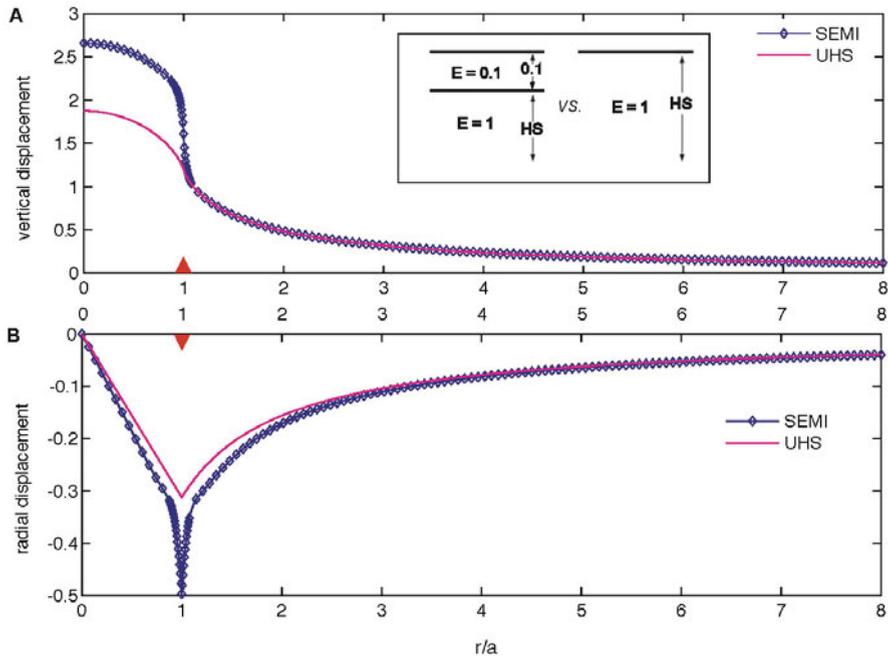


Fig. 8. A comparison between the: (a) vertical, and (b) radial displacements caused by a uniform circular load imposed on: (i) a uniform half-space with $E = 1.0$, and (ii) a layer of thickness $t = 0.1 a$ and $E = 0.1$ overlying a half-space with $E = 1.0$, as depicted in the inset in sub-plot (a).

Next we consider the “opposite” problem from that just discussed, in which the first layer of the 2-layered space is much “thinner” than the radius of the circular load (specifically $T = 0.1$). In this case one might suspect that the influence of the thin upper layer would be restricted, and so it is useful to compare this response to the UHS response in which $E_{\text{UHS}} = E_2 = 1$. Comparing these response profiles we see that the vertical surface displacement profiles are very distinct within the loading domain. Since the LHS has a very compliant upper layer, the surface beneath the load is deflected to a much greater extent than in the case of the UHS (Fig. 8a). But for $r > a$, the vertical response of the LHS is very nearly identical to that of the UHS. In other words, the surface response of the layered space is strongly influenced by the first layer within the loading area, but is dominated by the lower layer a short distance outside of the load. If we examine the radial component of displacement for this same problem (Fig. 8b) we see that in the medium field (say $r > 3a$) the surface response of the LHS is dominated by the lower layer (it matches the response of a UHS with the same properties as the 2nd layer of the LHS). The LHS and UHS responses differ to their

greatest extent at the boundary of the circle ($r = a$). As we can see from the UHS response curve, the load tends to pull material near the edge of the circle inwards as well as downwards. In the case of the LHS the very compliant upper layer leads to an enhanced inwards displacement, and the width of the “spike” in the LHS response curve is influenced by the thickness of the first layer.

Lastly we consider the accuracy of the SEMI algorithm itself, in the primitive context of evaluating the displacements due to a single circular load of unit radius. Here the question is how accurate is the approximation delivered by the SEMI method in comparison with (much slower) direct evaluation? We investigated this issue by generating a suite of elastic structure models using a Monte Carlo approach. We generated 50 models consisting of five layers over a half-space, and 50 models consisting of 9 layers over a half-space. Each layer thickness was generated randomly, and within our ensemble thicknesses ranged between 0.070 and 8282 m. Young’s modulus and Poisson’s ratio for each layer were also randomly generated, with the former falling in the range 0.11-5147 Nm^{-2} , and the later in the range 0.01-0.49. In every case the load had a radius of 1 m, and the surface displacement field was evaluated at 1250 stations from $r = 0$ to $r = 1000$ m, with more than half of the stations falling in the range 0-40 m. For each subsurface model the vertical and horizontal displacement components we computed directly and using the SEMI approach, these quantities were compared. This amounted to a total of 125 000 comparisons for each component of displacement. We defined the “relative error” as the difference between the SEMI and the directly computed value for displacement divided by the directly computed value. The RMS relative errors were 5.1×10^{-3} for the radial component and 3.1×10^{-3} for the vertical component. This result was obtained using our standard SEMI code which employs a total of 134 knots.

It should be kept in mind that when we model a general surface load using many circular loading elements, the interpolation errors associated with the different loading elements will tend to cancel at a specific station, particularly in the near field of the load where the displacements are largest.

5. WEIGHING THE ICE SHEETS WITH GPS: THE IMPACT OF SHALLOW GEOLOGICAL STRUCTURE

We now use our computer code to implement a more substantial and interesting calculation. We wish to demonstrate that including thin, near-surface layers of compliant sedimentary rock layers into our Earth model can produce significant changes to the computed response for a given loading or unloading scenario. We will illustrate this possibility by considering the

solid Earth's elastic adjustment to ice load changes in Greenland. Our goal here is not so much to compute the best possible model for such a response, but rather to show that the fine details of near-surface elastic structure can actually make a difference to Earth's response to typical (*i.e.*, realistic) changes in ice mass. That is, our purpose here is to perform a sensitivity study.

We use the ice loss grids of Krabill *et al.* (2000) that represent the spatial pattern of ice surface height changes (dH/dt) in Greenland during the interval 1995-2000. We chose this grid because it covers all of Greenland in a consistent way. We consulted with Bill Krabill and Bob Thomas to establish estimates of the near surface densities (depending on surface elevation) that allow us to convert height rates (dH/dt) to mass rates (dM/dt). We are well aware that the rate of ice loss in Greenland has accelerated quite dramatically since the year 2000. So our ice mass rate (dM/dt) grid represents something of a baseline measurement. Nevertheless, this calculation will serve to illustrate the potential impact of shallow geological structure on the Earth's unloading response. We approximated the surface loading field (dM/dt) using 7178 disk loads for the purpose of computing Earth's elastic response.

We use the crustal structure model CRUST 2.0 (<http://igppweb.ucsd.edu/~gabi/crust2.html>) to provide a reasonable estimate for the average crustal structure beneath Greenland. CRUST 2.0 provides a layered model for the crust and uppermost mantle, in each 2° by 2° square, specifying both the P and S wave velocity for each layer, as well as density. Given these three parameters it is a simple matter to estimate the two elastic parameters for each layer – either the two Lamé parameters, or, equivalently, Young's modulus (E) and Poisson's ratio (ν). We averaged the results we obtained over most of Greenland and adopted this nominal structure for the purpose of modeling the Earth's elastic response to changing ice loads. We combined this average CRUST 2.0 structure for the crust and upper mantle (above 400 km depth) with the PREM structure for the deeper Earth (below 400 km depth) – we refer to this as the HYBRID elastic structure model. Whereas our discretized version of PREM has a top layer 3 km thick composed of a material with $E = 68$ GPa and $\nu = 0.28$, the top layer in HYBRID is only 359 m thick and has $E = 8.17$ GPa and $\nu = 0.35$.

CRUST 2.0 identifies several areas in coastal Greenland in which the near-surface compliant layer is 1 km thick rather than just 359 m thick. This is basically an intelligent guess based on local surface geology. We have produced a third elastic structure model, called HYBRID/S, by modifying HYBRID so that its first (presumably sedimentary) layer is 1 km thick. We used our computer code to compute the vertical velocity of the Earth's crust in response to the ice mass rate field derived from Krabill *et al.* (2000), using the PREM, HYBRID, and HYBRID/S models for elastic structure. The re-

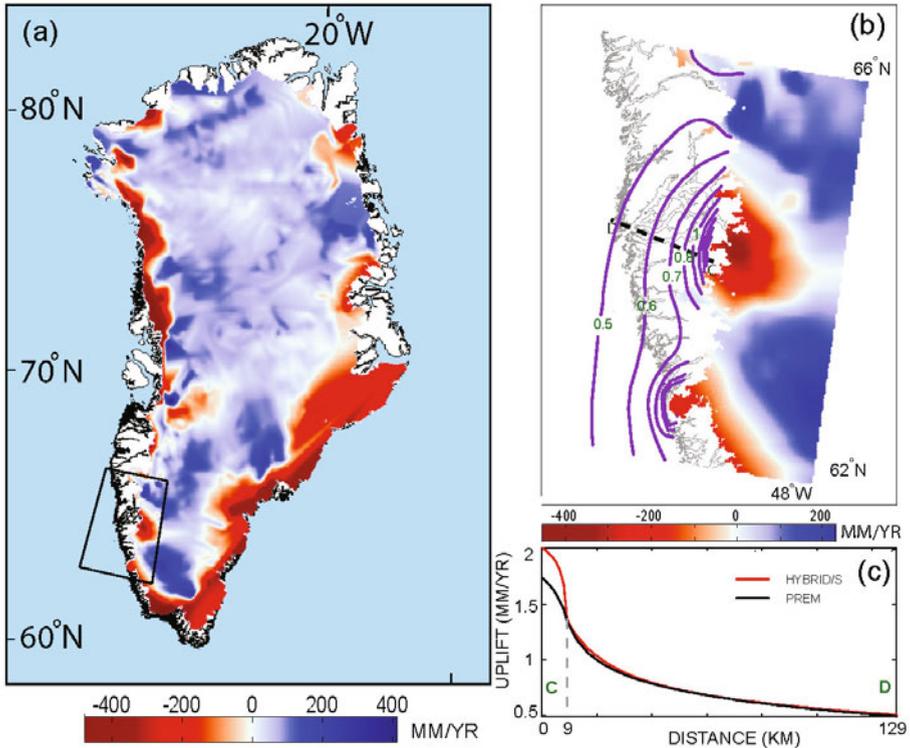


Fig. 9: (a) The average rate of ice surface elevation change in the time interval 1995-2000, according to Krabill *et al.* (2000). This height rate field (dH/dt) was converted into a mass rate field (dM/dt), in order to compute the Earth's elastic response to load changes. (b) A blow up, depicting part of SW Greenland, showing vertical crustal velocity contours adjacent to the ice sheet, and the location of the velocity profile D-C. The contours are those predicted using PREM structure. (c) The elastic rebound velocities along the profile D-C are predicted using the structure models PREM and HYBRID/S. The last model invokes a 1 km thick layer of relatively compliant rocks (presumably sediments) immediately beneath the surface of the crust, as suggested by model CRUST 2.0. The dashed vertical line represents the ice front. Note that when we use HYBRID/S rather than PREM, the rate of elastic rebound coincident with the load is increased by about 10%. This figure is modified from Zhou (2008).

bound velocities predicted using PREM and HYBRID are everywhere similar (differences are less than a few percent), but the results obtained using HYBRID/S are quite distinct in the near-field of the zones of major ice loss (Fig. 9). This finding indicates, not surprisingly, that a near-surface layer of relatively low stiffness enhances the rate of elastic rebound within and immediately adjacent to the changing load.

Whereas it is reasonable to use global Earth structure models such as PREM to model the medium- and far-field elastic response to surface loading or unloading, and even much of what is considered the near-field, this choice of structure model will usually cause the very-near-field loading response to be underestimated, since PREM does not resolve the strong near-surface decline in stiffness characteristic of many continental areas. While lateral heterogeneity of shallow elastic structure may be something of a complication for space geodetic studies of ice mass balance, it might also be something of an opportunity in that an unusually compliant shallow subsurface regime might act as an “amplifier” of the vertical crustal motion signals produced by local ice loss.

6. DISCUSSION

Approximating a spatially finite but otherwise general pattern of surface loading with a suite of circular loading elements (of equal diameter) allows us to exploit the computational acceleration associated with the SEMI method presented in Pan *et al.* (2007). This paper builds on their results and utilizes their computer codes. In turn, Pan *et al.* (2007) was motivated by the application and the algorithm presented here. The key idea discussed here is the reformulation of all but a computationally minor component of the problem of finding the displacements at m stations due to n circular loads into the “parallel” problem of computing the displacement produced at $n m$ stations in response to a single circular load. This is precisely the problem that the SEMI method was designed to address. Note, however, that the problem reformulation presented herein is in no way tied to the specific interpolation strategies employed by a code that implements the SEMI algorithm for a single circular load. Our code for computing the displacement field due to a suite of circular loading elements calls a distinct code that implements the SEMI method for solving the “primitive” or elementary problem of a single circular load. This modularity makes it easy to incorporate any improvements that may be achieved in the SEMI method.

The methodology described in this paper is very simple to state and to implement. We have stated the algorithm in great, and perhaps a surprising level of detail. This is because the whole point of this paper, and Paper 1, is computing a solution as rapidly as possible. It would be easy to present our algorithm using equations that are more compact and/or more evocative than those used above, and yet do so in a way that potentially wastes computer time. We have avoided the use of the transformation matrix \mathbf{T} that appears in Eq. 2, for example, because 4 of its 9 elements are zero and we do not want to waste time computing products and sums that contribute nothing towards the final solution. Similarly we have avoided computing trigonometric or in-

verse trigonometric functions. At the risk of appearing pedantic, we have presented the computation in what seems to us a nearly optimal approach for coding the algorithm.

The computational acceleration associated with the strategy presented here, combining problem reformulation and the SEMI method, is particularly useful in solving inverse problems, since inverse methods often involve solving the forward problem thousands or even hundreds of thousands of times. Having solved the inverse problem using the approach described in this paper, one could use the direct method described in Paper 1, to explore the solution in more detail, for example by evaluating quantities, such as the subsurface strain field, that did not feature in the inverse problem. Perhaps the most serious drawback of the SEMI method is that it is restricted to computing fields at the surface of the half-space.

Acknowledgements. MB and HZ were supported by NASA contract NNG04GF28G and NSF grant ARC-1111882. EP, FH, and RZ were partially supported by the Ohio DOT.

References

- Becker, J.M., and M. Bevis (2004), Love's problem, *Geophys. J. Int.* **156**, 2, 171-178, DOI: 10.1111/j.1365-246X.2003.02150.x.
- Bevis, M., E. Kendrick, A. Cser, and R. Smalley Jr. (2004), Geodetic measurement of the local elastic response to the changing mass of water in Lago Laja, Chile, *Phys. Earth Planet. In.* **141**, 2, 71-78, DOI: 10.1016/j.pepi.2003.05.001.
- Bevis, M., D. Alsdorf, E. Kendrick, L.P. Fortes, B. Forsberg, R. Smalley Jr., and J. Becker (2005), Seasonal fluctuations in the mass of the Amazon River system and Earth's elastic response, *Geophys. Res. Lett.* **32**, 16, L16308, DOI: 10.1029/2005GL023491.
- Bevis, M., E. Kendrick, R. Smalley Jr., I. Dalziel, D. Caccamise, I. Sasgen, M. Helsen, F.W. Taylor, H. Zhou, A. Brown, D. Raleigh, M. Willis, T. Wilson, and S. Konfal (2009), Geodetic measurements of vertical crustal velocity in West Antarctica and the implications for ice mass balance, *Geochem. Geophys. Geosyst.* **10**, 10, Q01005, DOI: 10.1029/2009GC002642.
- Bevis, M., J. Wahr, S.A. Khan, F.B. Madsen, A. Brown, M. Willis, E. Kendrick, P. Knudsen, J.E. Box, T. van Dam, D.J. Caccamise II, B. Johns, T. Nylén, R. Abbott, S. White, J. Miner, R. Forsberg, H. Zhou, J. Wang, T. Wilson, D. Bromwich, and O. Francis (2012), Bedrock displacements in Greenland manifest ice mass variations, climate cycles and climate change, *Proc. Natl. Acad. Sci. USA* **109**, 30, 11944-11948, DOI: 10.1073/pnas.1204664109.

- Blewitt, G., D. Lavallée, P. Clarke, and K. Nurutdinov (2001), A new global mode of Earth deformation: Seasonal cycle detected, *Science* **294**, 5550, 2342-2345, DOI: 10.1126/science.1065328.
- Boussinesq, J.V. (1885), *Application des Potentiels à l'Etude de l'Equilibre et du Mouvement des Solides Elastiques*, Gauthier-Villars, Paris, 508 pp. (in French).
- Davis, J.L., P. Elósegui, J.X. Mitrovica, and M.E. Tamisiea (2004), Climate-driven deformation of the solid Earth from GRACE and GPS, *Geophys. Res. Lett.* **31**, 24, L24605, DOI: 10.1029/2004GL021435.
- Dong, D., P. Fang, Y. Bock, M.K. Cheng, and S. Miyazaki (2002), Anatomy of apparent seasonal variations from GPS-derived site position time series, *J. Geophys. Res.* **107**, B4, 2075, DOI: 10.1029/2001JB000573.
- Dziewonski, A.M., and D.L. Anderson (1981), Preliminary reference Earth model, *Phys. Earth Planet. In.* **25**, 4, 297-356, DOI: 10.1016/0031-9201(81)90046-7.
- Hager, B.H. (1991), Weighing the ice sheets using space geodesy: A way to measure changes in ice sheet mass, *EOS Trans. AGU* **72**, 17, 91.
- Heki, K. (2001), Seasonal modulation of interseismic strain buildup in northeastern Japan driven by snow loads, *Science* **293**, 5527, 89-92, DOI: 10.1126/science.1061056.
- Khan, S.A., J. Wahr, L.A. Stearns, G.S. Hamilton, T. van Dam, K.M. Larson, and O. Francis (2007), Elastic uplift in southeast Greenland due to rapid ice mass loss, *Geophys. Res. Lett.* **34**, 21, L21701, DOI: 10.1029/2007GL031468.
- Krabill, W., W. Abdalati, E. Frederick, S. Manizade, C. Martin, J. Sonntag, R. Swift, R. Thomas, W. Wright, and J. Yungel (2000), Greenland ice sheet: High-elevation balance and peripheral thinning, *Science* **289**, 5478, 428-430, DOI: 10.1126/science.289.5478.428.
- Lamb, H. (1901), On Boussinesq's problem, *Proc. Lond. Math. Soc.* **34**, 1, 276-284, DOI: 10.1112/plms/s1-34.1.276.
- Love, A.E.H. (1929), The stress produced in a semi-infinite solid by pressure on part of the boundary, *Philos. Trans. Roy. Soc. Lond. A* **228**, 377-420, DOI: 10.1098/rsta.1929.0009.
- Mooney, W.D., G. Laske, and T.G. Masters (1998), CRUST 5.1: A global crustal model at $5^\circ \times 5^\circ$, *J. Geophys. Res.* **103**, B1, 727-747, DOI: 10.1029/97JB02122.
- Pan, E., M. Bevis, F. Han, H. Zhou, and R. Zhu (2007), Surface deformation due to loading of a layered elastic half-space: A rapid numerical kernel based on a circular loading element, *Geophys. J. Int.* **171**, 1, 11-24, DOI: 10.1111/j.1365-246X.2007.03518.x.
- Sasgen, I., D. Wolf, Z. Marinenc, V. Klemann, and J. Hagerdorn (2005), Geodetic signatures of glacial changes in Antarctica: Rates of geoid height change

- and radial displacement due to present and past ice-mass variations, GFZ Scientific Technical Rep. STR05/01, GeoForschungsZentrum, Potsdam, Germany.
- Terazawa, K. (1916), On the elastic equilibrium of a semi-infinite solid under given boundary conditions, with some applications, *J. Coll. Sci. Imp. Univ. Tokyo* **37**, 7, 1-64.
- van Dam, T., J. Wahr, P.C.D. Milly, A.B. Shmakin, G. Blewitt, D. Lavallée, and K.M. Larson (2001), Crustal displacements due to continental water loading, *Geophys. Res. Lett.* **28**, 4, 651-654, DOI: 10.1029/2000GL012120.
- Zhou, H. (2008), Layered Cartesian half-space models for Earth's elastic response to contemporary surface loading phenomena, Ph.D. Thesis, Ohio State University, Columbus, USA, 160 pp.

Received 14 March 2014

Received in revised form 28 October 2014

Accepted 29 October 2014

Permeability Evolution and Rock Brittle Failure

Qiang SUN¹, Lei XUE², and Shuyun ZHU¹

¹School of Resources and Geosciences, China University of Mining and Technology, Xuzhou, China; e-mail: sunqiang04@126.com

²Key Laboratory of Shale Gas and Geoengineering, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China

Abstract

This paper reports an experimental study of the evolution of permeability during rock brittle failure and a theoretical analysis of rock critical stress level. It is assumed that the rock is a strain-softening medium whose strength can be described by Weibull's distribution. Based on the two-dimensional renormalization group theory, it is found that the stress level λ_c (the ratio of the stress at the critical point to the peak stress) depends mainly on the homogeneity index or shape parameter m in the Weibull's distribution for the rock. Experimental results show that the evolution of permeability is closely related to rock deformation stages: the permeability has a rapid increase with the growth of cracks and their surface areas (*i.e.*, onset of fracture coalescence point), and reaches the maximum at rock failure. Both the experimental and analytical results show that this point of rapid increase in permeability on the permeability-pressure curve corresponds to the critical point on the stress-strain curve; for rock compression, the stress at this point is approximately 80% of the peak strength. Thus, monitoring the evolution of permeability may provide a new means of identifying the critical point of rock brittle fracture.

Key words: microfracturing, critical information, permeability evolution.

1. INTRODUCTION

Rock permeability is strongly influenced by the deformation of the rock and controlled by the evolution of the geometrical structure of cracks and pore space (volume, connectivity, tortuosity, shape, *etc.*). Depending on the initial porosity and the stress acting on the rock, compaction and strain hardening or, on the contrary, dilation and strain softening may occur (Sulem and Oufroukh 2006).

A better understanding of the evolution of permeability was initiated through the experimental work of Brace *et al.* (1968), who first applied the pulse decay method in triaxial tests of granites to measure their permeability under high pressure. In the past half century, significant achievements have been made in systematic studies on the evolution of permeability during rock cracking. The permeability may decrease slightly under the application of relatively small stresses, but when the stress is further increased, a threshold level is usually found, above which the permeability increases very markedly, even under confining pressure (*e.g.*, Brace *et al.* 1968, Mordecai and Morris 1971, Zoback and Byerlee 1975a, b; Walsh and Brace 1984, Wong *et al.* 1997, Zhu and Wong 1996, 1997, 1999, Yang *et al.* 2008, David *et al.* 2001, Lock *et al.* 2002, Paterson and Wong 2005, Liu *et al.* 2009, Zhang *et al.* 2013). This threshold may be closely correlated with the onset of marked Acoustic Emission (AE) (Paterson 1978). Jiang *et al.* (2002) gave the rock permeability-stress mathematical expression before the peak stress is reached, and investigated two characteristic parameters – the critical anti-permeability strength of rock and the initial permeability – according to the rock stress-strain curve and the permeability-strain curve obtained from rocks under complete loading-to-unloading cycles. Wang *et al.* (2006) found that the permeable pressure complies with a negative exponential function of time during rock deformation and failure.

Under an external load, the evolution of rock permeability is influenced by the stress-strain state (Sun *et al.* 2012). In this study, we monitored three parameters: stress, strain, and permeability, for compressed rocks undergone brittle failure, and the evolution of rock permeability is studied experimentally, especially during disintegration and rupture of the rock after the failure stress has been reached.

The renormalization group probability of destruction is also employed to analyze the stress and strain at the critical point of rock brittle failure. The fracturing of stressed materials is analogous to a critical phenomenon at a second-order phase transition or the percolation phenomenon (Martin 1997, Sornette 2000). The moment of rupture is similar to a critical point, so the fracturing process can be described by a renormalization-group scheme (Anifrani *et al.* 1995), and a critical region exists in the vicinity of the critical point of rupture (Sornette and Andersen 1998).

2. PERMEABILITY EVOLUTION AND STRESS-STRAIN OF INTACT ROCKS UNDER COMPRESSION

Several characteristic stresses important for understanding the damage process can be identified from the stress-strain curve shown in Fig. 1. σ_a is the crack closure stress, σ_b is the crack initiation stress, σ_c is the crack damage stress corresponding to long-term rock strength (Martin 1993, 1997), or the critical stress; and σ_d is the peak stress. The crack initiation or threshold stress σ_b is defined as the onset of stable crack growth, which is sometimes defined as the point where the crack volumetric strain deviates from zero. The permeability increases with the growth of cracks. When the damage initiation limit is exceeded, existing cracks start to propagate and new cracks initiate in a stable fashion, but critical rock mass damage is not encountered until the density of crack is sufficient for cracks to coalesce to form shear bands or tensile spalls. This state is defined as the “crack damage stress” (Martin 1997) or “permeability threshold”. The crack damage stress σ_c is generally defined as the point when volumetric strain reversal occurs and unstable crack growth begins. It also corresponds to the stress when a drastic increase of permeability is observed. In laboratory tests, σ_c can be considered as the wall strength in massive rocks. Table 1 is the damage stress level (*i.e.*, critical stress level) obtained from triaxial compression tests by permeability-strain and stress-strain curves.

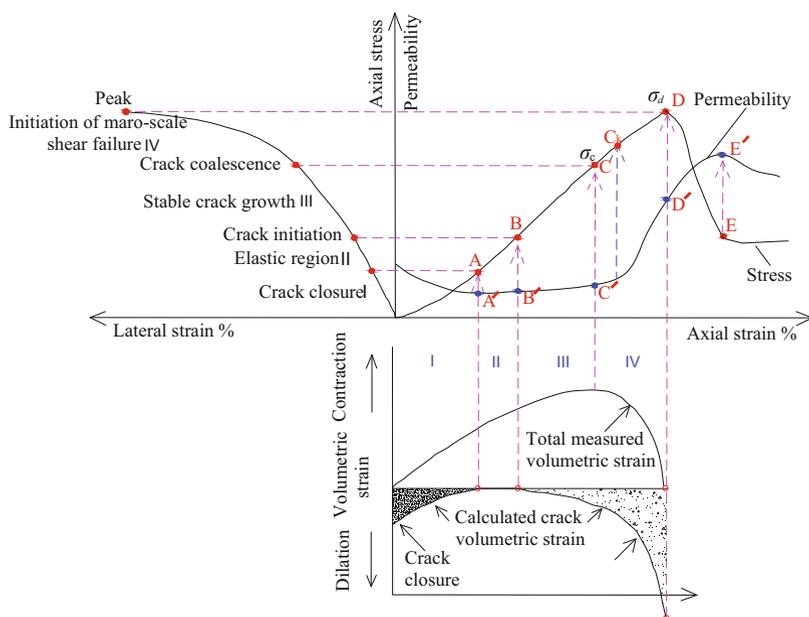


Fig. 1. Corresponding relationship between rock failure and permeability evolution.

Table 1

Relationship between the parameters of rock stress, strain, and permeability

No. of rock samples	Rock type	Confining pressure [MPa]	Pore pressure [MPa]	Differential pressure [MPa]	Point C	
					Strain level [%]	Stress level [%]
1	Coarse sandstone	8	2.8	1.5	82.7	89.1
2	Coarse sandstone	12	2.8	1.5	76.7	83.7
3	Coarse sandstone	16	2.8	1.5	80.8	89.5
4	Coarse sandstone	20	2.8	1.5	62.0	81.2
5	Sandstone	12	2.8	1.5	74.3	83.8
6	Limestone	4	2.8	1.5	65.7	68.7
7	Limestone	4	2.8	1.5	65.3	73.7
8	Medium sandstone	4	2.8	1.5	83.4	75.2
9	Medium sandstone	4	3.8	1.8	80.7	76.4
10	Medium sandstone	4	3.8	1.8	85.2	82.2
11	Fine sandstone	4	3.8	1.8	88.1	82.6
12	Fine sandstone	4	3.8	1.8	73.7	73.5
13	Fine sandstone	20	12.0	1.8	77.4	79.5
Average					76.6	79.9

Changes in permeability were observed as a function of pressure for a variety of crystalline rock samples subjected to confining pressure and pore pressure. For most samples, the permeability dropped slightly when the pressure was less than about 10% of the fracture stress σ_d (stage OA shown in Fig. 1). It started to increase at σ_b and then increased drastically after σ_c was reached, until the maximum permeability was usually reached near to σ_e , which is located at the knee point of the constitutive curve. The stress levels, *i.e.*, σ_c and σ_d , represent important stages in the development of macroscopic failure of intact rocks.

From the laboratory tests, three characteristics of the evolution of permeability were identified (in the following text, strain and stress levels are the normalized strain and stress relative to their peak values, respectively, and permeability level refers to the permeability normalized by its initial value):

- The permeability increases at stress levels by approximately 30-50%.
- The permeability increases drastically near to the critical point C (Fig. 2); however, sometimes the stress level corresponding to this permeability threshold is greater than the stress level at point C (shown as point C_k in Fig. 1), and is as large as approximately 80% of the peak strength. This

result is roughly consistent with the observation of Lockner *et al.* (1992) on the onset of fracture coalescence, which was found to occur at stress levels of approximately 70-80% of the peak strength.

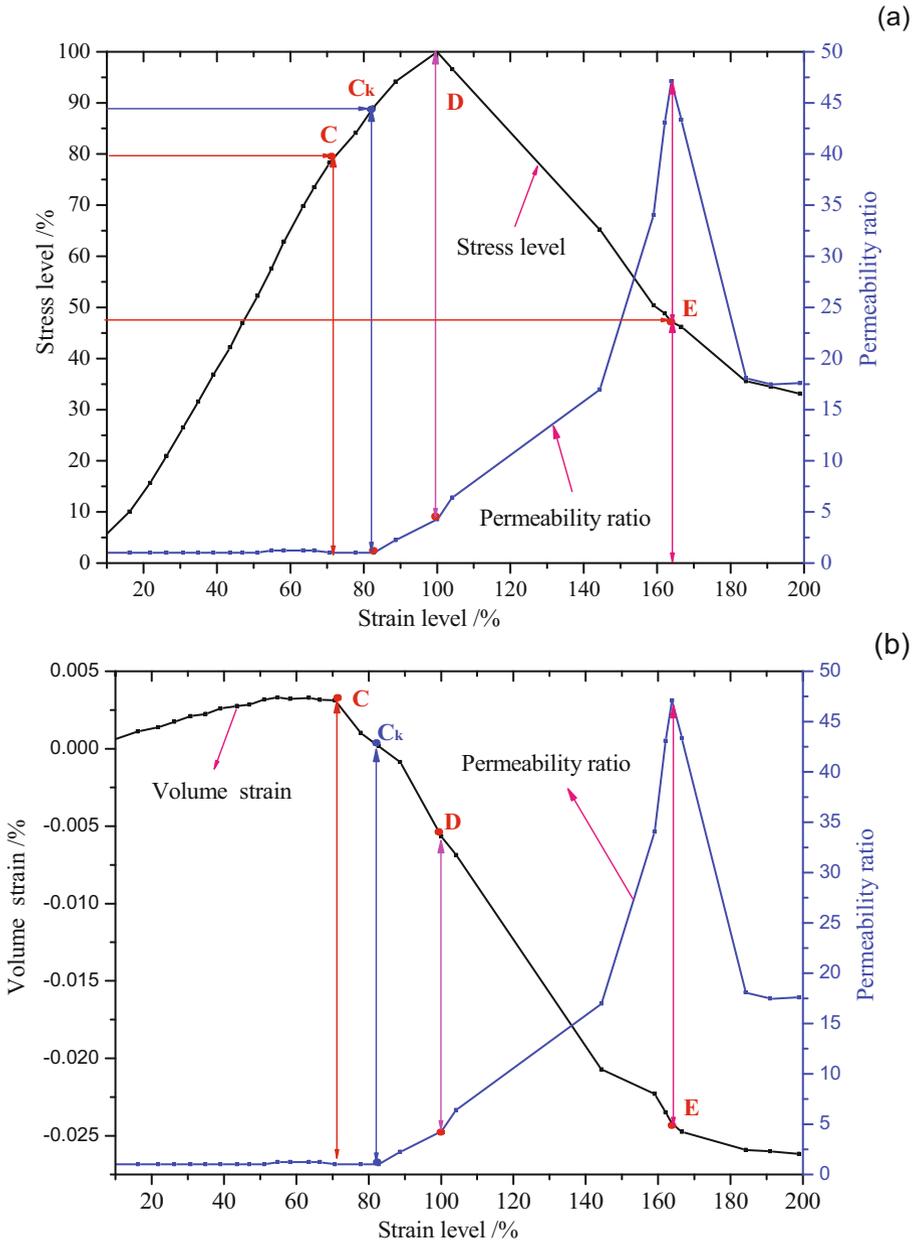


Fig. 2. Continued on next page.

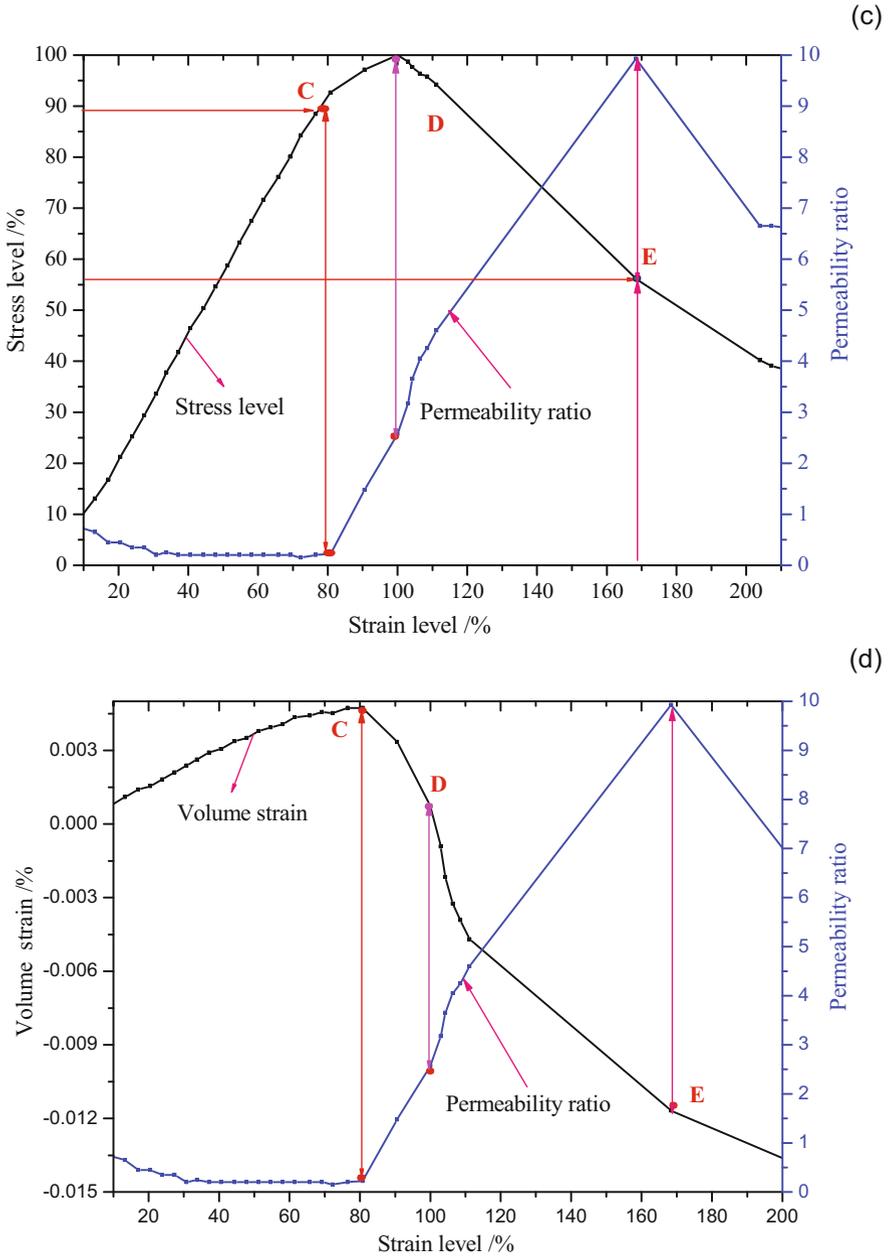


Fig. 2. Relationships between permeability ratio *versus* stress level during loading process for coarse sandstone: (a) coarse sandstone (sample 1 of Table 1), b) coarse sandstone (sample 1 of Table 1), (c) coarse sandstone (sample 3 of Table 1), and (d) coarse sandstone (sample 3 of Table 1).

- Macrocracks or shear-band formation normally occurs at point E, after which the permeability decreases.

3. THEORETICAL ANALYSIS OF STRESS AND STRAIN AT THE CRITICAL POINT

The rock material is considered as a system with a large number of intact rocks resulted from the complexity of formation environment and long-term geological processes. It is well accepted that as the applied load gradually increases, cracks will develop and coalesce around the potential macroscopic fracture plane (Xue 2014). Essentially, only when the crack density of the potential macroscopic fracture plane approaches a certain level, will the macroscopic fracture occur. Therefore, the forming process of the macroscopic fracture plane is simplified and considered as a two-dimensional fracture process in the present work, and two-dimensional renormalization group theory is introduced to describe the failure process of the rock samples (as shown in Fig. 3). Thus, the macroscopic fracture plane can be renormalized into many cells and different order blocks. Here, only the blocks of the first three orders are shown. For example, the first order block comprises four cells, while the second order block consists of four first order blocks. Similarly, the third order block is composed of four second order blocks. In fact, the same combination can be continued to an infinite scale. Furthermore, there exist five possible states for each of the different order blocks, *i.e.*, B4U0, B3U1, B2U2, B1U3, and B0U4 (as shown in Fig. 3d). The uppercase letter “B” denotes that the cells or blocks have broken (colored box), which is followed by the digit corresponding to the number of broken cells or blocks. The uppercase letter “U” means that the cells or blocks have not broken (white box), and the following digit corresponds to the number of unbroken cells or blocks. The corresponding broken probabilities of the five possible states are listed in column B, where p_1 means the broken probability of each cell.

As suggested by Smalley *et al.* (1985), Tang *et al.* (2000), and Wong *et al.* (2006), it is assumed that the strength of each individual cell is σ_{cell} , which obeys a Weibull distribution and depends on the number of micro cracks in the cells. When an external load is applied to a rock sample, it is assumed that each cell of the rock will be subjected to a corresponding local stress. When the strength of the cell σ_{cell} is less than its local stress, the cell will fail and its broken probability is p_α , which can be expressed as:

$$p_\alpha = p(\sigma_f < \alpha\sigma) = 1 - \exp\left[-\left(\frac{\alpha\sigma}{\sigma_0}\right)^m\right], \quad (1)$$

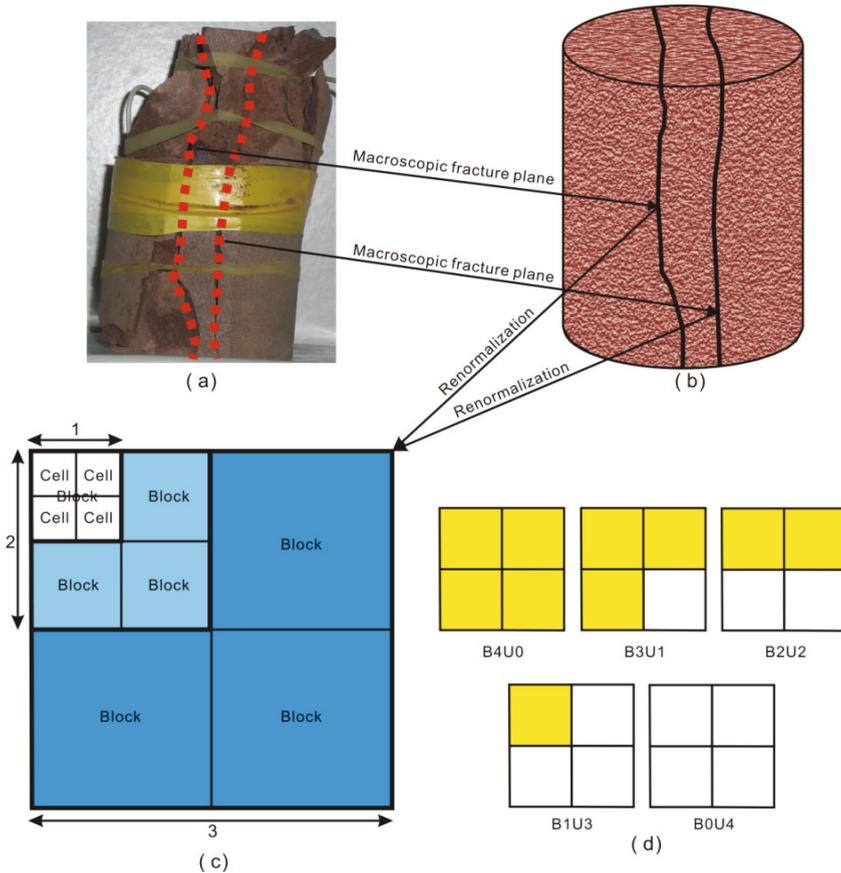


Fig. 3. Illustration of the two-dimensional renormalization group model of macroscopic fracture plane: (a) rock sample with two macroscopic fracture planes after failure under a uniaxial compression test, (b) sketch showing two macroscopic fracture planes, (c) the two-dimensional RG model, and (d) the five possible states for each block (Xue 2014).

where σ is the stress, α is a scale parameter, σ_0 is a measure of average strength, and m is the shape parameter that can be used to evaluate the discreteness of material strength.

The failure probability of elementary blocks corresponding to $\alpha = 1$ can be expressed as

$$p_1 = p(\sigma_f < \sigma) = 1 - \exp \left[- \left(\frac{\sigma}{\sigma_0} \right)^m \right]. \tag{2}$$

Combining Eqs. 1 with 2, one has

$$P_\alpha = 1 - (1 - P_1)^{\alpha^m} . \quad (3)$$

The conditional probability $p_{\alpha,b}$ for the occurrence of failure (Allegre *et al.* 1982) when a stress $(a - b)\sigma$ is transferred to an unbroken block under stress $b\sigma$ can be written as

$$p_{\alpha,b} = \frac{p(b\sigma < \sigma_f < a\sigma)}{p(\sigma_f > b\sigma)} = \frac{P_a - P_b}{1 - P_b} . \quad (4)$$

According to the renormalization group theory, the probability of destruction at the threshold point of destruction C, p_c , can be derived as

$$p_c = f_c(m) . \quad (5)$$

Equation 2 could also be expressed as a probability function of strain, that is,

$$p = 1 - \exp \left[- \left(\frac{\varepsilon}{\varepsilon_0} \right)^m \right] , \quad (6)$$

where ε is the axial strain of the rock, and ε_0 is a measurement of average strain.

As the stress-strain constitutive model of rocks can be expressed as (Qin *et al.* 2010a, b)

$$\sigma = E_0 \exp \left[- \left(\frac{\varepsilon}{\varepsilon_0} \right)^m \right] , \quad (7)$$

where E_0 is the initial elastic modulus, the strain value corresponding to the peak strength can be derived by taking the first derivative of Eq. 7 with respect to the stress ε , *i.e.*

$$\varepsilon_d = \left(\frac{1}{m} \right)^{1/m} \varepsilon_0 . \quad (8)$$

Inserting Eq. 5 into 6, and then combining with Eq. 8, the strain at the critical failure point C can be derived as

$$\varepsilon_c = \left\{ -\ln [1 - f_c(m)] \right\}^{1/m} \varepsilon_0 . \quad (9)$$

Divide Eq. 9 by 8, then we have

$$\lambda_c = \frac{\sigma_c}{\sigma_d} \times 100\% = \left\{ -m \ln [1 - f(m)] \right\}^{1/m} \exp \left[\frac{1}{m} + \ln [1 - f(m)] \right] \times 100\% . \quad (10)$$

Equation 10 gives the relationship between the critical stress level λ_c and the shape parameter.

As one or multiple macroscopic breaking surfaces with sags and crests are usually found on hard and brittle rocks, we assume that the micro-to-macroscopic destruction of rocks samples without joint fissure in the laboratory should be developed in 2D. In order to calculate $f(m)$ in the above equations, the rock material is considered as a system with a large number of intact rocks resulting from the complexity of formation environment and long-term geological processes. The critical behavior of brittle fracture of the rock material will be studied with the model as shown in Fig. 3, which is a 2D lattice renormalization. It is shown in Fig. 3 that a large block is composed of four small blocks, and it has the unique performance. And four large blocks can comprise a larger block, and so forth, that is, according to the process of the renormalization theory. For a group containing two blocks which are either broken or unbroken, the five states are possible: [BBBB], [BBBU], [BBUU], [BUUU], and [UUUU], where the letters “B” and “U” represent a broken block and an unbroken block, respectively. It is easy to conclude that the failure probability $p_1^{(2)}$ of the second-order blocks, which is expressed as Eq. 11, where p_1 is the failure probability of block and $(1 - p_1)$ is the unbroken probability.

$$p_1^{(1)} = p_{b_4u_0} + p_{b_3u_1} + p_{b_2u_2} + p_{b_1u_3} \quad , \quad (11)$$

where

$$p_{b_4u_0} = p_1^4 \quad , \quad (12)$$

$$p_{b_3u_1} = C_4^3 p_1^3 (1 - p_1) p_{4,1} \quad , \quad (13)$$

$$p_{b_2u_2} = C_4^2 p_1^2 (1 - p_1)^2 \left[p_{2,1}^2 + C_2^1 p_{2,1} (1 - p_{2,1}) p_{4,2} \right] \quad , \quad (14)$$

$$p_{b_1u_3} = C_4^1 p_1 (1 - p_1)^3 \left[\begin{array}{l} p_{4/3,1}^3 + C_3^2 p_{4/3,1}^2 (1 - p_{4/3,1}) p_{4,4/3} \\ + C_3^1 p_{4/3,1} (1 - p_{4/3,1})^2 \\ \left[p_{2,4/3}^2 + C_2^1 p_{2,4/3} (1 - p_{2,4/3}) p_{4,2} \right] \end{array} \right] \quad . \quad (15)$$

The threshold probability of destruction under 2D renormalized condition is computed, and the result, $p^* = f(m)$, is also plotted in Fig. 4. It is shown that the peak strength of rock increases with the decrease in shape parameter m , and the calculated 2D mean stress level is 83.2% for m equal to 1, 2, 3, 4, 5, 6, and 7, which is very close to the value of the Darley sandstone tested by Mordecai and Morris (1971), as shown in Fig. 5.

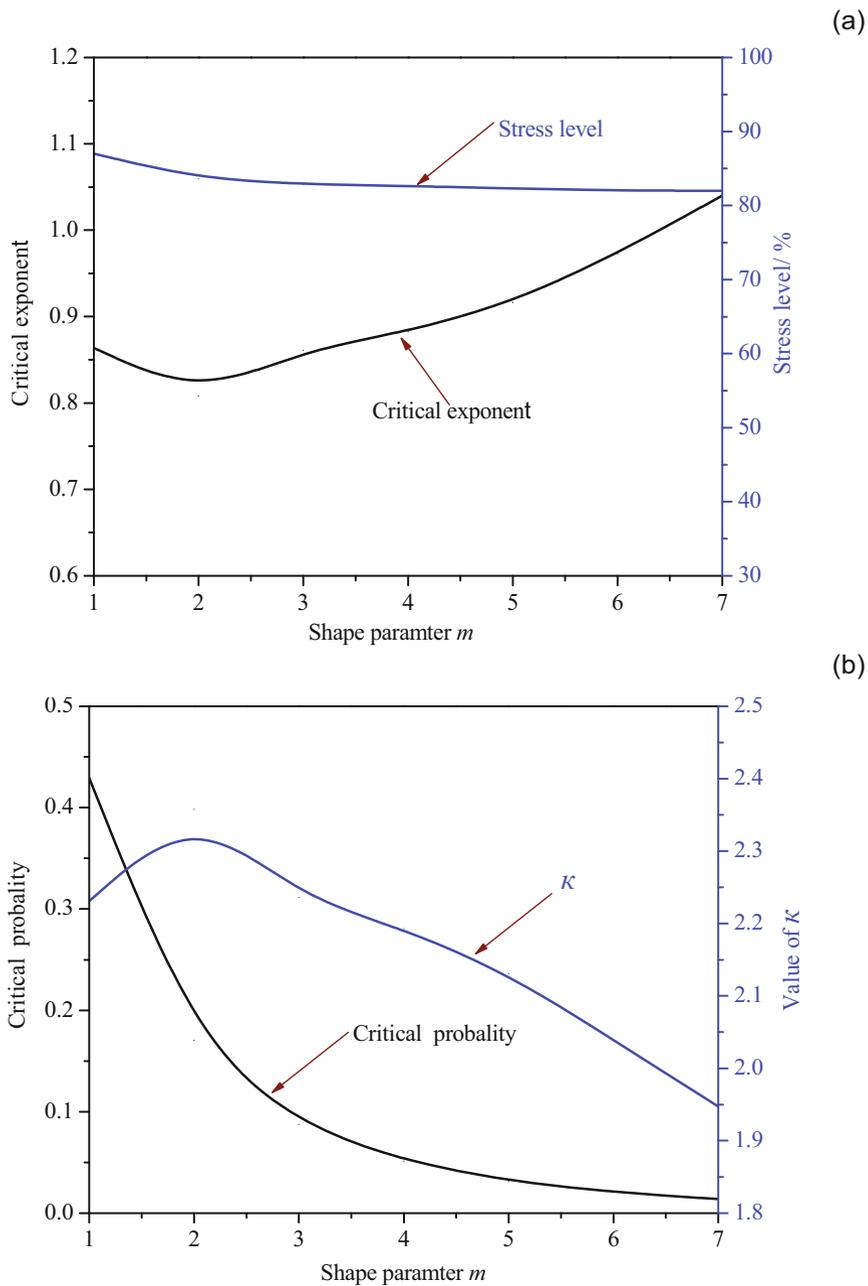


Fig. 4. The variation of critical probability, critical exponent, and critical stress level with the shape: (a) critical exponent and critical stress level, and (b) value of κ and critical probability.

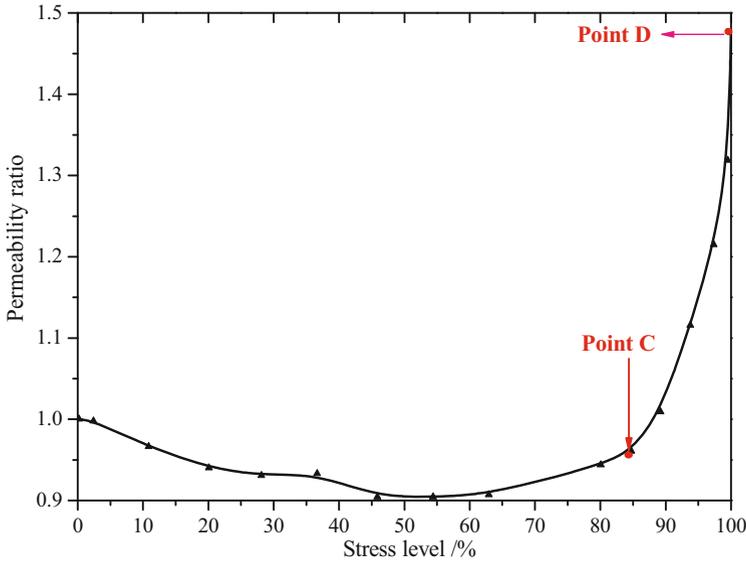


Fig. 5. Relationships of stress level *versus* permeability of Darley sandstone.

Correlation length is an important physical quantity for the renormalization group theory, which is used to mark the characteristics of permeability group. Near the threshold, the correlation length ξ can be expressed as

$$\xi = \xi_0 |p_c - p_1^{(n)}|^{-\nu} \quad (16)$$

where ν is the index of correlation length, and ξ_0 is the proportional constant corresponding to unit interval size.

Considering the scale invariance of the renormalization group theory, we have

$$\xi = \xi'_0 |p_c - p_1^{(n+1)}|^{-\nu} \quad (17)$$

where $\xi'_0/\xi_0 = b$; b is the scale factor, which is obtained by $b = N^{d-1}$; N is the block number of initial level in the renormalization transformation; and d is the spatial dimension.

According to Fig. 3, the value of b is 2. Combining Eqs. 16 with 17, one has

$$\nu = \frac{\ln b}{\ln \left(\frac{p_c - p_1^{(n)}}{p_c - p_1^{(n+1)}} \right)} \quad (18)$$

When the broken probability p is near to the critical value, there is

$$\frac{p_c - p_1^{(n)}}{p_c - p_1^{(n+1)}} \rightarrow \left. \frac{dp_{n+1}}{dp_n} \right|_{p^*}. \quad (19)$$

Let $\kappa = dp_{n+1}/dp_n$. When $\kappa > 0$, it means the unstable fixed point (critical point). Then we have:

$$v = \ln b / \ln \kappa. \quad (20)$$

The critical exponents of two-dimension renormalization group are shown by Fig. 4.

When the broken probability is lower than the percolation threshold p_c , the broken cells are almost isolated from each other, with less interaction. Contrarily, when the rupture probability is close to the percolation threshold p_c , the correlation length of broken cells increases suddenly, which results in the random, disordered broken cells cluster?

The renormalization group probability of destruction is then employed to calculate the stress at the threshold point of destruction, or the critical point, the stress corresponding to the peak strength, and then the critical stress level. As the permeability increases drastically near to the critical point C, the observation of the evolution of permeability may provide a new approach to identifying the critical point C in the stress-strain curve.

4. CASE STUDIES AND DISCUSSION

Case 1: Triaxial compression test of different sedimentary rock samples

Rock samples (shown in Table 1) were obtained from Yanzhou, Shandong, and cut into cylinders 50 mm in diameter and 100 mm long. They were tested on an electro-hydraulic servo-controlled testing machine (MTS815).

The test result is shown in Figs. 2 and 6. Based on the level of permeability (stress) change *versus* strain level (the stress level is normalized relative to its peak value, and the permeability ratio refers to the value of permeability normalized by the initial permeability obtained from experiments on rock samples), the variation of permeability with stress-strain level can be separated into 3 phases:

(i) In the compaction and elastic phases, permeability decreases with a very small degree with increasing stress (strain), which is mainly due to the closing of original micro fissures and pores.

(ii) After the elastic limited stress is reached, with the increase of loading, micro cracks appear and gradually expand. In this process, the permeability remains unchanged or slightly increases.

(iii) Near to the Point C, plenty of micro cracks initiate and grow rapidly, and micro breaking is developed spatially in rows along the potential break-

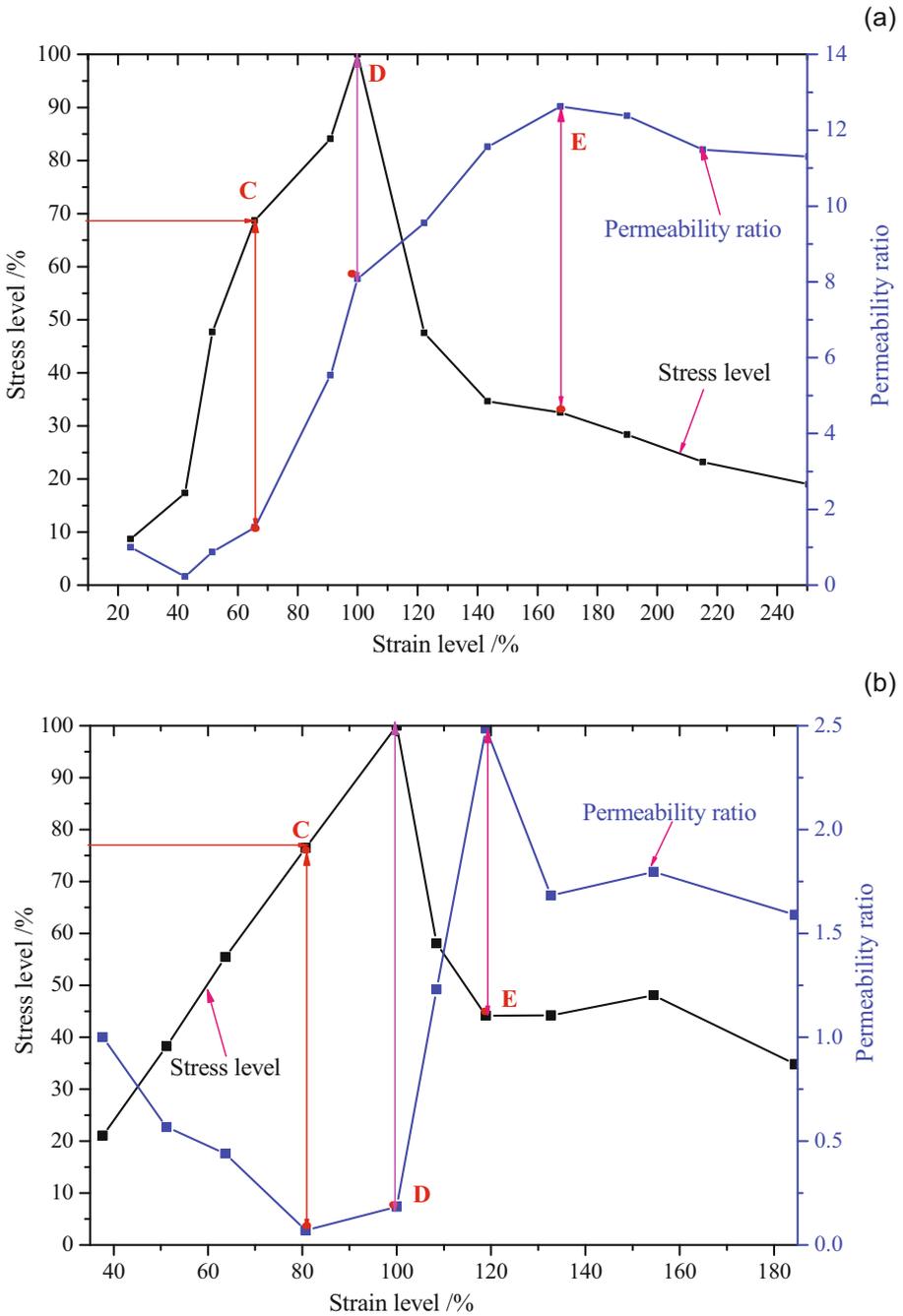


Fig. 6. Continued on next page.

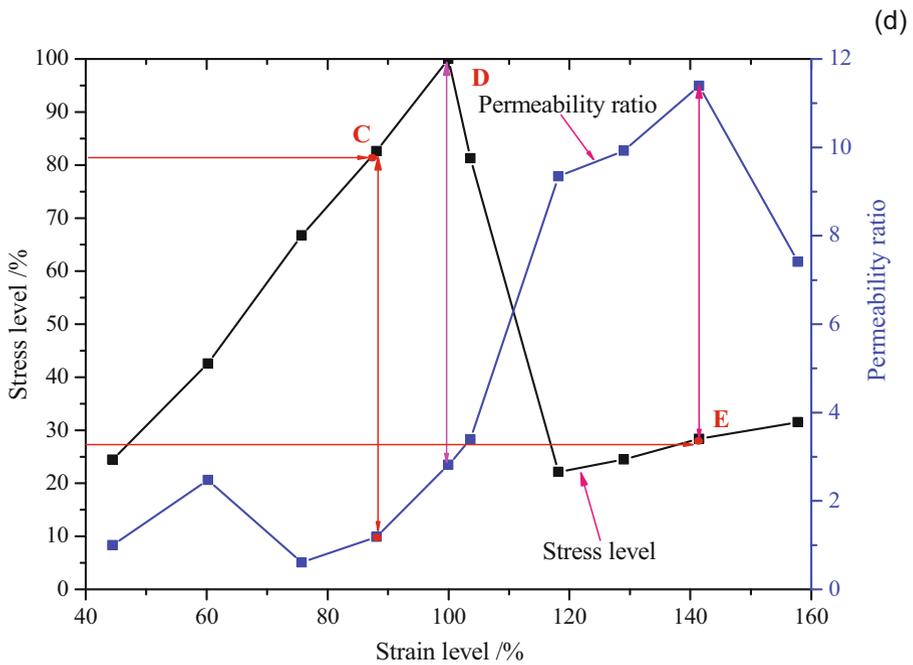
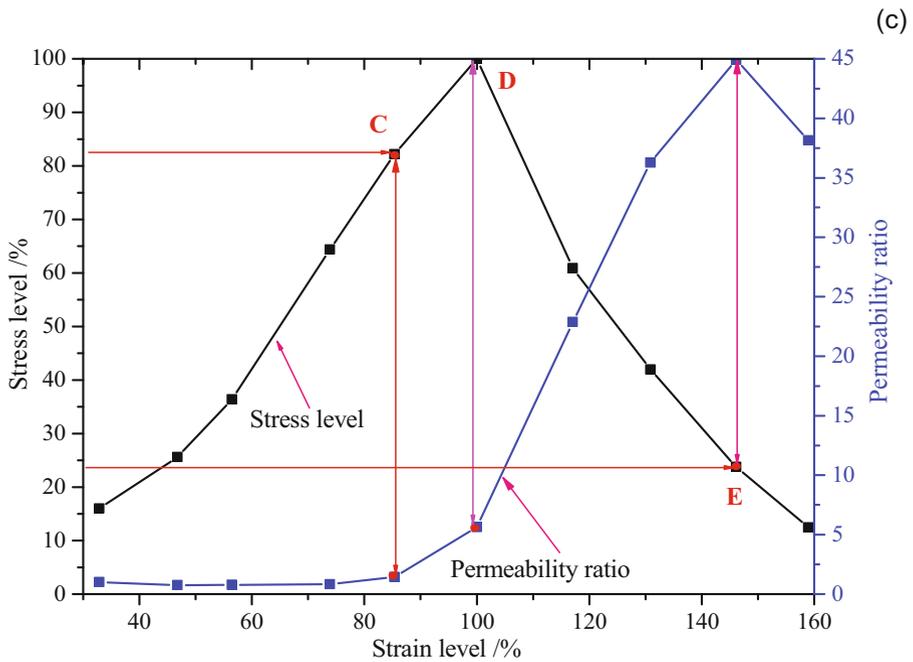


Fig. 6. Continued on next page.

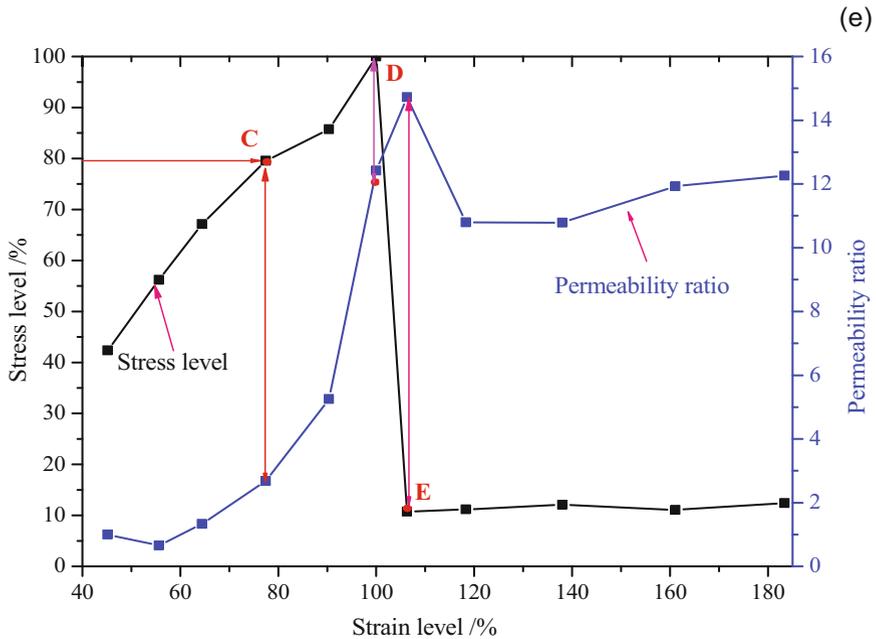


Fig. 6. Relationship between permeability, stress level and strain level: (a) limestone (sample 6), (b) medium sandstone (sample 9), (c) medium sandstone (sample 10), (d) fine sandstone (sample 11), and (e) fine sandstone (sample 13).

ing planes until the micro cracks eventually become connected with each other (Fig. 3), as indicated by the location of strain, and accelerated increase of volumetric strain at the cracks. In this phase, permeability has a precipitous increase.

From Fig. 6 and Table 1, we can infer that the stress level at the threshold point of destruction C is approximately 80% of the experimental result. This also suggests that the variation of rock permeability can reflect the changes of stress state, which may provide very useful information on rock stability.

Case 2: Triaxial compression test of granite sample

Rock samples were obtained from the Jining, Shandong, and cut into $\Phi 50 \times 100$ mm cylinders. They were tested on an electro-hydraulic servo-controlled testing machine (MTS815).

In the compaction and elastic phases, the permeability declines slightly with the increase of stress, mainly resulting from the closing of original micro fissures and pores (Fig. 7). When the ratio between the loading stress and peak stress near the point C (the average stress level is about 80%), the permeability increases sharply, with a jump in the vicinity of the yield point.

(a)

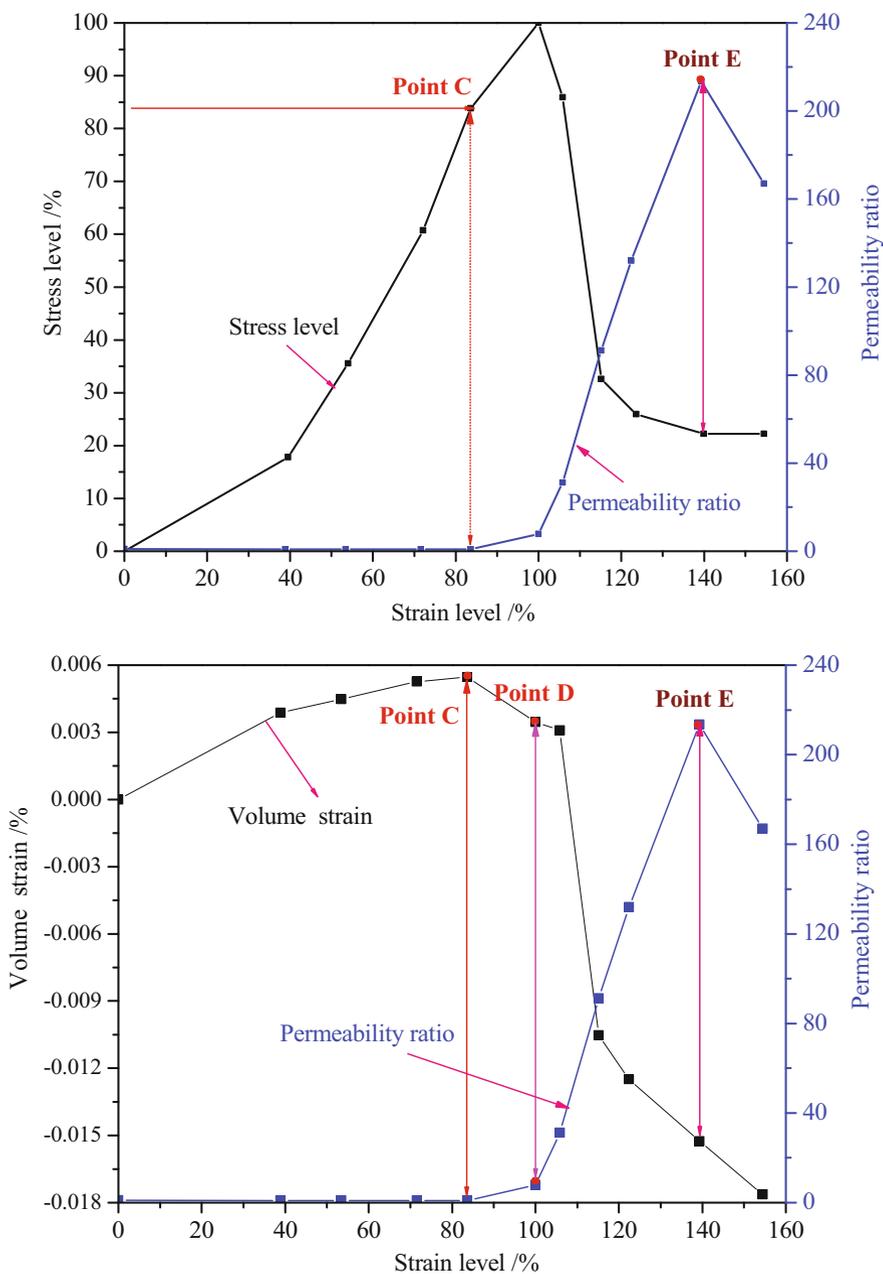


Fig. 7. Continued on next page.

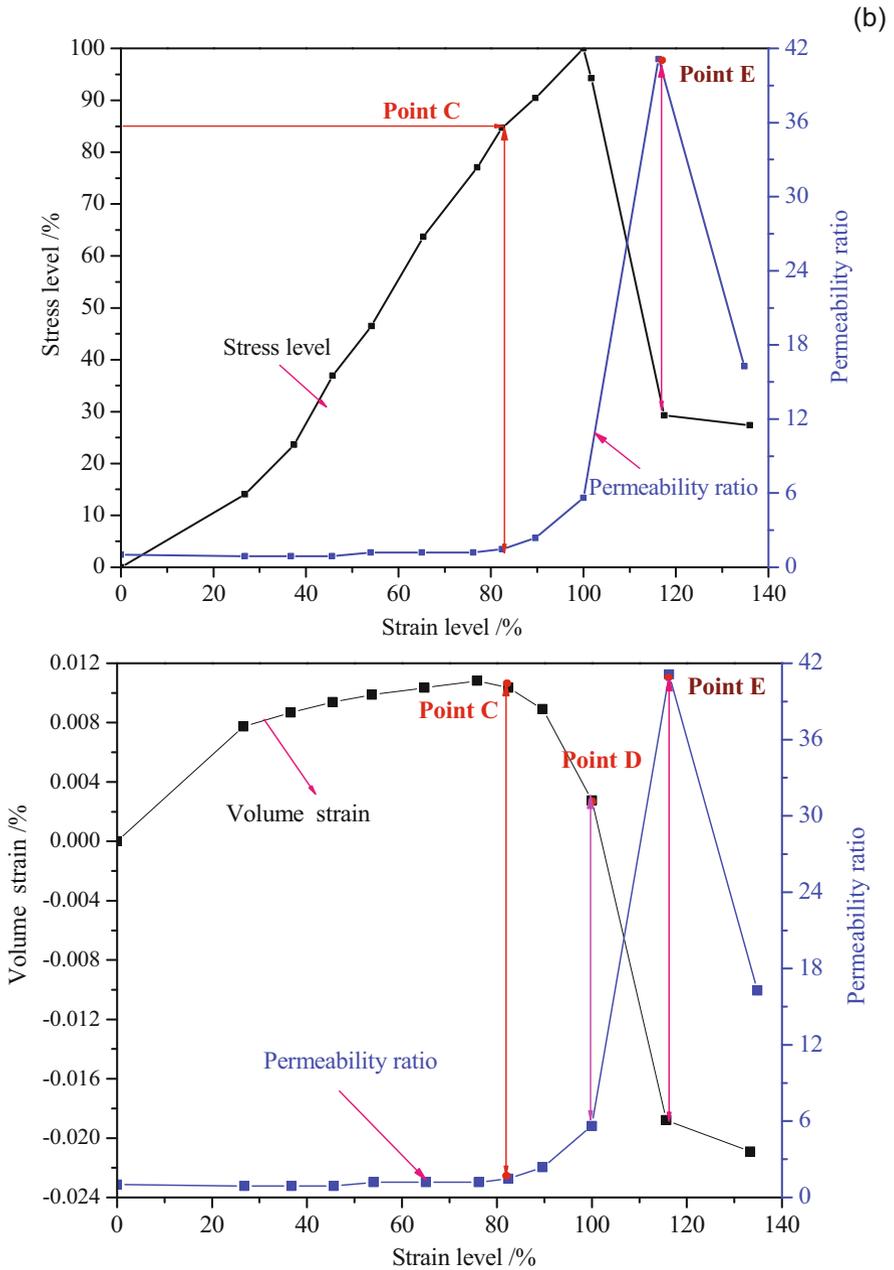


Fig. 7. Relationship between permeability, stress level, and strain level of granite samples: (a) sample 1 (confining pressure 4.0 MPa, pore pressure 2.8 MPa), and (b) sample 2 (confining pressure 6.0 MPa, pore pressure 2.8 MPa).

5. CONCLUSIONS

Generalized crack damage threshold is studied in this paper by monitoring stress, strain, and the evolution of permeability during loading processes and theoretically analyzing the stress and strain at the critical point, and the following conclusions are obtained:

- The evolution of permeability varies with the stages of rock fracturing during the process of compression. In the vicinity of the threshold point C (crack coalescence), with the rapid generation and growth of micro cracks inside the rocks, the permeability would increase in a large amount.
- In agreement with the statistical mean of the data collected from laboratory tests, onset of fracture coalescence starts at stress levels of approximately 80% of the peak strength, as predicted with the renormalization group probability of destruction.
- In most cases, the permeability increases sharply at the critical point in rock stress-strain curve. Observing the evolution of permeability may provide a new approach to identifying the critical point (yield point) C in the stress-strain curve.

These conclusions need more experiments for validation, and further investigation on their robustness under various conditions, so that physical predicative methods can be developed for recognizing the threshold information of destruction, promoting the interaction and incorporation of experimental science, rock mechanics, geophysics, and other subjects.

Acknowledgments. This research was supported by the State Basic Research and Development Program of China (No. 2013CB036003), the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, the Project supported by the National Science Youth Foundation of China (Grants Nos. 41102201, 41302233, 51309222) and the Project funded by China Postdoctoral Science Foundation (No. 2012M520376).

References

- Allègre, C.J., J.L. Le-Mouél, and A. Provost (1982), Scaling rules in rock fracture and possible implications for earthquake prediction, *Nature* **297**, 5861, 47-49, DOI: 10.1038/297047a0.
- Anifrani, J.-C., C. Le Floch, D. Sornette, and B. Souillard (1995), Universal log-periodic correction to renormalization group scaling for rupture stress pre-

- diction from acoustic emissions, *J. Phys. I France* **5**, 6, 631-638, DOI: 10.1051/jp1:1995156.
- Brace, W.F., J.B. Walsh, and W.T. Frangos (1968), Permeability of granite under high pressure, *J. Geophys. Res.* **73**, 6, 2225-2236, DOI: 10.1029/JB073i006p02225.
- David, C., B. Menendez, W. Zhu, and T.F. Wong (2001), Mechanical compaction, microstructures and permeability evolution in sandstones, *Phys. Chem. Earth A*, **26**, 1-2, 45-51, DOI: 10.1016/S1464-1895(01)00021-7.
- Jiang, Z., L. Ji, R. Zuo, and L. Cao (2002), Correlativity among rock permeability and strain, stress under servo-control condition, *Chin. J. Rock Mech. Eng.* **21**, 10, 1142-1446 (in Chinese).
- Liu, H.L., T.H. Yang, Q.L. Yu, and S.-K. Chen (2009), Experimental study on fluid permeation evolution in whole failure process of tuff, *J. Northeast. Univ. (Nat. Sci.)* **30**, 7, 1030-1033 (in Chinese).
- Lock, P.A., X. Jing, and R.W. Zimmerman (2002) Predicting the permeability of sandstone from image analysis of pore structure, *J. Appl. Phys.* **92**, 10, 6311-6319, DOI: 10.1063/1.1516271.
- Lockner, D.A., J.D. Byerlee, V. Kuksenko, A. Ponomarev, and A. Sidorin (1992), Observations of quasistatic fault growth from acoustic emissions. **In:** B. Evans, and T.F. Wong (eds.), *Fault Mechanics and Transport Properties of Rocks*, International Geophysics, Vol. 51, Academic Press, New York, 3-31, DOI: 10.1016/S0074-6142(08)62813-2.
- Martin, C.D. (1993), The strength of massive Lac du Bonnet granite around underground openings, Ph.D. Thesis, University of Manitoba, Manitoba, Canada.
- Martin, C.D. (1997) Seventeenth Canadian Geotechnical Colloquium: The effect of cohesion loss and stress path on brittle rock strength, *Can. Geotech. J.* **34**, 5, 698-725, DOI: 10.1139/t97-030.
- Mordecai, M., and L.H. Morris (1971), Investigation into the changes of permeability occurring in a sandstone when failed under triaxial stress conditions. **In:** G.B. Clark (ed.), *Dynamic Rock Mechanics (Proc. 12th Ann. Rock Mech. Symp., 16 November 1970, Rolla, USA)*, AIME, New York, 221-239.
- Paterson, M.S. (1978), *Experimental Rock Deformation – The Brittle Field*, Minerals and Rocks, Springer, Berlin Heidelberg.
- Paterson, M.S., and T.F. Wong (2005), *Experimental Rock Deformation – The Brittle Field*, 2nd ed., Springer, Berlin Heidelberg, 347 pp.
- Qin, S.Q., Y.Y. Wang, and P. Ma (2010a), Exponential laws of critical displacement evolution for landslides and avalanches, *Chin. J. Rock Mech. Eng.* **29**, 5, 873-880, DOI: 1000-6915(2010)05-0873-08 (in Chinese).
- Qin, S.Q., X.W. Xu, P. Hu, Y.Y. Wang, X. Huang, and X.H. Pan (2010b), Brittle failure mechanism of multiple locked patches in a seismogenic fault and exploration of a new approach to earthquake prediction, *Chin. J. Geophys.* **53**, 4, 1001-1014, DOI: 10.3969/j.issn.0001-5733.2010.04.02 (in Chinese).

- Smalley, R.F., D.L. Turcotte, and S.A. Solla (1985), A renormalization group approach to the stick-slip behavior of faults, *J. Geophys. Res.* **90**, B2, 1894-1900, DOI: 10.1029/JB090iB02p01894.
- Sornette, D. (2000), *Critical Phenomena in Natural Sciences. Chaos, Fractals, Self-organization and Disorder: Concepts and Tools*, Springer series in synergetics, Springer, Berlin Heidelberg.
- Sornette, D., and J.V. Andersen (1998), Scaling with respect to disorder in time-to-failure, *Eur. Phys. J. B* **1**, 3, 353-357, DOI: 10.1007/s100510050194.
- Sulem, J., and H. Ouffroukh (2006), Shear banding in drained and undrained triaxial tests on a saturated sandstone: Porosity and permeability evolution, *Int. J. Rock Mech. Min. Sci.* **43**, 2, 292-310, DOI: 10.1016/j.ijrmms.2005.07.001.
- Sun, Q., Z.Q. Jiang, and S.Y. Zhu (2012), Experimental study on permeability of soft rock of Beizao Coal Mine, *Chin. J. Geotech. Eng.* **34**, 3, 540-545, DOI: 1000-4548(2012)03-0540-06 (in Chinese).
- Tang, C.A., H. Liu, P.K.K. Lee, Y. Tsui, and L.G. Tham (2000), Numerical studies of the influence of microstructure on rock failure in uniaxial compression – Part I: effect of heterogeneity, *Int. J. Rock Mech. Min. Sci.* **37**, 4, 555-569, DOI: 10.1016/S1365-1609(99)00121-5.
- Walsh, J.B., and W.F. Brace (1984), The effect of pressure on porosity and the transport properties of rock, *J. Geophys. Res.* **89**, B11, 9425-9431, DOI: 10.1029/JB089iB11p09425.
- Wang, H.L., W.Y. Xu, and S.Q. Yang (2006), Experimental investigation on permeability evolution law during course of deformation and failure of rock specimen, *Rock Soil Mech.* **27**, 10, 1703-1708 (in Chinese).
- Wong, T.F., C. David, and W. Zhu (1997), The transition from brittle faulting to cataclastic flow in porous sandstones: Mechanical deformation, *J. Geophys. Res.* **102**, B2, 3009-3025, DOI: 10.1029/96JB03281.
- Wong, T.F., R.H.C. Wong, K.T. Chau, and C.A. Tang (2006), Microcrack statistics, Weibull distribution and micromechanical modeling of compressive failure in rock, *Mech. Mater.* **38**, 7, 664-681, DOI: 10.1016/j.mechmat.2005.12.002.
- Xue, L. (2014), A potential stress indicator for failure prediction of laboratory-scale rock samples, *Arab. J. Geosci.*, DOI: 10.1007/s12517-014-1456-1.
- Yang, Y.J., J. Chu, D.Z. Huan, and L. Li (2008), Experimental of coal's strain-permeability rate under solid and liquid coupling condition, *J. China Coal Soc.* **33**, 7, 760-764, DOI: 10.3321/j.issn:0253-9993.2008.07.008 (in Chinese).
- Zhang, R., Z.Q. Jiang, Q. Sun, and S.Y. Zhu (2013), The relationship between the deformation mechanism and permeability on brittle rock, *Nat. Hazards* **66**, 2, 1179-1187, DOI: 10.1007/s11069-012-0543-4.

-
- Zhu, W., and T.F. Wong (1996), Permeability reduction in a dilating rock: Network modeling of damage and tortuosity, *Geophys. Res. Lett.* **23**, 22, 3099-3102, DOI: 10.1029/96GL03078.
- Zhu, W., and T.F. Wong (1997), The transition from brittle faulting to cataclastic flow: Permeability evolution, *J. Geophys. Res.* **102**, B2, 3027-3041, DOI: 10.1029/96JB03282.
- Zhu, W., and T.F. Wong (1999), Network modeling of the evolution of permeability and dilatancy in compact rock, *J. Geophys. Res.* **104**, B2, 2963-2971, DOI: 10.1029/1998JB900062.
- Zoback, M.D., and J.D. Byerlee (1975a), Permeability and effective stress, *Bull. Am. Assoc. Petrol. Geol.* **59**, 154-158.
- Zoback, M.D., and J.D. Byerlee (1975b), The effect of microcrack dilatancy on the permeability of westerly granite, *J. Geophys. Res.* **80**, 5, 752-755, DOI: 10.1029/JB080i005p00752.

Received 9 December 2013

Received in revised form 18 May 2014

Accepted 9 July 2014



Estimating the Shapes of Gravity Sources through Optimized Support Vector Classifier (SVC)

Mohammad E. HEKMATIAN^{1,2}, Vahid E. ARDESTANI³,
Mohammad A. RIAHI³, Ayyub M.K. BAGH^{2,4}, and Jalal AMINI⁵

¹Faculty of Basic Sciences of Science and Research Branch, Islamic Azad University, Tehran, Iran; e-mail: mhekmatian@aeoi.org.ir (corresponding author)

²Nuclear Fuel Cycle Research School of Nuclear Science and Technology Research Institute (NSTRI), Tehran, Iran

³Institute of Geophysics, University of Tehran, Tehran, Iran
e-mails: ebrahimz@ut.ac.ir, mariahi@ut.ac.ir

⁴Faculty of Engineering of South Tehran Branch, Islamic Azad University, Tehran, Iran; e-mail: amemar@aeoi.org.ir

⁵Faculty of Engineering, University of Tehran, Tehran, Iran; e-mail: jamini@ut.ac.ir

Abstract

In gravity interpretation methods, an initial guess for the approximate shape of the gravity source is necessary. In this paper, the support vector classifier (SVC) is applied for this duty by using gravity data. It is shown that using SVC leads us to estimate the approximate shapes of gravity sources more objectively. The procedure of selecting correct features is called feature selection (FS).

In this research, the proper features are selected using inter/intra class distance algorithm and also FS is optimized by increasing and decreasing the number of dimensions of features space. Then, by using the proper features, SVC is used to estimate approximate shapes of sources from the six possible shapes, including: sphere, horizontal cylinder, vertical cylinder, rectangular prism, syncline, and anticline. SVC is trained using 300 synthetic gravity profiles and tested by 60 other synthetic and some real gravity profiles (related to a well and two ore bodies), and shapes of their sources estimated properly.

Key words: gravity sources shapes, SVC, feature, gravity profile, FS.

1. INTRODUCTION

Usual methods for estimation of gravity source shape are the forward method and the non linear inversion method. However, for making the shape estimation of gravity source, both of them need to have an initial assumption of the shape of the gravity source. There is no usual method for making the initial assumption of shape of gravity source and practically this assumption is prepared by non-gravity (for example geological) data. An attempt to estimate an approximate shape of gravity sources (as the initial assumption) using artificial neural network was done by Gret and Klingele (1998). Support vector classifier (SVC) algorithm could be used to estimate approximate shapes of gravity sources more objectively. SVC is one of the pattern recognition (PR) algorithms. When the input values (or objects) are different and these values have different labels (or are related to different classes), PR finds the classes of the above-mentioned objects based on some specific algorithms. An example of PR could be to determine whether a given gravity profile is created by a spherical or non-spherical gravity source. PR has different types according to the kind of learning procedure used to define the classes of the objects. Learning procedures consist of supervised learning and unsupervised learning approaches. In the supervised learning approach, the training data should be available, consisting of some objects that have been properly classified. In the unsupervised learning approach, training data are not available and the procedure tries to find some patterns in the data that can be used to determine different kinds of input values. SVC is considered as a supervised learning algorithm of PR (Heijden *et al.* 2004). In gravity, producing training data is possible and therefore it is advisable to use a supervised learning algorithm. While producing the training data in gravity, it should be noted that we need to create several synthetic gravity profiles caused by synthetic gravity sources with different definite shapes. In this research, a set of synthetic gravity data caused by definite synthetic source shapes such as sphere, vertical cylinder, horizontal cylinder, anticline, syncline, and rectangular prism are prepared and considered as properly classified input values (training set). Then values of proper features from each of synthetic gravity profiles are extracted (see Sections 2.2 and 3). These values are used to train the supervised learning algorithm (SVC) and this trained algorithm would be able to estimate the shapes of sources of other gravity profiles including the real ones; as shown in Section 4.2, some real gravity sources have been estimated by SVC. For example, the shape of a well (a real gravity source) is estimated by the trained SVC with a shape of vertical cylinder. Estimation of gravity source shapes is a classification in which each gravity source shape is a class.

Between different PR algorithms for gravity interpretation and other geophysical applications, unlike the SVC, neural network has been used frequently (Gret and Klingele 1998, Baan and Jutten 2000, Osman *et al.* 2006). Before this research, SVC has been used for depth estimation of faults (Hekmatian *et al.* 2013). However, according to our knowledge, SVC has not been used up to now for estimation of approximate shape of gravity sources. SVC algorithm has very suitable properties, such as a single global solution and ability to use different kernels. In this research we have shown that SVC algorithm can be adopted as a suitable method for estimating the approximate shapes of gravity sources.

2. THEORIES AND ALGORITHMS

2.1 Support vector classifier (SVC) algorithm

2.1.1 Linear SVC

Prior to linear SVC description we should note that support vector machine (SVM) is a concept in computer science for a set of related supervised learning methods that analyse data and recognize patterns used for classification and regression analysis; the original SVM formulations for classification is named SVC. SVC is clarified briefly in this and next sections, according to Heijden *et al.* (2004).

Linear (and non-linear) SVC considers only 2 classes of information that can be separated by a linear border. One class is located on one side of the linear border and the other is located on the other side of the linear border. If there are more than 2 classes, all classes except of one should be considered as one class and the remaining one as the other class. For example, we can consider the spherical shape of the gravity source as one class and all other possible shapes of the gravity source as another class. We should keep in mind that in SVC the border between the two classes is not a line but a margin and this border line is in the centre of the margin. The linear discriminant function that is used in linear SVC is (Heijden *et al.* 2004):

$$\mathbf{g}(\mathbf{Z}_n) - \mathbf{w}^* \mathbf{Z}_n + b, \quad (1)$$

where \mathbf{Z}_n is the matrix of the values of features (measurement vectors), \mathbf{w} is the vector of coefficient, and b is a scalar value. The rows of \mathbf{Z}_n are the features of each object (each gravity profile) and the columns of this matrix show the kind of features in all objects. Two classes ($C_n = 1$ and $C_n = -1$ are labels of two classes) are defined by Heijden *et al.* (2004) as:

$$\mathbf{w}^T \mathbf{Z}_n + b \geq 1 \quad \text{if } C_n = +1, \quad (2)$$

For all n

$$\mathbf{w}^T \mathbf{Z}_n + b \leq -1 \quad \text{if } C_n = -1, \quad (3)$$

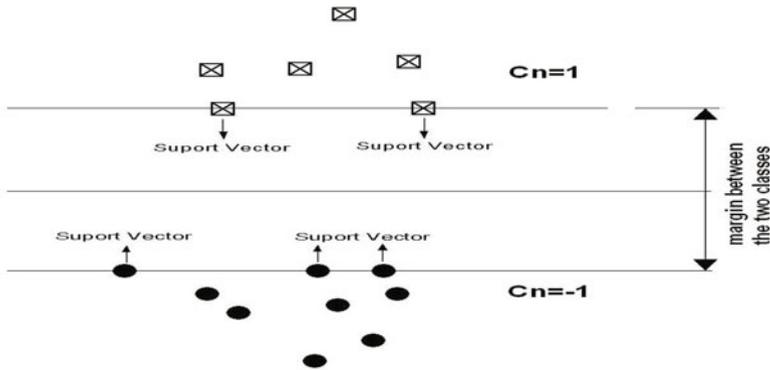


Fig. 1. The linear support vector classifier. The boundary between 2 classes is not only a margin but also a margin with maximum width and this shows the single global solution of SVC.

where \mathbf{w}^T means transpose of \mathbf{w} . The length of the margin between these two classes is shown in Fig. 1.

For more explanations regarding linear SVC, one of proper references is Heijden *et al.* (2004).

2.1.2 Non linear SVC

At first, some definitions should be recalled. The decision function is a function that maps the measurements space (features space) onto the set of possible classes. It should be emphasized that in SVC only 2 classes exist. Therefore, in linear and non-linear SVC, decision function consists of only 2 discriminant functions, $\mathbf{g}(\mathbf{Z}) \geq 1$ and $\mathbf{g}(\mathbf{Z}) \leq -1$; see Eqs. 1-3.

To have non-linear boundaries in SVC, the discriminant function should be defined as (Heijden *et al.* 2004):

$$\mathbf{g}(\mathbf{Z}) = \mathbf{w}^T \mathbf{y}(\mathbf{Z}) = \sum_{n=1}^{N_s} \mathbf{k}(\mathbf{Z}, \mathbf{Z}_n), \quad (4)$$

where $\mathbf{k}(\mathbf{Z}, \mathbf{Z}_n)$ is the kernel (such as polynomial one), and N_s is the number of samples in the training set. In Equation 4, $\mathbf{g}(\mathbf{Z}) = 1$ and $\mathbf{g}(\mathbf{Z}) = -1$ show the upper and lower limits of the margin of non-linear SVC, respectively (see Fig. 1 for comparing with linear SVC). Gaussian kernel could be used to develop a non-linear SVC and it can have a weighting matrix equal to $\sigma^2 \mathbf{I}$ (\mathbf{I} is the unit matrix). In this case, we have radial basis function kernel (RBF kernel) that has been applied in this study (Heijden *et al.* 2004):

$$\mathbf{k}(\mathbf{Z}, \mathbf{Z}_n) = \exp\left(-\frac{\|\mathbf{z} - \mathbf{z}_n\|^2}{\sigma^2}\right), \quad (5)$$

where σ^2 is a parameter for decreasing or increasing the power of exponential function of Eq. 5.

2.2 The features and feature selection (FS) algorithm

Feature is a kind of measurement value which is obtained from each object. In this paper, object is a gravity profile which is created by a specific gravity source. Features that can be used for defining the shape of the gravity sources are derivable from a principal gravity profile. Gret and Klingele (1998) noted that the principal gravity profile is the gravity profile passing through the maximum value along the anomaly and crossing the anomaly lines perpendicularly. Gravity profiles in this research mean principal gravity profiles.

Some examples of the different features used to characterize the gravity anomalies are defined as (Gret and Klingele 1998):

$$F1 = X_{g50} / X_{g75} \quad , \quad (6)$$

$$F2 = X_{g75} / X_{g50} \quad , \quad (7)$$

$$F3 = X_{g_{inf}} / X_{g75} \quad , \quad (8)$$

$$F4 = (X_{g50} - X_{g66}) / (X_{g80} - X_{g66}) \quad , \quad (9)$$

where g_y is the value of gravity (g) at $y\%$ of maximum value of gravity g_{max} and X_{g_y} is the value of X at g_y (to be more specific, in our research X is the horizontal distance in the direction of the profile from the location of g_{max} in the profile). Also, $X_{g_{inf}}$ is the value of X at g_{inf} and g_{inf} is the value of gravity (g) at inflection point of the gravity profile. In Figure 2 an example of a

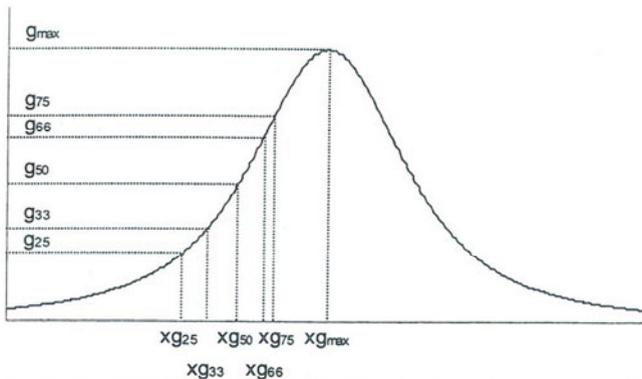


Fig. 2. Some of the parameters of features describing the shape of gravity sources. In this figure, g_y is the value of gravity (g) at $y\%$ of maximum value of gravity g_{max} and X_{g_y} is the value of X at g_y (Gret and Klingele 1998).

gravity profile and some of the parameters of the features describing the shape of gravity sources are shown.

Choosing proper features (feature selection, FS) is very important while dealing with high misclassification in testing our classifier (more proper features are those features which, when used in classification, lead to better separation of the classes). The high misclassification in our selected features space means that the different classes are not separated properly. One of the enable FS algorithms is “inter/intra class distance” (IICD) and we have used this algorithm in our research. The IICD algorithm, like all other FS algorithms, has a criterion for selection of more proper features. This criterion is called performance measure (PM). PM increases with increasing suitability of the features for classification, and *vice versa*; in other words, the selection of more proper features is done when their PM is bigger. PM of IICD is based on Euclidean distance between each of the two objects in the training set (all objects used for training SVC) in the features space. For more explanation about “inter/intra class distance” (IICD) algorithm, we should mention that, as Heijden *et al.* (2004) noted, T_s is a training set with N_s samples (objects). The classes w_k are represented by subsets T_k (subsets T_k are subsets of objects of T_s), each class having N_k samples. Measurement vectors in T_s without reference to their class are denoted by \mathbf{Z}_n (For each object we measure a set of features. This set of features is called measurement vector. The point location of each object in features space is defined with its measurement vector.). Measurement vectors in T_k (*i.e.*, vectors coming from class w_k) are denoted by $\mathbf{Z}_{k,n}$. The sample mean of class w_k is (Heijden *et al.* 2004):

$$\hat{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{Z}_{k,n} . \quad (10)$$

The sample mean of the entire training set is (Heijden *et al.* 2004):

$$\hat{\mu}_k = \frac{1}{N_s} \sum_{n=1}^{N_s} \mathbf{Z}_n . \quad (11)$$

We can represent the distances between objects or samples (of the training set) by means of scatter matrices. A scatter matrix gives some information about the dispersion of a population of samples around their mean. For instance, the matrix that describes the scattering of vectors from class w_k is (Heijden *et al.* 2004):

$$\mathbf{S}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} (\mathbf{Z}_{k,n} - \hat{\mu}_k)(\mathbf{Z}_{k,n} - \hat{\mu}_k)^T , \quad (12)$$

where \mathbf{S}_k supplies information about the average distance of the scattering in class w_k . Averaged over all classes, the scatter matrix is (Heijden *et al.* 2004):

$$\mathbf{S}_w = \frac{1}{N_s} \sum_{k=1}^K N_k \mathbf{S}_k = \frac{1}{N_s} \sum_{k=1}^K \sum_{n=1}^{N_k} (\mathbf{z}_{k,n} - \hat{\mu}_k)(\mathbf{z}_{k,n} - \hat{\mu}_k)^T . \quad (13)$$

This matrix is “the within-scatter matrix”, as it describes the average scattering within classes. In addition, there is “the between-scatter matrix”, \mathbf{S}_b , that describes the scattering of the class-dependent sample means around the overall average (Heijden *et al.* 2004):

$$\mathbf{S}_b = \frac{1}{N_s} \sum_{k=1}^K N_k (\hat{\mu}_k - \hat{\mu})(\hat{\mu}_k - \hat{\mu})^T . \quad (14)$$

We know that an individual number is a kind of matrix. So, when “the within-scatter matrix” (\mathbf{S}_w) is an individual number, it shows the squared average distance of the location of each object in one class from the average location of all objects of that class in features’ space for all the classes. In this case, we name \mathbf{S}_w as intraclass distance, and when “the between-scatter matrix” (\mathbf{S}_b) is an individual number, it refers to the squared average distance of the average location of all objects of each class from the average location of all objects of all classes in features space. In this case, we name \mathbf{S}_b as interclass distance. The features that are more suitable for classification are the ones for which their interclass distance is greater than their intraclass distance. The performance measure (PM) suited to express the separability of classes is the ratio between interclass and intraclass distance. Therefore, features with greater PM are preferred for classification (according to IICD algorithm).

Also we should take into consideration the fact that the number of possible subsets, $q(D)$, consisting of D features between N features is (Heijden *et al.* 2004):

$$q(D) = \frac{N!}{(N-D)!D!} . \quad (15)$$

We have used the FS approach, as will be described in what follows.

3. THE PROCEDURE

In this procedure it was necessary to develop a software, which was done. For producing synthetic gravity profiles from different gravity sources and for extracting features from each gravity profile, a software package was produced using Visual Basic language. The algorithms of gravity calculation which we used for preparing the above-mentioned software were obtained from different references. The algorithm of gravity calculation with sources of spherical shape was obtained from Telford *et al.* (1976), and that of anticline or syncline shape from Talwani *et al.* (1959), with rectangular prism

shape obtained from Plouff (1976), horizontal cylinder shape obtained from Blakely (1996) and vertical cylinder shape obtained from Belikov (1978). For features selecting (FS) we produced 3 softwares using Visual Basic language. The algorithms of FS which we used for preparing our softwares were obtained from Heijden *et al.* (2004). These softwares are, respectively, able to select more suitable 2, 3, and 4 features out of 10 features. Also, we produced and trained tens of SVC codes for estimation of the gravity sources shapes. For producing the codes we used Matlab environment and also some tools of PRTools 4.1 and its manual (Duin *et al.* 2007) which can be found in website <http://www.prtools.org/>.

The procedure of this study includes 6 steps as follows:

(1) Three hundred synthetic gravity profiles (training set) were created by gravity sources with 6 definite shapes and different dimensions, depths and density contrasts for training SVC. For each of the 6 shapes, 50 synthetic gravity profiles were created. The specifications of the sources of these 300 synthetic gravity profiles are presented in Table 1.

Table 1

Gravity sources with different shapes, dimensions, depths, and density contrasts used for creating synthetic gravity profiles (to be used for training set)

Shapes	Radius [m]	Dimensions except radius [m]	Minimum depth [m]	Maximum depth [m]	Density contrast [gr/cc]
Sphere	6 to 365		5 to 600		0.5 to 1.5
Horizontal cylinder	5 to 240		5 to 600		0.5 to 1.2
Vertical cylinder	5 to 500	5 to 100 (height)	5 to 500		0.5 to 1.1
Rectangular prism		6 to 1100 (in direction of the profile) 2 to 100 (perpendicular to the profile)	3 to 600	5 to 700	0.4 to 1.2
Anticline		4 to 185 (vertical extension) 20 to 400 (horizontal extension)	5 to 500		1.0
Syncline		4 to 185 (vertical extension) 20 to 400 (horizontal extension)	5 to 500		1.0

(2) We extracted 10 features from each of the above-mentioned 300 synthetic gravity profiles. These ten features are shown in Table 2.

Table 2

The extracted features

Number of the extracted features	Extracted features
1	(X_{g50}/X_{g80})
2	(X_{g60}/X_{g80}) ,
3	(X_{ginf}/X_{g80})
4	(X_{g50}/X_{g70})
5	(X_{g60}/X_{g70})
6	(X_{ginf}/X_{g70})
7	(X_{g50}/X_{g90})
8	(X_{g60}/X_{g90})
9	(X_{ginf}/X_{g90})
10	$((X_{g50} - X_{g80})/X_{g90})$

Note: All the parameters are introduced in Section 2.2.

As shown in Table 2, we specified each of the extracted features with a number from 1 to 10. By having these ten features from each of the profiles, we would be able to select more suitable features for classification according to IICD algorithm (see Section 2.2). We should mention that these 10 features are good references for FS because they are related to the shape of the gravity sources and are independent of the depths and dimensions of the sources and also consist of all the X_{gy} from X_{g50} to X_{g90} , systematically.

(3) We selected more properly 2, 3, or 4 features out of 10. According to Eq. 15, the possible numbers of 2, 3, or 4 features (subsets) are 45, 120, and 210, respectively. But from these 45, 120, and 210 subsets (a total of 375 subsets), respectively, only 10, 17, and 14 have relatively large performance measures (see Section 2). By the values of the features of these 10, 17, and 14 subsets we trained, respectively, 10, 17, and 14 SVC codes for classification or, better to say, for estimation of the gravity sources shapes.

(4) The best trained SVC codes were related to only one subset of 3 features (includes (X_{g50}/X_{g80}) , (X_{ginf}/X_{g80}) , and $((X_{g50} - X_{g80})/X_{g90})$, *i.e.*, features 1, 3, and 10 regarding to Table 2) that was able to classify more than 0.70 of objects in the training set correctly (therefore, the best subset of features is the subset not only with big PM but also should classify more objects correctly than the other subsets of features). So we tested this best trained SVC with 60 more synthetic gravity profiles (testing set). The above-

mentioned testing set consists of 10 profiles for each of the 6 shapes which we want to classify. The best trained SVC codes were able to classify the testing set properly (so that the best trained SVC code is the one that is able to classify the testing set most properly). But for more accurate estimation, we decided to train more SVC codes with a different approach (explained in the next step). Before going to the next step, let us make two comments. First, it should be mentioned that in testing all the trained SVCs, we inserted, as a test, some noises in some of gravity profiles with different values equal to about one twentieth of the average value of the gravity profiles. These noises caused a little effect on the values of the extracted features (and so caused a little effect on the success of classification) except on those features (of Table 2) that consist of $X_{g \text{ inf}}$ (see Section 2.2). This is so because $X_{g \text{ inf}}$ is the value of X at $g \text{ inf}$ and $g \text{ inf}$ is the value of gravity (g) at inflection point of the profile and, by inserting noise, new inflection points will be created and it will become too hard to locate the principal inflection point in this situation. The second comment is related to the learning procedure. For learning SVC, it is needed to create a training dataset consisting of the features of each object in the dataset and a number (1 or 2) indicating the class of each object (SVC is considering 2 classes only). Then SVC (with the kernel (RBF) and its parameter σ (see Eq. 5) will use the training dataset to designate the boundary between the two classes. By trials and errors, it is found that the proper value for the parameter σ is 0.05. With RBF kernel with the above-mentioned value for its parameter, the misclassification was (with respect to other kernels and parameters) low. The SVC code with this kernel and the best subset of features can be assumed as the best SVC code.

(5) For better estimation of gravity sources shapes, we introduced 4 classification groups, such that each group consists of two classes and each class consists of three shapes. These 4 groups are shown in Table 3.

In other words, for better classification, the six shapes are classified by four classification groups and each of these 4 classification groups consists of two classes, while each class consists of three shapes. These four groups and related information are shown in Table 3. For each of the classification groups we used a couple of codes. For example, according to Table 3, for the first classification group one SVC code is trained to separate Elongated shapes from Non-Elongated shapes and one other SVC code trained to separate Limited shapes from Non-Limited shapes. These two codes do the same duty (separating Elongated shapes from Limited shapes) but one of these two codes tries to detect Limited shapes (those objects that are not detected as Limited shapes will be introduced as Non-Limited shapes) and the other code does the same duty for Elongated shapes. If we use all these 4 couples of codes, we will be able to classify all the 6 gravity sources shapes, because each of the 6 shapes has different responses according to all the 4 couples of

Table 3

Four classification groups of 2 three shapes classes

Number of group	Name of class 1	The related gravity sources shapes of class 1	Name of class 2	The related gravity sources shapes of class 2
1	Elongated shapes	Horizontal cylinder, Anticline and Syncline	Limited shapes	Vertical cylinder, Rectangular prism and Sphere
2	Up curved	Horizontal cylinder, Anticline and Sphere	Up Flat	Vertical cylinder, Rectangular prism and Syncline
3	Down curved	Horizontal cylinder, Syncline and Sphere	Down Flat	Vertical cylinder, Rectangular prism and Anticline
4	Circle section	Horizontal cylinder, Vertical cylinder and Sphere	Non Circle section	Rectangular prism, Anticline and Syncline

codes. For example, horizontal cylinder based on the codes related to Group 1 through Group 4 is classified, respectively, as: Elongated shapes, Up curved, Down curved, and Circle section. Likewise, anticline is classified, respectively, as: Elongated shapes, Up curved, Down flat, and Non Circle section. In this way, all the 6 shapes have different classes regarding all the four groups. Therefore, only by using these 4 couples of codes we are able to classify all the 6 shapes of the gravity sources. By using these 4 couples of codes, most of the 60 synthetic gravity profiles that were used for testing were classified correctly and testings with real gravity profiles were good. Considering Table 3, the six possible shapes of gravity sources in our research have the classes as presented in Table 4.

On the other hand, using all these 4 couples of codes will give us opportunity to confirm or reject the result of each other. For example, if the result of one of these 4 couples of codes is Elongated shape and results of 2 others are Up flat and Down flat, we will understand that the gravity profile is not interpretable by these 4 couple of codes. The reason is that none of the 6 shapes have these 3 properties (Elongated shape, Up flat, Down flat) simultaneously (see Table 4.). In such conditions, the gravity profile may not be a principal one (see Section 2.2).

(6) Although the above-mentioned codes until now have been sufficient for classifying the 6 shapes, for classifying with greater certainty, we trained more SVC codes. Based on Table 3 we have 8 classes and each class consists of 3 shapes. Thus, we trained twenty four ($8 \times 3 = 24$) other codes so

Table 4

The classes of the 6 possible shapes of gravity sources
in our research based on Table 3

The shape of the gravity source	The related classes
Anticline	Elongated shape, Up curved, Down flat, Non Circle section
Horizontal cylinder	Elongated shape, Up curved, Down curved, Circle section
Rectangular prism	Limited shape, Up flat, Down flat, Non Circle section
Sphere	Limited shape, Up curved, Down curved, Circle section
Syncline	Elongated shape, Up flat, Down curved, Non Circle section
Vertical cylinder	Limited shape, Up flat, Down flat, Circle section

that each of them can estimate the shape out of 3 shapes. For more explanations see the next section.

Finally, it should be noted that, in this paper, the SVC has not only been used for estimating the approximate shape of the gravity sources, but we also modified our approach by training different SVC codes. And justification of this approach is that these different SVC codes can test the result of each other.

The flowchart of the above-mentioned 6 steps is as follows:

(1) 300 synthetic gravity profiles (training set) were created by gravity sources with definite shapes (6 shapes), dimensions, depths, and density contrasts for training SVC.

(2) We extracted 10 features from each of the 300 synthetic gravity profiles.

(3) Regarding our FS procedures:

(a) We selected the more proper 2, 3, and 4 features out of the 10 features based on IICD algorithm.

(b) We trained SVC codes (using different proper subsets) and tested them by the training set. We selected the best trained SVC codes (in which we use the best subset of features).

(4) We tested the best trained SVC codes with 60 more synthetic gravity profiles different from the training set. The results were relatively good but for better classification we modify our approach in the next step.

(5) We trained SVC codes for estimating the 6 shapes in 4 classification groups of 2 three shapes classes (in which we used the best subset of features). And the results of testing were better.

(6) For classifying with greater certainty we trained 24 more SVC codes (in which we used the best subset of features). Each of these 24 codes can estimate the shape between 3 shapes.

4. RESULTS AND DISCUSSIONS

4.1 Synthetic data (and some other discussions)

The results of testing with the training set showed us that using 2-feature subsets for our classification is not suitable. Examining the trained SVC codes related to 17 proper 3-feature subsets (see point 3 in Section 3) with the training set showed that only one of them is able to classify more than 0.70 of the objects in the training set correctly and we save this subset for more testing. Finally, testing the trained SVC codes related to the 14 proper 4-feature subsets with the training set showed us that they are able (with their best subset) to classify only 0.60 of objects in the training set correctly. Therefore, for classification we returned to only one subset of 3 features that was able to classify more than 0.70 of the objects in the training set correctly. Before discussing the above-mentioned best or most suitable subset of 3 features, in the following we show some results of using 2-feature subsets.

In the best case of testing with the training set using 2-feature subsets, only 0.42 of 50 anticlines in the training set are classified correctly (as shown in Fig. 3).

The results illustrate that using only 2 features for our classification is not suitable. So we do not continue discussion about using only 2 features for our classification.

As mentioned in point 3 of Section 3, according to Eq. 15, the possible number of 3-features subsets is 120. But from these 120 subsets only 17 have relatively big performance measures (see Section 2.2). These 17 subsets are the more proper subsets of features and these are shown in Table 5.

As we mentioned earlier, the best subset of features for classification (estimation) of the gravity source shape is one of the 3-feature subsets. As it was mentioned in point 4 of Section 3, this subset includes (X_{g50}/X_{g80}) ,

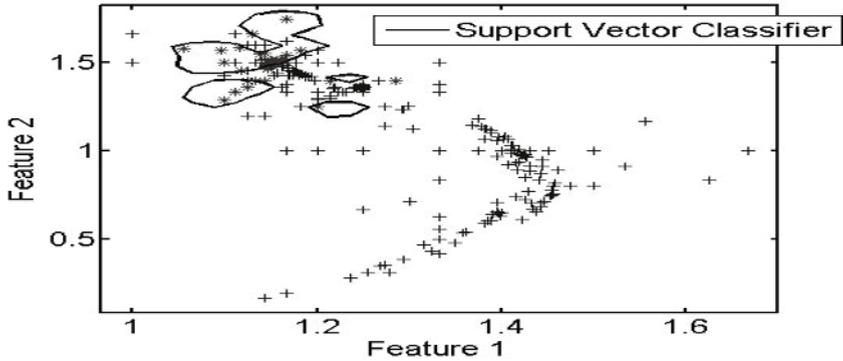


Fig. 3. The border of one of the trained SVC (related to the best 2-feature subset) and points showing the features of 300 training synthetic gravity profiles. This figure shows the trained SVC for detecting anticline (star symbols) and non-anticline (cross symbols) gravity source shape in 2-dimensional features space.

Table 5

17 subset of 3 features (between 120 possible subsets)
with relatively big performance measures (PM)

Feature 1 number	Feature 2 number	Feature 3 number	PM
5	6	10	3.397
3	6	10	3.394
4	6	10	3.371
1	6	10	3.358
2	6	10	3.180
1	3	10	3.169
1	5	10	3.156
1	4	10	3.150
3	5	10	3.148
1	3	6	3.143
3	4	10	3.141
4	5	10	3.123
1	5	6	3.106
1	4	6	3.098
3	4	6	3.036
3	5	6	3.033
1	2	10	3.008

Note: Regarding feature numbers, see Table 2.

$(X_{g\text{inf}}/X_{g80})$ and $((X_{g50} - X_{g80})/X_{g90})$, which are features 1, 3, and 10, regarding Table 2. In Table 6 we present the portion of each of the 6 gravity sources shapes which classified correctly using the best subset of features and using the related trained SVC codes.

Using the best 3-feature subset, 0.74 of 50 horizontal cylinders in the training set are classified correctly and are presented in Fig. 4.

As stated earlier about non-linear SVC, the kernel which used is RBF kernel (Eq. 5). The value of σ in the kernel is selected to be equal to 0.05.

As mentioned in point 5 of Section 3, for better classification, the six shapes are classified by four classification groups and each of these 4 classification groups consists of two classes, each class consisting of three shapes.

Table 6

The portion of each of the 6 gravity source shapes (in the training and testing set) which classified correctly using the most suitable subset of features and using the related trained codes

The shape of gravity source	The portion of the shape (in the training set) which classified correctly	The portion of the shape (in the testing set) which classified correctly
Anticline	$(36/50) = 0.72$	$(5/10) = 0.5$
Horizontal cylinder	$(37/50) = 0.74$	$(5/10) = 0.5$
Rectangular prism	$(50/50) = 1$	$(3/10) = 0.3$
Sphere	$(49/50) = 0.98$	$(10/10) = 1.0$
Syncline	$(48/50) = 0.96$	$(5/10) = 0.5$
Vertical cylinder	$(49/50) = 0.98$	$(6/10) = 0.6$

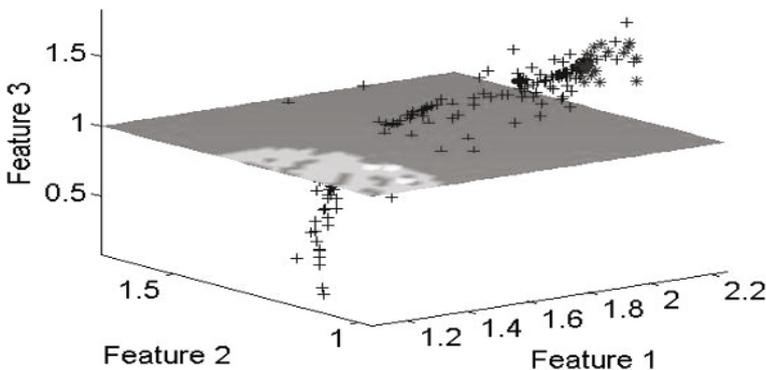


Fig. 4. One of the trained SVC (related to the best 3-feature subset) and points showing the features of 300 training synthetic gravity profiles. This figure shows the trained SVC for detecting horizontal cylinder (star symbol) and non-horizontal cylinder (cross symbol) gravity source shape in 3-dimensional features space.

These four groups and related information are shown in Table 3. For each of the 4 above-mentioned classification groups we used a couple of codes. The results of testing of one of these 4 couples of codes by training set and testing set are shown in Table 7.

Table 7

The portion of each of the 6 gravity sources shapes (in the training and testing set) which classified correctly using the most suitable subset of features and using one of the 4 couples of codes

The shape of gravity source	The portion of the shape (in the training set) which classified correctly	The portion of the shape (in the testing set) which classified correctly
Anticline	$(49/50) = 0.98$	$(9/10) = 0.9$
Horizontal cylinder	$(50/50) = 1$	$(10/10) = 1.0$
Rectangular prism	$(50/50) = 1$	$(7/10) = 0.7$
Sphere	$(49/50) = 0.98$	$(10/10) = 1.0$
Syncline	$(50/50) = 1$	$(5/10) = 0.5$
Vertical cylinder	$(50/50) = 1$	$(9/10) = 0.9$

The results of testing 3 others of the 4 couples of codes are suitable and almost all of the shapes were classified correctly.

As stated before, for each of the discussed 4 classification groups we have 2 classes. So all together we have eight classes, each of them consisting of 3 shapes. Therefore, for the six gravity sources shapes, eight groups of the three shapes (all together 24 shapes) exist. In this regard, we produced and trained 24 SVC codes (8 groups of codes, each consisting of 3 codes) so that each of these codes can separate one shape out of the three shapes. In Table 8 these eight groups of three shapes are represented.

Table 8

Names of the eight groups of the 3 shapes for which the related 24 codes can classify (estimate) the shape of the gravity source between 3 possible shapes using the most suitable subset of features

Names of the eight of the 3 shapes
Anticline, Horizontal cylinder, Syncline
Rectangular Prism, Sphere, Vertical cylinder
Anticline, Horizontal cylinder, Sphere
Rectangular Prism, Syncline, Vertical cylinder
Horizontal cylinder, Sphere, Syncline
Anticline, Rectangular prism, Vertical cylinder
Horizontal cylinder, Sphere, Vertical cylinder
Anticline, Rectangular prism, Syncline

In the next section we have used some of these 24 SVC codes in the shape estimation.

It should be mentioned that we created some type of SVC codes so that each of them individually is able to estimate the approximate shape of gravity source. But if we test each type of these codes with the others and if some or all of these types of SVC codes give us the same result (the estimate of the approximate shape of the gravity source is the same), we will be more assured about the estimated approximate shape of the gravity source than when we only use one type of our SVC codes for shape estimation. So logically the modified approach we introduced (using not only one but some or all of our different types of SVC codes for shape estimation) leads us to more liable results. In any way, the results of classifications of 2 types of our trained SVC codes are presented in Tables 6 and 7. So they can be compared.

Finally, we should say that, as we noted in the Introduction, there is no usual method and no unique result for estimation of initial assumption of shape of gravity source (approximate shape of gravity sources) and practically this assumption is prepared by non gravity data. So any trial in this regard is useful and therefore using SVC (especially with more proper features) will be a step towards better estimation of initial assumption of shapes of gravity sources.

4.2 Real data

In this Section we have used two sets of real data for testing our approach. The approximate location of these 2 sets of real data in Iran is shown in Fig. 5.

The first set of real data belongs to the small grid gravity network at the Institute of Geophysics of the University of Tehran. The aim was to detect and model the tunnels of an old buried channel. The grid space of $2 \times 2 \text{ m}^2$ is used for measurements. In both, first (Fig. 6) and second (Fig. 7) sets of real data, a Scintrex (CG3M) gravimeter with 1 micro-gal resolution is used and the coordinates of the points are measured by the Total Station (Leika 750) with an accuracy to a few centimetres in the x , y , and h coordinates. And also in both sets of real data, after gravity corrections due to the standard formulas, the Bouguer gravity anomalies are computed and, by removing the trend effect with polynomial fitting method, the residual anomalies are computed. The first set of real data is shown in Fig. 6.

A gravity profile (line AB in Fig. 6) is selected for testing. This profile was collected above a well, *i.e.*, the gravity source of this gravity profile is a well filled with air. We knew that the well is very similar in shape to a vertical cylinder or a long vertical rectangular prism.

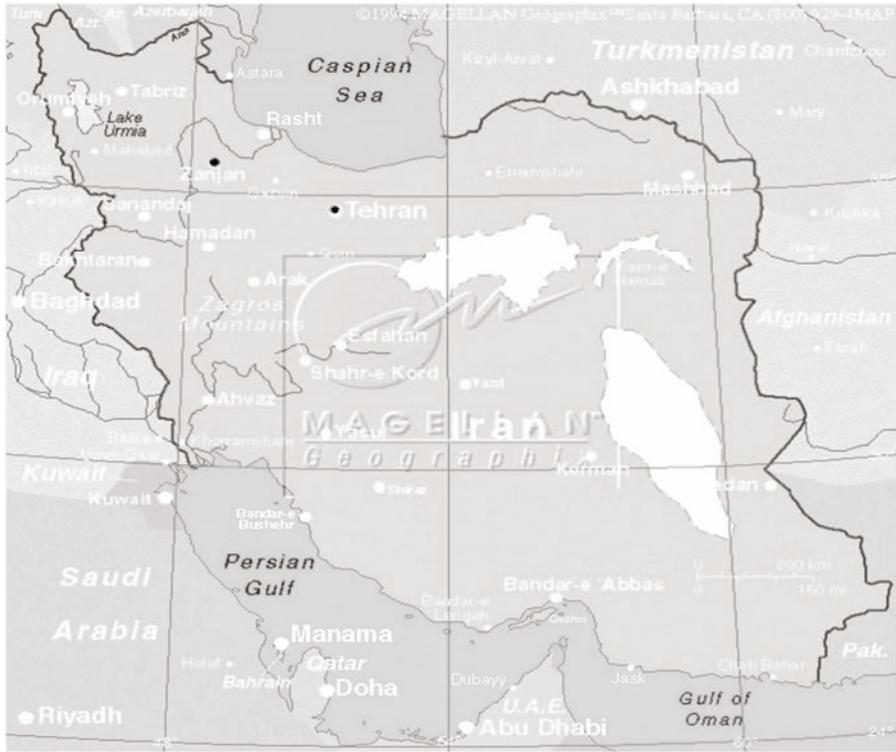


Fig. 5. Dark spots show the approximate locations of the 2 sets of real data in Iran. The geographic map of Iran is extracted from Ardestani (2008).

The second set of real data belongs to the grid gravity network carried out in a mining site close to Zanzan (a city in Iran) for searching for a manganese ore body. The grid space of $10 \times 10 \text{ m}^2$ is used. This set of real data is shown in Fig. 7.

The three real gravity profiles which we used in this set of real data are presented in Fig. 7 as profiles CD, EF, and GH. The 3D inversion according to the approach of Camacho *et al.* (2002) was applied on window 2 (this window is the big rectangle that is around not only the 3 above-mentioned profiles but also other areas in Fig. 6) of the second set of gravity data and this inversion is shown in Fig. 8.

In the above-mentioned gravity inversion, the aim has been to detect density differences of the sources with background and to detect approximate locations and dimensions of the sources.

The results of executing the trained SVC codes related to the 4 classification groups on the above-mentioned real gravity profiles are shown in Table 9.

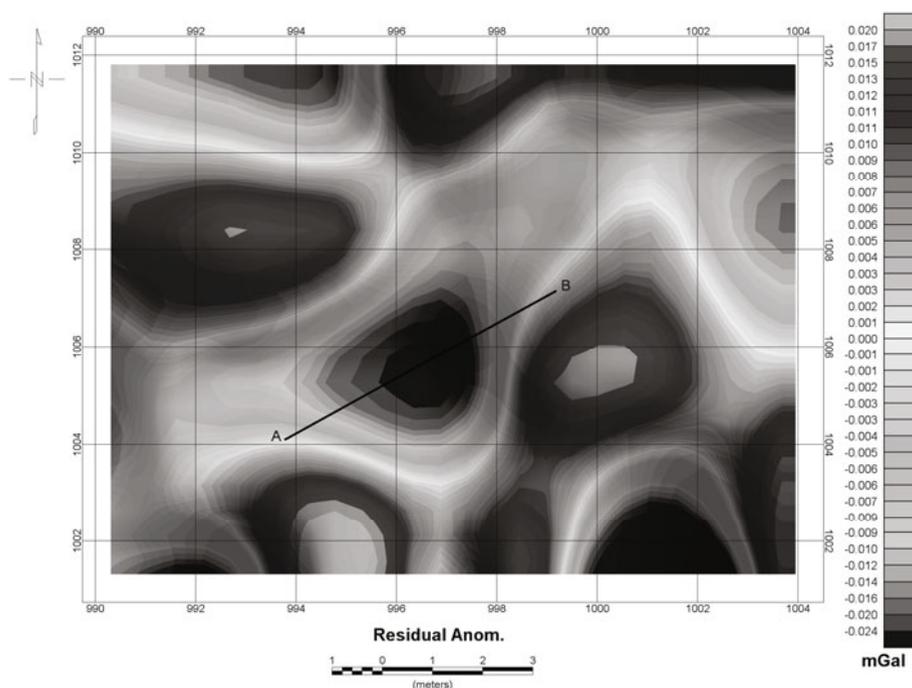


Fig. 6. Location of the gravity profile AB over the well. The map of this figure is a residual gravity map of a part of the Institute of Geophysics of University of Tehran and the coordinates are the local one (Ardestani 2009).

Regarding the calculation of probability of being of one specific class (for example, the result of executing one of the SVC codes on profile EF is that its source has Limited shape (a specific class) with probability 0.77 (as shown in Table 9)), we should refer to Fig. 1 for more explanations. In Figure 1, all the objects in the area above the line showing the upper limit of the margin are related to the class with the label $C_n = 1$ with probability 1 (100%), but if we have (in Fig. 1) any object between the line showing the upper limit of the margin and the line showing the middle of the margin, then this object would be related to the class with the label $C_n = 1$ with a probability between 100 and 50%. Also, if we have any object between the line showing the middle of the margin and the line showing the lower limit of the margin, then this object would be related to the class with the label $C_n = 1$ with a probability between 50 and 0%. And in areas lower than the line showing the lower limit of the margin, any object is related to the class with the label $C_n = 1$ with a probability equal to 0%. For all the objects, it should be mentioned that they are related to the class with the label $C_n = -1$ with probability equal to “1 – probability of being related to the class with

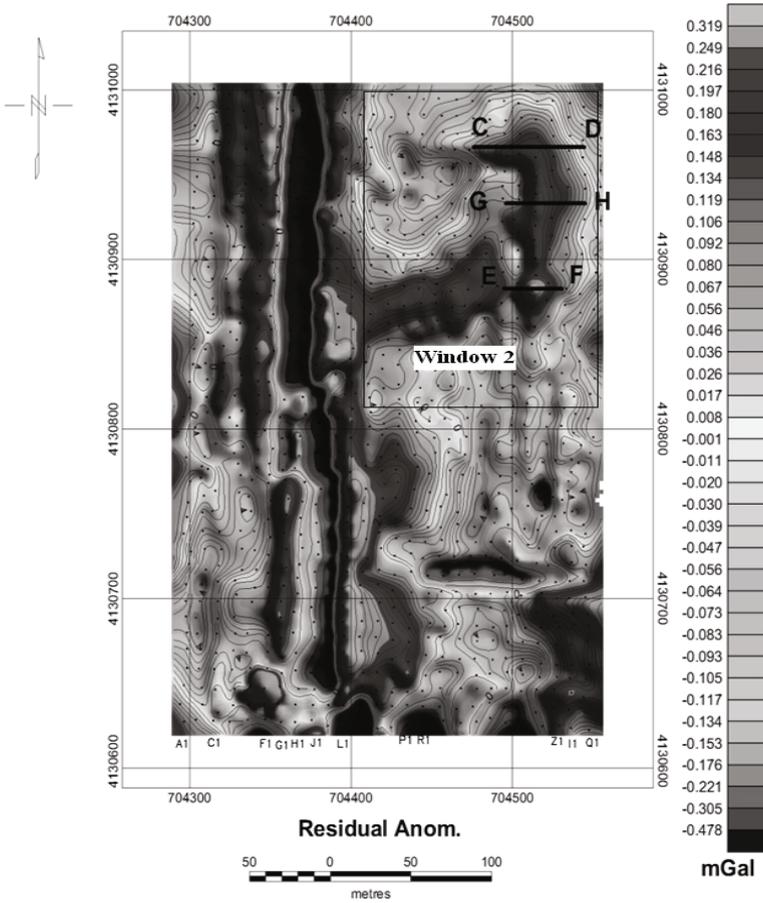


Fig. 7. Locations of the gravity profiles CD, EF, and GH. The map of this figure is a residual gravity map of a mining site near Zanjaan (a city in Iran) and the coordinates are UTM.

the label $C_n = 1''$. Also it should be mentioned that there is no rule regarding suitability for a definite value of probability to specify that, by having this value of probability or greater, we will reach the assurance for relating an object to one specific class. But subjectively the probability equal to 0.7 (70%) or more is good for relating an object to a specific class.

As seen in Table 9, the results of executing the codes on profile CD are not compatible with any of the 6 possible shapes of gravity sources (see Table 4). So the profile CD is not interpretable by SVC. Considering the location of CD profile in Fig. 7, we conclude that it is not a principal profile and it is not suitable for interpreting by SVC through our approach (see Sec-

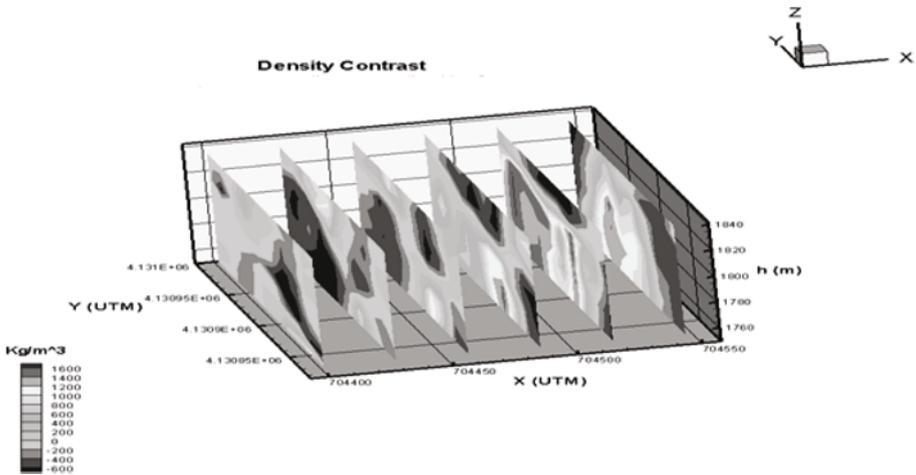


Fig. 8. The applied 3D inversion (based on the method of Camacho *et al.* 2002) on window 2 of Fig. 7. In the above figure, h is the height above sea level.

tion 2.2). Here a question may arise why the probability of being related to any specific class for CD profile is high (see Table 9). The answer is that the classes of CD profile are estimated only according to the values of its features and these values are only meaningful if they are extracted from a principal profile.

The results of executing the codes on 3 other profiles (AB, EF, and GH) show that these profiles are probably related to rectangular prism shapes (see

Table 9

The results of executing the trained SVC codes related to the 4 classification groups (see point 5 of Section 3) on the real gravity profiles

The results of executing the codes on profile			
AB	CD	EF	GH
Limited shape with probability 0.77	Elongated shape with probability 0.99	Limited shape with probability 0.77	Limited shape with probability 0.77
Up flat shape with probability 0.85	Up flat shape with probability 0.99	Up flat shape with probability 0.85	Up flat shape with probability 0.85
Down flat shape with probability 0.56	Down curved shape with probability 0.84	Down flat shape with probability 0.57	Down flat shape with probability 0.57
Non Circle section shape with probability 0.58	Circle section shape with probability 0.99	Non Circle section shape with probability 0.58	Non Circle section shape with probability 0.58

Table 4). Regarding AB profile, we state that a vertical cylinder is very similar to vertical long rectangular prism and also the shape of the well inside the ground may in some places more resemble a vertical long rectangular prism than a vertical cylinder. Secondly, the results in Table 9 show that the shape of the source of AB profile has a non-circle section shape with probability 0.58, thus AB profile has a circle-section shape with probability equal to $1 - 0.58 = 0.42$ and the probability of having a vertical cylinder source shape is about equal to having a rectangular prism source shape. The conditions of profile EF are almost the same as profile AB. To eliminate the ambiguity we tested these two profiles with 2 codes related to Table 8. According to our tests on 24 codes of Table 8, two codes of the related 24 codes are the best ones for estimating the rectangular prism and the vertical cylinder. By using these 2 codes, the results are as follows:

□ By executing the one which is the most suitable for estimating the rectangular prism on profiles AB and EF, the result was that with probability 0.18 the shapes of gravity sources of both profiles, AB and EF, are rectangular prisms.

□ By executing the one which is the most suitable for estimating the vertical cylinder on profiles AB and EF, the result was that with probability 0.75 the shapes of gravity sources of profiles AB and EF are vertical cylinders.

□ The results of executing the codes on profile GH show that this profile is related to rectangular prism gravity source shape with probability 0.7.

The elongated shape of the gravity anomaly of GH profile (see Fig. 7) confirms that the shape of the gravity source cannot be a vertical cylinder. Finally, the estimated shapes of the gravity sources of the above-mentioned real gravity profiles by our different trained SVC codes are given in Table 10.

Table 10

Shapes of gravity sources of the real gravity profiles
and the results of their testing by our different trained SVC codes

Name of the gravity profile	Shape of gravity source	Estimated shape of the source by the trained SVC codes
AB	Similar to vertical cylinder	Vertical cylinder
CD	Possibly rectangular prism	Not interpretable because of non being principal profile
EF	Possibly vertical cylinder or rectangular prism	Vertical cylinder
GH	Possibly rectangular prism	Rectangular prism

We see that the estimation of the shapes of the real gravity sources with SVC is compatible with the reality.

5. CONCLUSION

The SVC was used for estimation of the gravity sources shapes. Also, the results were improved by optimizing SVC by selecting proper features.

Also, by increasing the dimensions of features space from 2 to 3 and 4, we showed that among 375 subsets of features, where 210 subsets consisted of 4 features, only one subset, with 3 features, is the best. This means that increasing the dimensions of features space is sometimes not proper.

In this research, another important step was taken is developing several softwares that can test the results of each other. As you have seen in Table 6 (using the related trained codes) the highest portion of the corrected classified shapes in the testing set is related to the spherical shapes and the lowest portion is related to the rectangular prism shapes. Also, as you have seen in Table 7 (using the related 4 couples of codes) the highest portion of the corrected classified shapes in the testing set is related to both the horizontal cylinder and spherical shapes and the lowest portion is related to the syncline shapes. The related codes of Table 6 can be used independent of the related codes of Table 7 and *vice versa*. However, if we use all the above-mentioned codes together and if all the results confirm each other, the ill-posing of the gravity inversion will be treated relatively and results that are compatible among all the previously mentioned codes can be noticed approximately as the unique response. Therefore, we suggest that this approach can be applied for geoscientific investigations as a new approach for the estimation of the approximate shape of the gravity sources in exploration projects. We also suggest that this approach can be used in other branches of geophysics.

Acknowledgements. We thank all the faculty members who assisted us in preparing this paper. Our special thanks to Dr. Hashemi, for introducing us to the Pattern Recognition Laboratory, Delft University of Technology (whose group produced and designed PRTools which we used in this research). Also by studying his papers (Hashemi 2010, Hashemi *et al.* 2008), we decide to perform this research. We thank Prof. Duin (Pattern Recognition Laboratory, Delft University of Technology), for his helpful guidance at the start of this research. We thank the geophysics group of the Faculty of Basic Sciences of Science and Research Branch, Islamic Azad University, for guiding us so well during this research. Also we thank the Research Council of the University of Tehran.

References

- Ardestani, V.E. (2008), Modelling the karst zones in a dam site through micro-gravity data, *Explor. Geophys.* **39**, 4, 204-209, DOI: 10.1071/EG08022.
- Ardestani, V.E. (2009), Residual gravity map of a part of Institute of Geophysics of University of Tehran, University of Tehran, Tehran, Iran.
- Baan, M., van der, and Ch. Jutten (2000), Neural networks in geophysical applications, *Geophysics* **65**, 4, 1032-1047, DOI: 10.1190/1.1444797.
- Belikov, M.V. (1978), Approximating of the external potentials of bodies rotation, M.Sc. Thesis on Physical-Mathematical Science, Institute of Theoretical Astronomy of the USSR Academy of Science, Leningrad State University (translation from Russian).
- Blakely, J.R. (1996), *Potential Theory in Gravity and Magnetic Applications*, Cambridge University Press, Cambridge, 441 pp.
- Camacho, A.G., F.G. Montesinos, and R. Vieira (2002), A 3-D gravity inversion tool based on exploration of model possibilities, *Comput. Geosci.* **28**, 2, 191-204, DOI: 10.1016/S0098-3004(01)00039-5.
- Duin, R.P.W., P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D.M.J. Tax, and S. Verzakov (2007), PRTools4. 1 – A Matlab toolbox for pattern recognition, Delft University of Technology, Delft, Netherlands, <http://www.prtools.org/>.
- Gret, A.A., and E.E. Klingele (1998), *Application of Artificial Neural Networks for Gravity Interpretation in Two Dimensions*, Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, Zürich, Switzerland.
- Hashemi, H. (2010), Logical considerations in applying pattern recognition techniques on seismic data: Precise ruling, realistic solutions, *CSEG Recorder* **35**, 4, 47-50.
- Hashemi, H., D.M.J. Tax, R.P.W. Duin, A. Javaherian, and P. de Groot (2008), Gas chimney detection based on improving the performance of combined multi-layer perceptron and support vector classifier, *Nonlin. Process. Geophys.* **15**, 6, 863-871, DOI: 10.5194/npg-15-863-2008.
- Heijden, F., van der, R.P.W. Duin, D. de Ridder, and D.M.J. Tax (2004), *Classification, Parameter Estimation and State Estimation, An Engineering Approach using Matlab*, John Wiley & Sons Ltd, Chichester.
- Hekmatian, M.E., V.E. Ardestani, M.A. Riahi, A.M. Koucheh Bagh, and J. Amini (2013), Fault depth estimation using support vector classifier and features selection, *Appl. Geophys.* **10**, 1, 88-96, DOI: 10.1007/s11770-013-0371-7.
- Osman, O., A.M. Albora, and O.N. Ucan (2006), A new approach for residual gravity anomaly profile interpretations: Forced Neural Network (FNN), *Ann. Geophys.* **49**, 6, 1201-1208.

- Plouff, D. (1976), Gravity and magnetic fields of polygonal prisms and application to magnetic terrain corrections, *Geophysics* **41**, 4, 727-741, DOI: 10.1190/1.1440645.
- Talwani, M., J.L. Worzel, and M. Landisman (1959), Rapid gravity computations for two-dimensional bodies with application to the Mendocino submarine fracture zone, *J. Geophys. Res.* **64**, 1, 49-59, DOI: 10.1029/JZ064i001p00049.
- Telford, W.M., L.P. Geldart, R.E. Sheriff, and D.A. Keys (1976), *Applied Geophysics*, Cambridge University Press, Cambridge.

Received 25 March 2013

Received in revised form 24 June 2014

Accepted 1 July 2014



Reef Reservoir Identification by Wavelet Decomposition and Reconstruction: A Case Study from Yuanba Gas Field in China

Bingjie CHENG^{1,2,3}, Tianji XU⁴, Benedict ROBBINS⁵,
and ZhongMing SHEN¹

¹State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation,
Chengdu University of Technology, Chengdu, China;
e-mail: chengbingjie09@cudt.cn

²Key Laboratory of Earth Exploration and Information Technology of
Ministry of Education, Chengdu University of Technology, Chengdu, China

³Meteorological Information and Signal Processing,
Key Laboratory of Sichuan Higher Education Institutes,
Chengdu University of Information Technology, Chengdu, China

⁴3rdGeophysical Institute, Exploration and Production Institute,
Southwest Oil and Gas Company, SINOPEC, Chengdu, China

⁵Fugro Aperio, Fugro Onshore Geotechnics, Wallingford, Oxfordshire, UK

Abstract

The organic reef is a special type of carbonate reservoir which always dominates the spatial distribution, reserves and accumulations of natural gas. However, it is difficult to determine the organic reef's internal structure and gas reservoirs due to numerous adverse factors such as the low resolution of seismic data, depth of burial, strong anisotropy, irregular spatial distribution and complex internal structure. A case study of wavelet decomposition and reconstruction technology applied to elucidate the features of organic reef reservoirs in the Changxing formation from Yuanba gas field shows that the seismic record reconstructed by high frequency signal can adequately describe the internal properties of

organic reef reservoirs. Furthermore, the root mean square amplitude ratio of both low and high frequency data obtained from the reconstructed seismic data clearly show spatial distribution of gas and water in reef reservoirs.

Key words: seismic signal, organic reef, wavelet decomposition and reconstruction, hydrocarbon detection, reef reservoir identification.

1. INTRODUCTION

Marine carbonate reservoirs are widely distributed in China. However, these carbonate reservoirs are deeply buried, and there is strong vertical and horizontal heterogeneity, which contribute to increased difficulty in hydrocarbon detection. Reef gas reservoirs have been an important domain for gas exploration since the Yuanba gas field was discovered in the Sichuan basin; the Yuanba gas field shows huge potential as a prospective carbonate reef area. The growth scale and distribution of reef reservoirs are mainly controlled by carbonate sedimentary facies in the Changxing formation in Yuanba gas field (Ma *et al.* 2007, Bi *et al.* 2007). The identification of reef reservoirs is challenging, and some key problems remain to be solved, such as distinguishing water from gas, identifying the thickness of slim /fine layers, and improving accuracy in reservoir recognition.

In the reef reservoirs of Changxing formation in Yuanba gas field, the reef body is generally composed of reef base, reef core, and reef cap (Cai 2011). Organic reef is a special type of carbonate rock formation which has unique characteristics in terms seismic response. The seismic records show dome-shaped reflections from the outside reef, strong reflections or draping structure reflections from the reef top, discontinuous chaotic or nearly blank reflections from the inner reef, low elevation reflections from the reef bottom, and onlap reflections from the wings of the reef (Xiong *et al.* 2009, Hu *et al.* 2010). All of these characters aid in the identification of reef shape, and accurate location of the reef reservoir. However, it is difficult to clearly describe the detailed characteristics of the internal reef reservoirs due to some adverse factors, such as irregular spatial distribution, deep burial, low resolution seismic responses, strong heterogeneity, and anisotropy, complex internal structure and small physical differences between country rock and reservoirs.

For hydrocarbon detection, many geophysicists are optimistic about the use of the time-frequent seismic signal, which probes the response characteristics in both time and frequency domains (Castagna *et al.* 2003, Cohen 1995). However, most of their search has focused on the frequency, energy, amplitude, and attenuation properties ascertained through time-frequency analysis to obtain the low-frequency shallows and high-frequency attenuation of the gas reservoirs rather than to reconstruct seismograms in different

frequencies (Burnett *et al.* 2003, Reine *et al.* 2009, Xu *et al.* 2011, Xue *et al.* 2013).

The technique of wavelet decomposition and reconstruction can adaptively decompose a seismic trace into a series of wavelets of different frequency, amplitude, and energy. Since these wavelets' attributes are closely related to geologic features and the properties of reservoirs, the method then reconstructs the atoms in terms of the information associated with the gas reservoirs, with certain targeted objectives in mind, and a new seismic trace can be reconstructed with a selected subset of wavelets, which is found to be highly informative for seismic data interpretation and reservoir characterization (An 2006).

2. METHODS AND NUMERICAL MODELING

2.1 The theory of seismic wavelet decomposition and reconstruction

Presently, many mathematic methods (such as Fourier transform, wavelet transform, S-transform, and Matching Pursuit) have been proposed to carry out the seismic wavelet decomposition and reconstruction techniques (Partyka *et al.* 1999, Battista *et al.* 2007, Zhang 2008, Sinha *et al.* 2009, Wang 2010, Zhang and Lu 2010, Lu and Li 2013). In recent years, the method of Matching Pursuit was developed and widely applied in atom decomposition theory.

Matching Pursuit adaptively decomposes a seismic trace into a series of constituent wavelets (Mallat and Zhang 1993, Wang 2007). Each of these wavelets, selected from a dictionary consisting of a large number of wavelets, also called atoms, has an optimal correlation coefficient with the trace.

When considering the digital signal in atomic decomposition theory, the seismic trace can be decomposed to a series of wavelets with different dynamic characters, and the seismic trace can be reconstructed by simple linear super imposition of these wavelets. Xu *et al.*'s (2010) research outlines that if the original seismic trace $f(t)$ is decomposed N times by the wavelet $M(t)$, then:

$$f(t) = \sum_{n=0}^{N-1} A_n M_n(t) + E^{(N)} f, \quad (1)$$

where A_n is the amplitude of the n th wavelet $M_n(t)$, $E^{(N)} f$ is residual error, when $n = 0$ (n th wavelet), $E^{(N)} f = f$.

The result of the seismic trace decomposition and reconstruction is highly correlated to wavelet choice; thus, it is very important to choose suitable wavelets. In this paper, we employ Morlet wavelets as atoms in wavelet decomposition and reconstruction because it is appropriate for seismic waves with energy attenuation and velocity dispersion (Morlet *et al.* 1982a, b). The Morlet wavelet has reliable time-frequency resolution and can match the

signal to be analyzed furthest (Goupillaud *et al.* 1984, Gao *et al.* 1996); therefore, it has been widely used for wavelet transform in seismic data analysis (Grubb and Walden 1997, Farge 1992). Liu and Marfurt (2005) proposed a Matching Pursuit decomposition using Morlet wavelets. In this paper, we use a similar process, but develop a decomposition and reconstruction method based on Morlet wavelets.

Morlet *et al.*'s (1982a, b) research established Eqs. 2 and 3 for time domain and frequency domain wavelets and are still being currently used (Liu and Marfurt 2005):

$$M(t) = \exp(-t^2 \omega_m^2 \cdot \ln 2/k) \cdot \exp(i2\pi \omega_m t), \quad (2)$$

$$M(\omega) = \frac{\sqrt{\pi/\ln 2}}{\omega_m} \cdot \exp\left(-k \cdot \frac{\pi^2 (\omega - \omega_m)^2}{\ln 2 \cdot \omega_m^2}\right), \quad (3)$$

where ω is the frequency, ω_m is the average frequency or dominant frequency, k is a constant which acts on the Gaussian part of the wavelet function to modulate the wave shape to obtain the best seismic wavelet.

In frequency or time domains, using Eqs. 2 or 3 to iteratively calculate Eq. 1, we can decompose the seismic trace $f(t)$ into a wavelet function library formed by several different Morlet wavelet functions. By using the linear combination of different Morlet wavelet functions in the wavelet library, the reconstructed seismic trace $f'(t)$ is:

$$f'(t) = \sum_{m=0}^{M-1} A_m M_m(t), \quad M \leq N, \quad (4)$$

where N is the number of different morlet wavelet functions of original seismic trace $f(t)$, M is the number of different Morlet wavelet functions of reconstructed seismic trace $f'(t)$. M is less than N and the closer the M matches N , the closer the reconstructed seismic trace $f'(t)$ matches the original seismic trace $f(t)$. When $M = N$, $f'(t) = f(t)$.

During the iteration for wavelet decomposition, when the iterating times are enough to make the residual error $E^{(N)}f = 0$, the reconstructed seismic trace $f'(t)$ of all wavelet functions in the wavelet library equals the original seismic trace $f(t)$. If we ignore the residual errors produced in the process of Morlet wavelet decomposition and the linear summation of all Morlet wavelet functions, then $f'(t) \approx f(t)$. However, we need also to consider the computational efficiency during real seismic trace decomposition; thus, it can always meet the requirement that $E^{(N)}f$ is small enough. Notably, in generalized cases, associated processing time and economic value are very important factors in terms of practical requirements. When $E^{(N)}f$ is small enough, the method of wavelet decomposition will cease.

2.2 Numerical modeling

To illustrate the effectiveness of seismic wavelet decomposition and reconstruction methods for hydrocarbon detection, we produce a model to simulate the seismic response based on Hudson's theory (Hudson 1980, 1981, 1988; Hudson *et al.* 1996). The reservoirs in Yuanba gas field are pore-fractured reservoirs which are characterized by extremely low porosity, low permeability, and high-angle fractures. Primarily, there are vertical fractures distributing in the target interval (Hu *et al.* 2010, Chen 2011). The simplest effective model of a formation containing vertical fractures is horizontal transverse isotropy (HTI). For models with penny-shaped cracks, anisotropic coefficients, and weaknesses are expressed by the micro-structural parameters using the theories of Hudson (Hudson 1981, 1988; Hudson *et al.* 1996) and Thomsen (Thomsen 1986, 1995). Here, we use Hudson's model to calculate the elastic stiffness of anisotropic model, and use the expressions of fracture anisotropic parameters derived from the HTI model (Bakulin *et al.* 2000) to calculate the anisotropic parameters, which can be expressed as:

$$\varepsilon = -\frac{8e}{3}, \quad \delta = -\frac{8e}{3} \left[1 + \frac{g(1-2g)}{(3-2g)(1-g)} \right], \quad \gamma = -\frac{8e}{3(3-2g)}, \quad (5)$$

where $e = Na^3/V$ is defined as the fracture density, and V denotes the volume, a is the fracture radius, $g = V_s^2/V_p^2$ is the square of S -wave to P -wave velocity ratio. The relationship between the fracture density e and the aspect ratio r can be obtained by volume percent $\zeta = 4\pi re/3$.

The dimensionless parameters are normal ΔN and tangential ΔT weaknesses (Bakulin and Molotkov 1998), which can be calculated by the following equation:

$$\Delta N = \frac{4e}{3g(1-g)}, \quad \Delta T = \frac{16e}{3(3-2g)}. \quad (6)$$

The design of the model is based on the seismic data and reservoir logging parameters of the Yuanba Gas Field in Sichuan Basin, Southwest China. The geological models include six formations, where layers ⑦ and ⑧ are the gas-saturated carbonate and water-saturated carbonate, respectively (Fig. 1). The parameters of each layer are shown in Tables 1 and 2. Table 3 presents the fracture and anisotropy parameters calculated with the given fracture intensity. The formation thicknesses are 100 m. We use a point resource to obtain the synthetic seismogram, whose dominant frequency is 25 Hz without any added noise.

Figure 2a shows the forward modeling seismogram of the fractured model in Fig. 1. The synthetic seismogram is obtained by the full-wave forward modeling of finite difference. The event (about 0.3 s) is noticeably

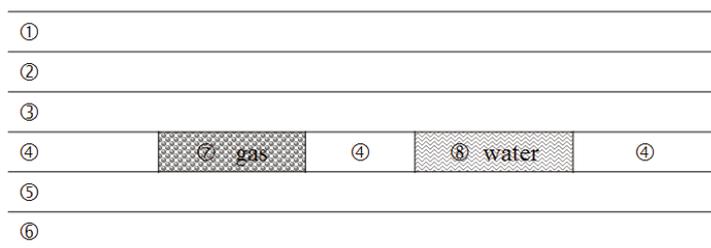


Fig. 1. A gas-saturated and water-saturated carbonate geological model.

Table 1
Rock properties for the geological model

Layer number	V_p [m·s ⁻¹]	V_s [m·s ⁻¹]	ρ [g·cm ⁻³]
①	3945	2650	2.48
②	4300	2487	2.50
③	4540	2949	2.65
④	4804	3081	2.65
⑤	5340	2949	2.75
⑥	5750	3358	2.71

Explanations: V_p – P -wave velocity, V_s – S -wave velocity, ρ – density.

Table 2
Rock properties for the fracture layer

Layer number	Stratum materials	V_p [m·s ⁻¹]	V_s [m·s ⁻¹]	ρ [g·cm ⁻³]	Φ [%]
⑦⑧	Carbonate rock	4.517	2.937	2.57	6.0
⑦	Mixed gas	0.7		0.00089	
⑧	Water	1.2		1.0	

Explanation: Φ is the porosity.

Table 3
The fracture and anisotropic properties of the gassy and hydrous layers

Layer number	Fracture parameters					Anisotropic parameters		
	E [m ⁻³]	r	ΔT	ΔN	θ [°]	ε	δ	γ
⑦⑧	0.02	0.01	0.05	0.11	30	-0.05	-0.05	-0.02

Explanation: θ is the obliquity.

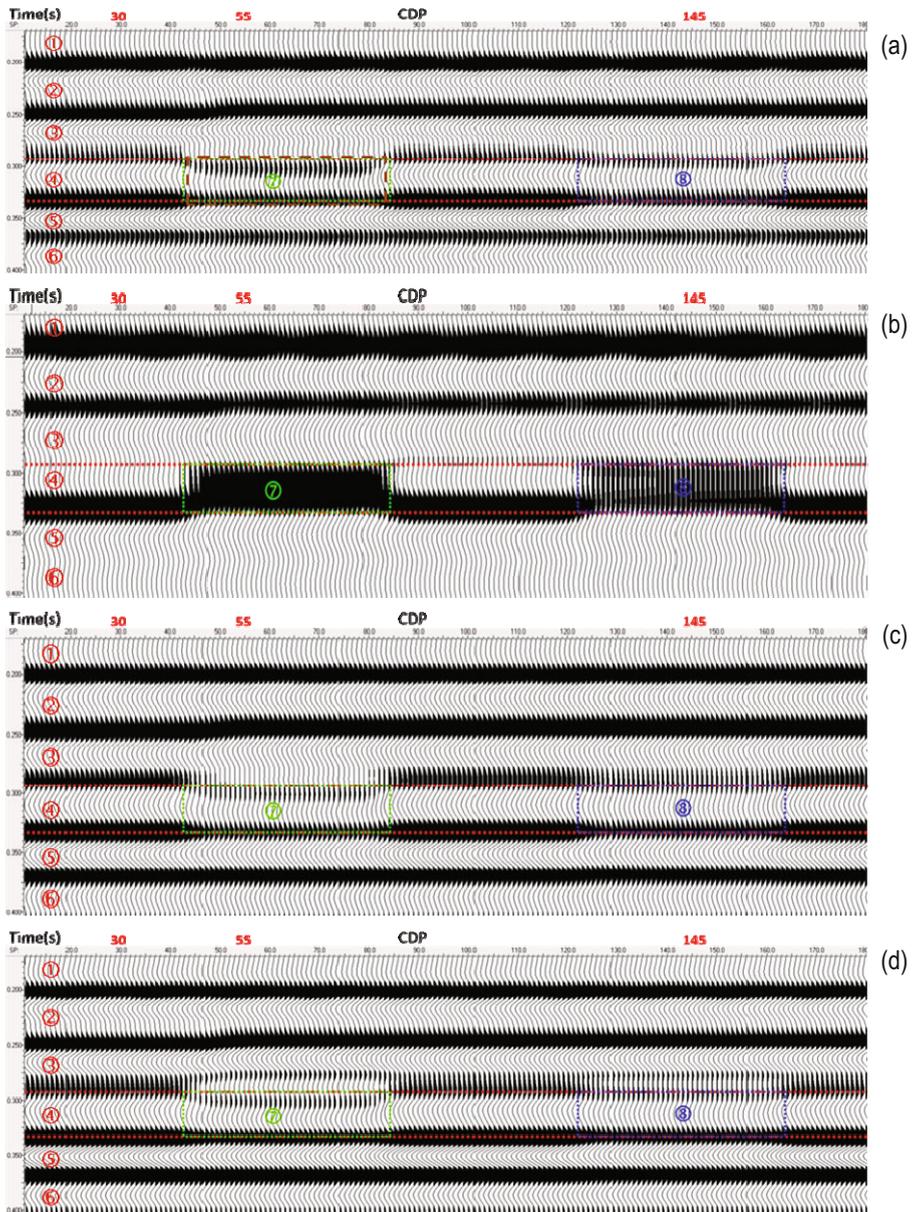


Fig. 2. Forward modeling and reconstructed seismic signal: (a) zero-offset synthetic seismogram of the geological model, (b) signal reconstruction at low frequency (from 5 to 10 Hz), (c) signal reconstruction at middle frequency (from 20 to 30 Hz), and (d) signal reconstruction at high frequency (from 40 to 50 Hz).

pulled down when the seismic waves propagate through the gas-saturated and water-saturated layers. Because of the physical property differences between the dry and gas or fluid saturated carbonate, the seismic reflections are different too. Layers ④, ⑦, and ⑧ have the same top and bottom boundaries (red dotted line); however, their seismograms are not able to display the same reflecting characteristics. Notably, the time delay for the gassy carbonate (green rectangular box) is much larger than the water filled carbonate (blue rectangular box). Furthermore, the seismic reflections of their top boundaries are not very distinct.

Figure 2b shows the 5-10 Hz reconstructed seismogram. Obviously, at low frequency, the top seismic reflection of the dry carbonate (④) layer is the weakest among ④, ⑦, and ⑧ layers, but the bottom seismic reflection is inverted. When comparing seismic events, the former layer ⑦ has more energy than the latter layer ⑧. The boundaries of gassy and hydrous carbonates are very clear, in the rectangular box where strong energy groups are observed.

Figure 2c shows the 20-30 Hz reconstructed seismogram. At a middle frequency, the gas or fluid saturated carbonates contain weaker seismic reflection events than the other layers. In particular, the reflection energy of gassy carbonate is weakest among them.

Figure 2d shows the 40-50 Hz reconstructed seismogram. It is well known that while the impedances change from high to low, the polarity of seismic events will be negative. As shown in Fig. 2d, at high frequency, the top troughs of layers ⑦ and ⑧ display the top boundaries clearly.

For a prime example with clear contrast, from Fig. 2a, we select the 30th, 55th, and 145th traces to decompose and reconstruct in the different frequency zone. The selected three traces are related with the seismogram of the dry, gas-saturated and fluid-saturated carbonates respectively. Figure 3a-c shows that nearly all of the events have increased strength correlating to increased frequency with the exception of the original. Figure 3b shows that the top seismic reflections of the gassy carbonate become increasingly clear with increased frequency, while Fig. 3c shows this feature of the hydrous carbonate to be inferior.

Therefore, from the model testing results, we can draw the following conclusions:

- the low-frequency reconstructed seismogram highlights the reflecting energy of the reservoirs,
- the reflections of the gas-saturated and water-saturated carbonates are not strong in the middle-frequency reconstructed seismogram,
- the high-frequency reconstructed seismogram outlines the layer boundaries.

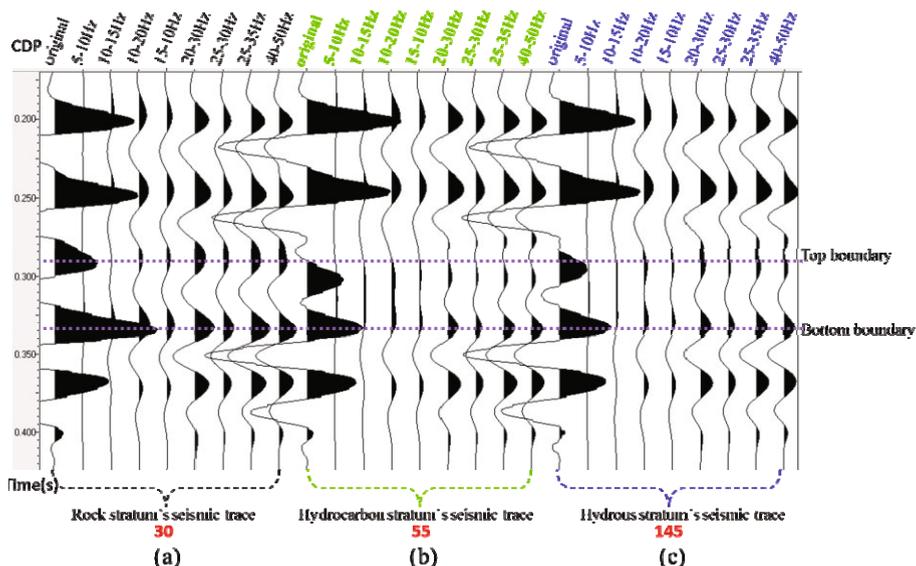


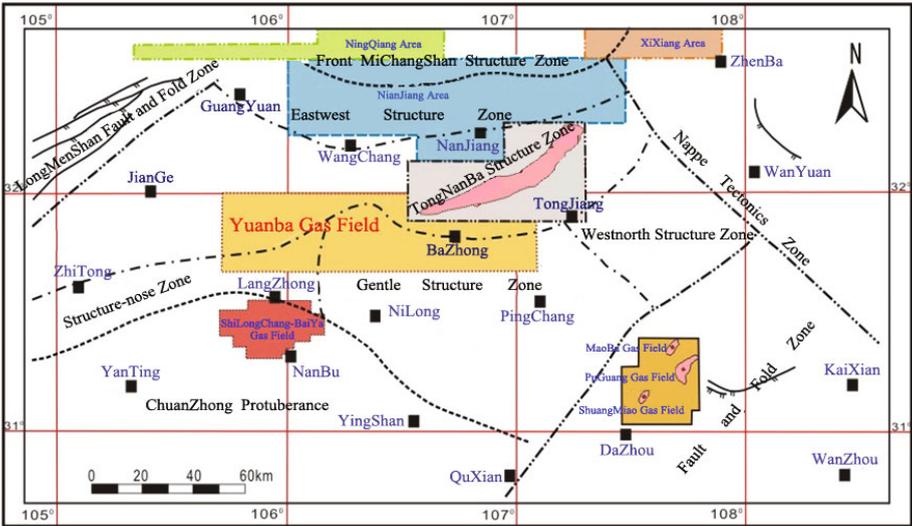
Fig. 3. Frequency-dependent seismic responses of the geological model: (a) reconstructed seismogram of the unfilled carbonate, (b) frequency-dependent seismogram of the carbonate filled with gas, and (c) frequency-dependent response of the carbonate filled with water.

3. APPLICATION IN GAS RECOGNITION OF REEF RESERVOIR IN CHANGXING FORMATION OF YUANBA GAS FIELD

3.1 Sedimentary facies features

Yuanba gas field is located in Sichuan Province in China, and is by far the most deeply buried marine gas field in China. Figure 4a shows the geographic location of Yuanba gas field. Changxing formation in Yuanba gas field is a representative formation combining reef and beach sedimentation. There are four sedimentary facies zones from northeast to southwest, namely continental shelf, slope, platform margin and open platform, respectively (Fig. 4b). In the early stage of Changxing's formation, the southeastern sections of the reservoir consisted of open bench terrace sediment, the northeast of the reservoir was slope and land shed sediment, and the landform was generally flat. The thin biodetritus beach and sandstone beach developed at parts of the open bench terrace. However, the reservoir was not productive and is under deep water. In these zones, there are only scattered high energy biodetritus beach reservoirs in locally high areas. In the medium term of Changxing's formation, the southwest portions gradually developed to a bench terrace edge, and produced thick, high energy biodetritus beach sediments which were widely distributed. In the later period of Changxing's

(a)



(b)

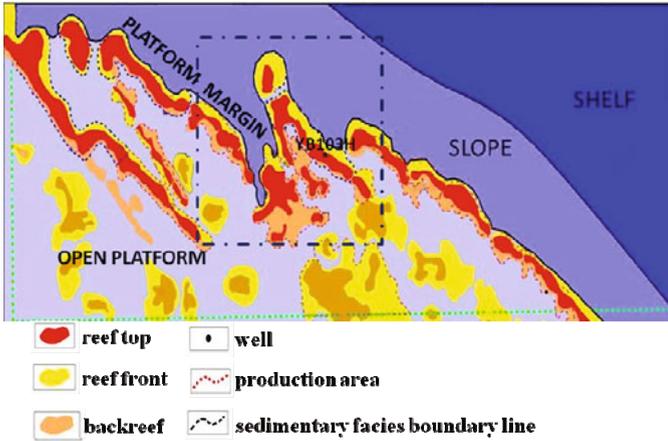


Fig. 4. The geographic location and the sedimentary facies of Yuanba gas field: (a) the geographic location of Yuanba gas field, and (b) the sedimentary facies of Yuanba gas field.

formation, the sediment topography changed greatly, and the organic reef began to develop at the margin of the continental shelf (Wang *et al.* 2012). The organic reefs in Changxing formation of Yuanba gas field mainly developed outside the bench terrace edge, and are distributed in a series of non-connecting bands. With the reef rock filling constantly into the back, a shallow bench will be developed and mixed gradually in it. Therefore, the organic reef and the shallow bench sections offer the greatest potential reservoir zones above the Changxing formation.

3.2 Reservoir characteristics

In Yuanba gas field, the Changxing formation consists of dolomite and limestone. Petrophysical properties of dolomite are more favorable for gas reservoirs than those of limestone. Autoclastic dolomite and aplite dolomite are important rock types for reservoir strata. The reef reservoir is formed of dolomitic residue and non-residue autoclastic solution pores, organic clastic limestone, and organic reef limestone. The beach reservoir mainly consists of gray solution pore dolomite and gray matter as well as residue autoclastic limestone dolomite. It also has non-dolomitic gray limestones formed of autoclastic, sandclastic, and conglomerate clastic. Notably, solution pore dolomite is the most likely rock type to contain the largest gas reserves.

Statistics of core testing show that the porosity of Changxing formation reef reservoir is 0.23 ~ 19.59%, with the average of 4.06%, most of the values range between 2 ~ 5% (Fig. 5a), the permeability is $0.0028 \sim 1720.7187 \times 10^{-3} \mu\text{m}^2$, the main peak values are $0.01 \sim 0.1 \times 10^{-3} \mu\text{m}^2$, 39% of samples' permeabilities are smaller than $0.1 \times 10^{-3} \mu\text{m}^2$, 44% of samples' permeabilities are $0.1 \sim 10 \times 10^{-3} \mu\text{m}^2$ (Fig. 5b). The porosities of most limestones are $< 5\%$, but most of the permeabilities are $> 1 \text{ mD}$, which indicates that the fractures are the most important gas-bearing spaces. Changxing formation generally belongs to pore reservoirs and fracture-pore reservoirs.

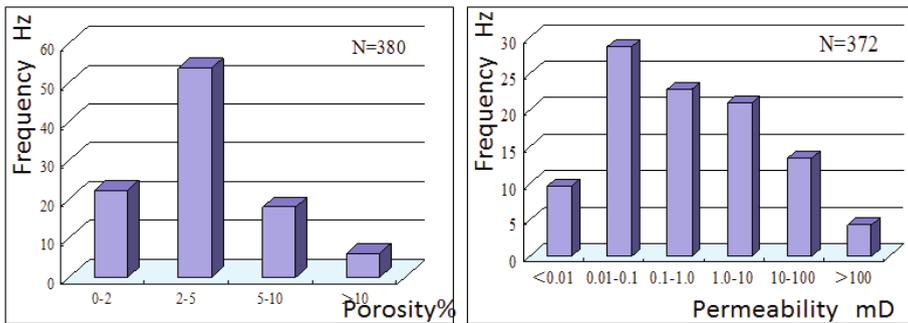


Fig. 5. The distributing histograms of organic reef reservoirs' physical property in Changxing formation of Yuanba gas field: (a) porosity, and (b) permeability. N is the number of samples.

3.3 Reef reservoir recognition

The reef reservoir is a special kind of carbonate rock formation which has unique seismic reflecting characteristics in Yuanba gas field. On the outline of the reef body, the seismic response shows dome-shaped reflections. On the top of the reef, the seismic records appear to contain a "draping structure" within strong reflections. Seismic records internal to the reef present

discontinuous, chaotic or nearly blank reflections. On the reef bottom, the seismic records appear with low elevation reflection response. On the flanks of the reef, some onlap reflections can be clearly identified. All of these characteristics help us to identify the shape and edge of the reef, and accurately locate the position of reef reservoirs. However, according to the original seismic data, it is difficult to deeply analyze the internal structures and growth periods of reefs because of the negative influences of discontinued, irregular, flat or blank reflections from the inner reef. So, we applied seismic wavelet decomposition and reconstruction technology to process the original seismic data. It is expected to determine the seismic frequency response laws of the organic reef and allow insight into the internal structure of the reef reservoir.

This paper shows a case study in part of Yuanba gas field which is shown in the blue rectangle in Fig. 4b. The study has focused on the well YB103H, in which open-flow capacity is $751.6 \times 104 \text{ m}^3/\text{d}$. From left to right, Fig. 6 shows the well logs, the PSTM seismic data, and the reconstructed seismic data. In Figure 6b, c, and d, from top to bottom, the blue curves T-cx-5x, T-cx-3x, and T-cx-down represent the bottoms of different strata in the Changxing formation. Above the blue curve T-cx-5x, the first wave peak corresponds to the reef's top. The blue curve T-cx-3x shows the bottom of reef. The blue curve T-cx-down represents the bottom of the Changxing formation. The red arrow indicates mudstone strata overlying the reef. In the PSTM seismic data, the reflection events from the reef's outline is clear, but the reflection events from the overlying strata and the reef's interior are irregular and complex which makes the prediction and recognition of reef reservoirs extremely difficult. After decomposing the original seismic data, we reconstructed new seismic data with different dominant frequencies.

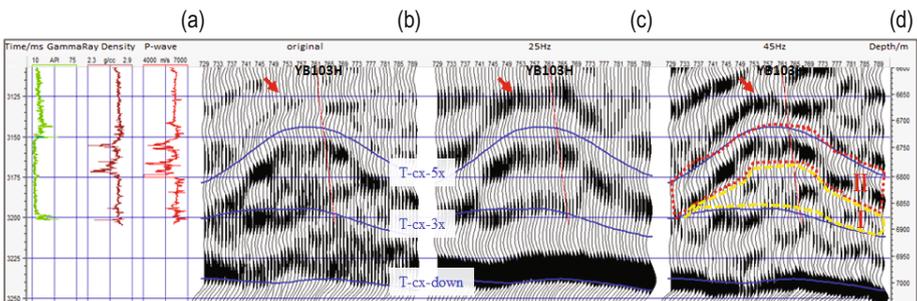


Fig. 6. Well seismic calibration and PSTM seismic data comparison with their reconstructed seismic data: (a) from left to right, well logs are natural gamma ray, density and acoustic logging curves respectively, (b) PSTM seismic data, (c) 25 Hz dominant frequency reconstructed seismic data, and (d) 45 Hz dominant frequency reconstructed seismic data.

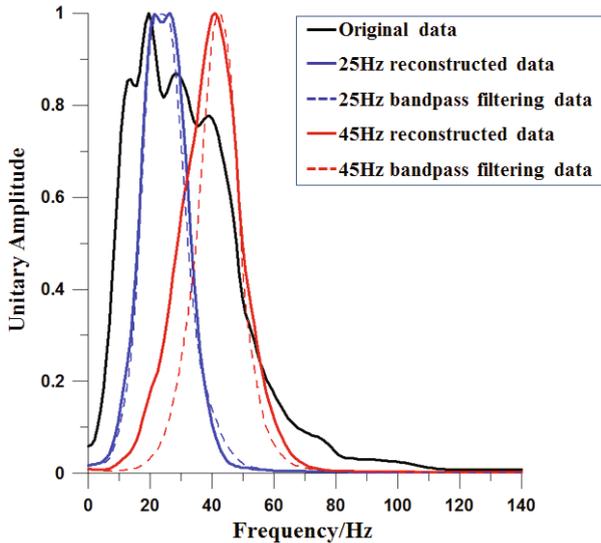


Fig. 7. The comparative amplitude spectrums of the original, bandpass filtering, and reconstructed seismic data.

The reconstructed seismic data with 25 Hz dominant frequency indicates that the mudstones covering and surrounding the reef are clearly identified. The reconstructed seismic data with 45 Hz dominant frequency, however, do not only clearly highlight the reflections from the reef's outline but also the internal evolution sequence. Figure 6d shows two evolution ages of the reef which are enclosed by red and yellow frames, respectively. The reef developed from base to core and cap in one age, and evolved from core to cap in the other. Seismic reflections alone are not able to reveal the reef's internal characteristics any further; however, the main evolution periods can be identified from them when combined with well logs. Therefore, by comparing the three seismic data sets, the reconstructed seismic data is more utilizable not only to predict the reef reservoir, but also to help us combine the well to seismic calibration information to improve the recognition precision of the reef body.

Figure 7 shows the amplitude spectra of the original PSTM seismic data (dark line), the reconstructed seismic data with 25 and 45 Hz dominant frequency (blue and red lines, respectively), and the band pass filtering seismic data with 25 and 45 Hz center frequency (blue and red dash lines, respectively). By comparing the amplitude spectra of the reconstructed seismic data with that of the band pass filtering with the same frequency, both of frequency amplitude spectra are similar. Thus, the residual error of reconstructed data is negligible.

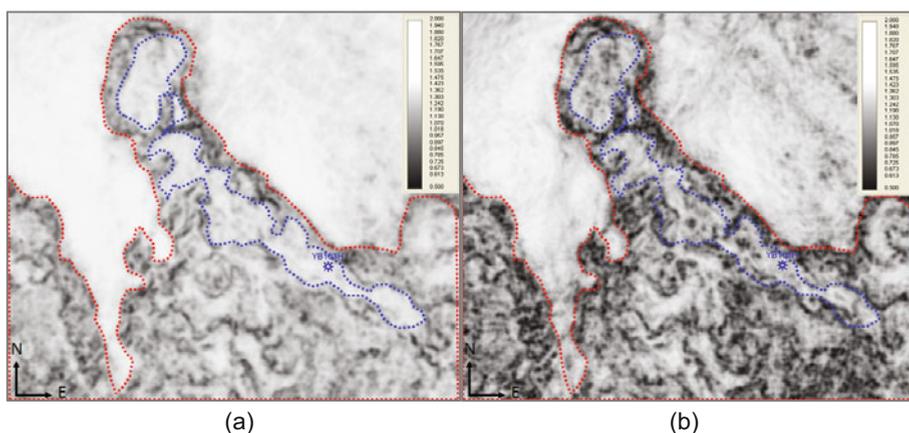


Fig. 8. Comparison of coherent attributes along T-cx-5x horizon between PSTM data and reconstructed PSTM data in Yuanba gas field: (a) coherence calculated by original seismic data, and (b) coherence calculated by 45 Hz dominant frequency reconstructed seismic data.

A seismic coherence attribute is usually used in fracture prediction. It is also useful for studying the reef-beach fracture-pore reservoirs in Changxing formation of Yuanba gas field. Figure 8 shows coherence attributes extracted from the original seismic data and the 45 Hz dominant frequency reconstructed seismic data on the T-cx-5x horizon in Yuanba gas field. Furthermore, the darker the color, the more probability of the developing fractures. On the T-cx-5x horizon in Changxing formation, the reservoir's spatial distribution can be highlighted by the coherence attributes which are circled by a red dotted line frame. However, when contrasting in Fig. 8 panels (a) with (b), it is evident that the latter has higher resolution than the former. In addition, in the area where the reef reservoir is circled by a blue dotted line frame, the abnormal coherence attribute of the latter is stronger than the former, and reveals the superior quality of the reef reservoir in which the coherence is poor and the probability of fracture growth higher.

In addition to coherence, the Root Mean Square (RMS) amplitude is another useful attribute to identify the reef. Figure 9 shows the RMS amplitudes of T-cx-5x horizon in Yuanba gas field. Figure 9a illustrates the RMS amplitudes calculated from the original seismic data and Fig. 9b shows the RMS amplitudes calculated from the 45 Hz dominant frequent reconstructed seismic data. Inside the blue dotted line frame, the red areas are considered to be the reefs. By comparing panels (a) with (b), it is easy to see that the latter describes the spatial distribution of reef reservoir more clearly than the former. In the layout chart of the T-cx-5x horizon, the well YB103H in Yuanba gas field lies in the reservoir zone extending from northwest to

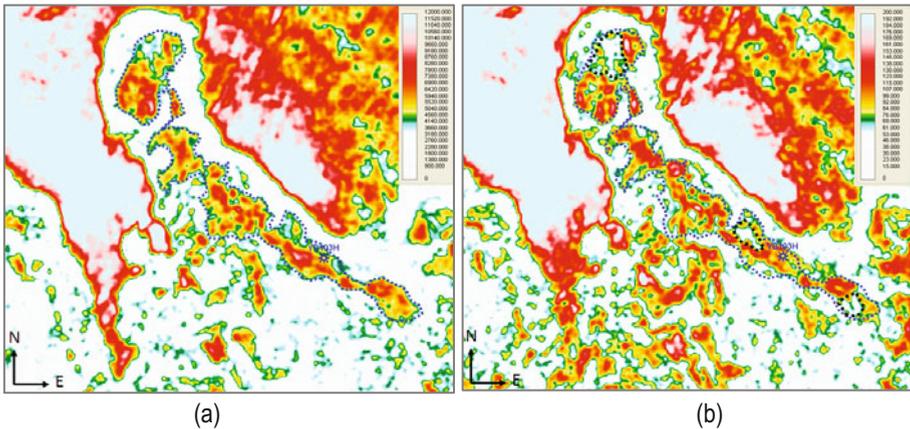


Fig. 9. Comparison of RMS amplitude attributes along the T-cx-5x horizon between PSTM seismic data and reconstructed seismic data in Yuanba gas field: (a) RMS amplitude calculated by original seismic data, and (b) RMS amplitude calculated by 45 Hz dominant frequency reconstructed seismic data.

southeast, having the same growth orientation as the reef zone. Although both, (a) and (b), can clearly describe the spatial distribution of a high quality reef reservoir, the latter can describe the inner boundary of the reef reservoir in more detail than the former, which is circled by the black dotted line frame.

Despite the reservoirs of Changxing formation in Yuanba gas field being very tight, the reef's porosity is relatively high, and its fractures are well developed, which is advantageous for hydrocarbon accumulation and migration. Notably, the reef reservoir has critical factors for spatial distribution of gas, reserves, and production. After comparing and analyzing the original and reconstructed seismic data, the results show that the reconstructed data can effectively highlight characteristics of seismic reflections within the reef reservoirs which allow in-depth analysis and understanding of their spatial distribution, internal structure, and evolution.

The RMS amplitude ratio of reconstructed data can also be applied to further understand the identified reef reservoirs. Figure 10 shows the RMS amplitude ratio of the 8 and 45 Hz reconstructed seismic data in Yuanba gas field, which can be used to distinguish between the strata consisting of gas-water or purely gas. In Figure 10, black shows the original seismic profile, and the red-yellow-green-white represents the RMS amplitude ratios from small to large. In the location of well YB103H, the reef reservoir top where fractures and pores grow well is mainly within the gassy strata (indicated by red arrows) which spread along the reef widely. The gas-water stratum is very narrow and lies in the bottom of reef cap showed in the white area be-

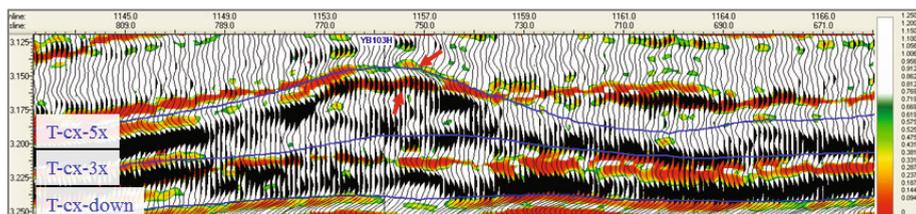


Fig. 10. Original seismic data and RMS amplitude ratio of 8 and 45 Hz reconstructed seismic data in Yuanba gas field.

tween red arrows. These results prove that the RMS amplitude ratio of reconstructed data is useful in solving the technical challenges of recognizing reef reservoirs' gas within the Changxing formation in Yuanba gas field.

To sum up, in Changxing formation, the gas prediction results based on the reconstructed data fit exactly with the logging results of well YB103H. Hence, the reconstructed data perform well for understanding the properties of the organic reef, and provide important evidence for highly productive reservoir prediction and gas recognition. Consequently, it is an extremely useful method in gas exploration and development.

4. CONCLUSION

In Changxing formation of Yuanba gas field, the applications of wavelet decomposition and reconstruction technology for reef reservoirs identification show that:

- the reconstructed seismic data of middle to high frequency (about 25 Hz) can highlight the mudstone overlying the reef, and exactly describe the reef's outline;
- the reconstructed seismic data of high frequency (about 45 Hz) can identify the reef's internal characteristics, such as the spatial distribution of the reef and *in situ* fracture development; these factors are paramount in order to recognize the reef's evolution history and predict the highly productive zones of the reservoir;
- by combining the low and high frequency reconstructed seismic data, we can effectively and efficiently determine spatial distribution of the gas reservoir.

In short, the wavelet decomposition and reconstruction technology can be successfully applied in reef reservoir prediction and gas recognition, proving that the technology is an extremely useful tool with promising potential applications to assist in future gas exploration and development.

Acknowledgments. This work was partially supported by NSFC Project No. 41204071 and No. U1262206, the National High Technology Research and Development Program of China (No. 2013AA064201), application program of science and Technology Department of Sichuan province, China (No. 2013JY0060), and Cultivating programme of excellent innovation team of Chengdu University of Technology.

References

- An, P. (2006), Application of multi-wavelet seismic trace decomposition and reconstruction to seismic data interpretation and reservoir characterization. **In:** *Proc. 76th SEG Annual Meeting, 1-6 October 2006, New Orleans, USA*, SEG-2006-0973.
- Bakulin, A.V., and L.A. Molotkov (1998), Effective seismic models of fractured and porous media, St. Petersburg University Press, St. Petersburg (in Russian).
- Bakulin, A., V. Grechka, and I. Tsvankin (2000), Estimation of fracture parameters from reflection seismic data—Part I: HTI model due to a single fracture set, *Geophysics* **65**, 6, 1788-1802, DOI: 10.1190/1.1444863.
- Battista, M.B., C. Knapp, T. McGee, and V. Goebel (2007), Application of the empirical mode decomposition and Hilbert–Huang transform to seismic reflection data, *Geophysics* **72**, 2, 29-37, DOI: 10.1190/1.2437700.
- Bi, C.C., L.X. Li, Y. Mei, Y.C. Zhang, and C. He (2007), Geological control factors of Changxing organic reef distribution and seismic prediction techniques, East Sichuan Basin, *Nat. Gas. Geosci.* **18**, 4, 2-7.
- Burnett, M.D., J.P. Castagna, E. Méndez-Hernández, G.Z. Rodríguez, L.F. García, J.T.M. Vázquez, M.T. Avilés, and R.V. Villaseñor (2003), Application of spectral decomposition to gas basins in Mexico, *The Leading Edge* **22**, 11, 1130-1134, DOI: 10.1190/1.1634918.
- Cai, X.Y. (2011), The subtly method of reservoir and exploration effects on the organic reef-beach body of Changxing Formation, Yuanba Area, Northeastern Sichuan, *Eng. Sci.* **13**, 28-33 (in Chinese with English abstract).
- Castagna, J.P., S.J. Sun, and R.W. Siegfried (2003), Instantaneous spectral analysis: Detection of low-frequency shadows associated with hydrocarbons, *The Leading Edge* **22**, 2, 120-127, DOI: 10.1190/1.1559038.
- Chen, Y. (2011), Prediction on reef reservoir in Changing group of Yuanba area, Northeast Sichuan Basin, *Geophys. Prospect. Petrol.* **50**, 2, 173-180 (in Chinese).
- Cohen, L. (1995), *Time-frequency Analysis*, Prentice Hall, Eaglewood Cliffs, 299 pp.
- Farge, M. (1992), Wavelet transforms and their applications to turbulence, *Ann. Rev. Fluid Mech.* **24**, 395-457, DOI: 10.1146/annurev.fl.24.010192.002143.

- Gao, J.H., W.B. Wang, and G.M. Zhu (1996), On the choice of wavelet functions for seismic data processing, *Acta Geophys. Sinica* **39**, 3, 392-400 (in Chinese).
- Goupillaud, P., A. Grossmann, and J. Morlet (1984), Cycle-octave and related transforms in seismic signal analysis, *Geoexploration* **23**, 1, 85-102, DOI: 10.1016/0016-7142(84)90025-5.
- Grubb, H.J., and A.T. Walden (1997), Characterizing seismic time series using the discrete wavelet transform, *Geophys. Prospect.* **45**, 2, 183-205, DOI: 10.1046/j.1365-2478.1997.00346.x.
- Hu, W.G., Y. Pu, Z.N. Zhao, and J.L. Xiao (2010), Identification of reef reservoir of Changxing formation in Yuanba area of Northeastern Sichuan Basin, *Geophys. Prospect. Petrol.* **49**, 1, 46-53 (in Chinese).
- Hudson, J.A. (1980), Overall properties of a cracked solid, *Math. Proc. Cambridge Philos. Soc.* **88**, 2, 371-384, DOI: 10.1017/S0305004100057674.
- Hudson, J.A. (1981), Wave speeds and attenuation of elastic waves in material containing cracks, *Geophys. J. Int.* **64**, 1, 133-150, DOI: 10.1111/j.1365-246X.1981.tb02662.x.
- Hudson, J.A. (1988), Seismic wave propagation through material containing partially saturated cracks, *Geophys. J. Int.* **92**, 1, 33-37, DOI: 10.1111/j.1365-246X.1988.tb01118.x.
- Hudson, J.A., E.Liu, and S. Crampin (1996), The mechanical properties of materials with interconnected cracks and pores, *Geophys. J. Int.* **124**, 1, 105-112, DOI: 10.1111/j.1365-246X.1996.tb06355.x.
- Liu, J.L., and K.J. Marfurt (2005), Matching pursuit decomposition using Morlet wavelets. **In:** *Proc. 75th SEG Annual Meeting, 6-11 November 2005, Houston, USA*, SEG-2005-0786.
- Lu, W.K., and F.Y. Li (2013), Seismic spectral decomposition using deconvolutive short-time Fourier transform spectrogram, *Geophysics* **78**, 2, 43-51, DOI: 10.1190/geo2012-0125.1.
- Ma, Y.S., C.L. Mou, Q.Y. Tan, Q. Yu, and R.H. Wang (2007), Reef-bank features and their constraint to reservoirs of natural gas, from Permian Changxing Formation to Triassic Feixianguan Formation in Daxian-Xuanhan Area of Sichuan Province, South China, *Earth Sci. Front.* **14**, 1, 182-192, DOI: 10.1016/S1872-5791(07)60007-4.
- Mallat, S.G., and Z. Zhang (1993), Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.* **41**, 12, 3397-3415, DOI: 10.1109/78.258082.
- Morlet, J., G. Arens, E. Fougereau, and D. Giard (1982a), Wave propagation and sampling theory – Part I: Complex signal and scattering in multilayered media, *Geophysics* **47**, 2, 203-221, DOI: 10.1190/1.1441328.
- Morlet, J., G. Arens, E. Fougereau, and D. Giard (1982b), Wave propagation and sampling theory – Part II: Sampling theory and complex waves, *Geophysics* **47**, 2, 222-236, DOI: 10.1190/1.1441329.

- Partyka, G., J. Gridley, and J. Lopez (1999), Interpretational applications of spectral decomposition in reservoir characterization, *The Leading Edge* **18**, 3, 353-360, DOI: 10.1190/1.1438295.
- Reine, C., M. van der Baan, and R. Clark (2009), The robustness of seismic attenuation measurements using fixed- and variable-window time-frequency transforms, *Geophysics* **74**, 2, 123-135, DOI: 10.1190/1.3043726.
- Sinha, S., P. Routh, and P. Anno (2009), Instantaneous spectral attributes using scales in continuous-wavelet transform, *Geophysics* **74**, 2, 137-142, DOI: 10.1190/1.3054145.
- Thomsen, L. (1986), Weak elastic anisotropy, *Geophysics* **51**, 10, 1954-1966, DOI: 10.1190/1.1442051.
- Thomsen, L. (1995), Elastic anisotropy due to aligned cracks in porous rock, *Geophys. Prospect.* **43**, 6, 805-829, DOI: 10.1111/j.1365-2478.1995.tb00282.x.
- Wang, Y.H. (2007), Seismic time-frequency spectral decomposition by matching pursuit, *Geophysics* **72**, 1, 13-20, DOI: 10.1190/1.2387109.
- Wang, Y.H. (2010), Multichannel matching pursuit for seismic trace decomposition, *Geophysics* **75**, 4, 61-66, DOI: 10.1190/1.3462015.
- Wang, Z.H., J. Deng, Q.Y. Tan, R.H. Wang., and J.X. Cheng (2012), Characteristics of typical sedimentary facies and their petrophysics of Changxing formation in Yuanba district, *J. Mineral. Petrol.* **32**, 2, 86-96.
- Xiong, X.J., Z.H. He, and D.J. Huang (2009), Numerical simulation on seismic response features of reef, *Acta Petrolei Sin.* **30**, 1, 75-79 (in Chinese).
- Xu, T.J., Z.M. Shen, and X.K. Wen (2010), Research and application of multi-wavelet decomposition and reconstructing technology, *J. Chengdu Univ. Technol. (Sci. Technol. Ed.)* **37**, 6, 660-665.
- Xu, T.J., B.J. Cheng, Z.P. Qian, J.M. Tang, Q.G. Gan, and Z.M. Shen (2011), Multi-scale frequency and absorption attributes of seismic signals, *J. Geophys. Eng.* **8**, 3, 457-463, DOI: 10.1088/1742-2132/8/3/006.
- Xue, Y.J., J.X. Cao, and R.F. Tian (2013), A comparative study on hydrocarbon detection using three EMD-based time-frequency analysis methods, *J. Appl. Geophys.* **89**, 108-115, DOI: 10.1016/j.jappgeo.2012.11.015.
- Zhang, Q., and W.K. Lu (2010), Spectral decomposition using deconvolutive short time Fourier transform spectrogram. **In:** *Proc. 80th SEG Annual Meeting, 17-22 October 2010, Denver, USA*, SEG-2010-1581.
- Zhang, R.F. (2008), Spectral decomposition of seismic data with CWPT, *The Leading Edge* **27**, 3, 326-329, DOI: 10.1190/1.2896622.

Received 8 October 2013

Received in revised form 6 August 2014

Accepted 21 October 2014

3D Gravity Inversion using Tikhonov Regularization

Reza TOUSHMALANI¹ and Hakim SAIBI²

¹Department of Computer, Faculty of Engineering, Kangavar Branch,
Islamic Azad University, Kangavar, Iran; e-mail: geoman110@gmail.com

²Laboratory of Exploration Geophysics, Department of Earth Resources Engineering,
Faculty of Engineering, Kyushu University, Fukuoka, Japan;
e-mail: saibi-hakim@mine.kyushu-u.ac.jp (corresponding author)

Abstract

Subsalt exploration for oil and gas is attractive in regions where 3D seismic depth-migration to recover the geometry of a salt base is difficult. Additional information to reduce the ambiguity in seismic images would be beneficial. Gravity data often serve these purposes in the petroleum industry. In this paper, the authors present an algorithm for a gravity inversion based on Tikhonov regularization and an automatically regularized solution process. They examined the 3D Euler deconvolution to extract the best anomaly source depth as *a priori* information to invert the gravity data and provided a synthetic example. Finally, they applied the gravity inversion to recently obtained gravity data from the Bandar Charak (Hormozgan, Iran) to identify its subsurface density structure. Their model showed the 3D shape of salt dome in this region.

Key words: gravity inversion, base salt, Tikhonov regularization, Euler deconvolution, Bandar Charak.

1. INTRODUCTION

Subsalt exploration for oil and gas is attractive in regions where 3D seismic depth-migration to identify the geometry of the salt base is difficult. The complexity of the seismic ray paths and the lack of sufficient seismic energy

penetrating the salt make identifying the salt base difficult because of the complex shape of the salt and the high impedance contrast with the surrounding sediments. This problem may lead to poor imaging and inaccurate interpretations of subsalt geological structures. Additionally, 3D seismic data processing, particularly 3D seismic depth migration, which must be iteratively performed, is costly.

Additional information reduces the ambiguity in seismic images and helps speed up the iterative migration process. Model building would also be beneficial. Gravity data serve these purposes in the petroleum industry (Cheng 2003). The gravity method was the first geophysical technique, and it is used in oil and gas exploration. Despite being eclipsed by seismology, it has continued to be an important and sometimes crucial constraint in a number of exploration fields. In oil exploration, the gravity method is particularly applicable in salt provinces, over thrust and foothills belts, underexplored basins, and targets of interest that underlie high-velocity zones. The gravity method is frequently used in mining to map subsurface geology and to calculate reserves for some massive sulfide ore-bodies. In addition, there is a modest increase in the use of gravity techniques in specialized investigations for shallow targets, and these techniques are also applied in archeology, hydrogeology, and geothermal studies (Saibi *et al.* 2008). Data reduction, filtering, and visualization, together with low-cost, powerful personal computers and color graphics, have transformed the interpretation of gravity data. Euler and Werner deconvolution depth and edge estimation techniques can help define the lateral location and depth of isolated faults and boundaries from gravity data. Complex geology with overlapping anomalies arising from different depths can limit the effectiveness of deconvolution fault-detection results (Nabighian *et al.* 2005, Touthmalani 2011).

Mapping the Earth's structure in two and three dimensions is being addressed through various inverse techniques. Many researchers prefer inverse methods to forward modeling methods because they offer quantitative solutions that can be determined more often than the trial-and-error approach of forward modelling (Touthmalani and Saibi 2014).

Researchers are working on the inverse gravimetric problems to seek an operator to act on observed gravity data and return a subsurface density distribution that can generate the observed field. Li and Oldenburg (1998) presented a 3D inversion of gravity data and Li (2012) discussed the recent advances in 3D generalized inversion of potential-field data.

1.1 All types of operators

Some researchers assume a known density contrast and design nonlinear operators to determine the geometry of the source. Wavenumber domain meth-

ods have proven effective to determine the depth to a particular density interface, despite the difficulties associated with wavenumber domain transformations. These types of methods are less effective in determining full 3D solutions with multiple sources and varying density contrasts.

Other researchers use an approach similar to the one presented in this research. This approach solves the unknown density distribution by applying a linear operator to the data. Linear space domain operators are designed by fixing the geometry of the sources, usually by dividing the source volume into a number of elementary shapes, and allowing the density of the cells to vary. The linear approach is effective to determine the nature of small-scale geological features when given reasonable starting models. Recent advancements in computational power have made it feasible to use for regional problems as well.

The previously discussed methods are based on combining linear, spatial, and wavenumber domain calculations. These methods are strongly dependent on the variance of the initial model and advise users to run the program with different variances and initial models before making any geologic interpretation (Bear *et al.* 1995).

In this paper, the authors present an inverse method to derive objective 3D density distributions given map gravity data. The researchers used an appropriate estimate for the error on the right side, which can be determined automatically by projecting the rows of the matrix onto the “usable” rows of an orthogonalized version of the system based on the regularized solutions for an ill-conditioned system (Jones 2006).

It is difficult to obtain high-resolution images using a 3D gravity inversion, because the problem is extremely underdetermined (*i.e.*, there are too many model parameters). To reduce the number of model parameters, a 3D gravity inversion scheme using Euler deconvolution as *a priori* information was proposed (Rim *et al.* 2007).

This paper is organized as follows: First, the Euler deconvolution method is described. Second, the “usable” rows of an orthogonalized version of the system based on regularized nonnegative solutions for an ill-conditioned system is discussed. Then, the techniques are applied to the solution of a general linear inverse problem. Finally, synthetic examples and a field data example are provided.

1.2 Euler deconvolution technique

Euler deconvolution is a technique that uses potential field derivatives to image the subsurface depth of a magnetic or gravity source (Thompson 1982, Hsu 2002). Mushayandebvu *et al.* (2001) described the 3D space Euler’s deconvolution equation as

$$(x - x_0) \frac{\partial g}{\partial x} + (y - y_0) \frac{\partial g}{\partial y} + (z - z_0) \frac{\partial g}{\partial z} = -N \Delta g, \quad (1)$$

where (x_0, y_0, z_0) is the coordinate position of the top of the body, z is the depth measured (positive is downward), x is the horizontal distance, Δg is the value of the residual field, and N is a structural index. The structural index is a measure of the rate of change or fall off rate with the distance of a field. Therefore, it is a function of the geometry of the causative bodies. The magnetic field of a point dipole falls off with the inverse cube with an index of three, while a vertical line source gives an inverse square field fall off and an index of two. Extended bodies form assemblages of dipoles and will have indices ranging from zero to three (El Dawi *et al.* 2004, Reid *et al.* 1990).

If Δg_i is the residual gravity field at a point in a gravity survey, with the point measured at (x, y, z) and the coordinate position of the top of the body (x_0, y_0, z_0) , then Eq. 1 can be written as

$$\left[\begin{array}{ccc} \frac{\partial \Delta g}{\partial x} & \frac{\partial \Delta g}{\partial y} & \frac{\partial \Delta g}{\partial z} \end{array} \right] \begin{bmatrix} (x - x_0) \\ (y - y_0) \\ (z - z_0) \end{bmatrix} = -N \Delta g_i. \quad (2)$$

By calculating the horizontal and vertical gradients of the field, Eq. 2 only has three unknowns x_0, y_0, z_0 , and N . The first three describe the location of the body. Many simultaneous equations can be obtained for various measurement locations, which yield one matrix equation (Hansen and Suci 2002).

$$\left[\begin{array}{ccc} \frac{\partial g^1}{\partial x} & \frac{\partial g^1}{\partial y} & \frac{\partial g^1}{\partial z} \\ \cdot & \cdot & \cdot \\ \frac{\partial g_i}{\partial x} & \frac{\partial g_i}{\partial y} & \frac{\partial g_i}{\partial z} \end{array} \right] \begin{bmatrix} (x - x_0) \\ (y - y_0) \\ (z - z_0) \end{bmatrix} = N \begin{bmatrix} \Delta g^1 \\ \cdot \\ \Delta g_i \end{bmatrix}. \quad (3)$$

The least squares method can be used to obtain the unknowns x_0, y_0 , and z_0 if the structural index N is known. The Euler deconvolution method is used to estimate the depth and form of an anomaly, gravity and magnetic resources, the value of two factors of the Structural Index (SI) and the width of moving window size (W size). These values are sometimes determined by the interpreter using previous experience or regional geology, because they are not always clear. In this section, despite previous works, two loops were included in codes written in Matlab. In the first loop, all Structural Index (SI) values of sets of 0-3, with an increase of 0.5, were included. In the second

loop, moving window size (W size) odd values of 3-19 with an increase of 2 were included. All possible depths were extracted. Analyzing histograms of all z values showed that accepted values are the most frequent regarding the depth of model (Toushmalani and Hemati 2013).

2. METHODOLOGY

The authors addressed the gravimetric problem by assuming that the Earth can be represented by a finite number of rectangular blocks. This parameterization allows us to calculate the gravitational attraction caused by each block. The sum of the contributions from each of the individual blocks produces the observed gravity field (Fig. 1).

The gravity observed at any station (+) can be calculated by summing the contributions from each block at any point in question (Bear *et al.* 1995). The vertical gravitational attraction of a homogeneous body of an arbitrary shape can be shown as

$$g_z(x, y, z) = \gamma \int_{a_1}^{a_2} \int_{b_1}^{b_2} \int_{c_1}^{c_2} \frac{\rho c d a d b d c}{(a^2 + b^2 + c^2)^{3/2}}, \tag{4}$$

where

$$a_1 = u_1^i - x_i, \quad a_2 = u_2^i - x_i, \quad b_1 = v_1^j - y_i, \quad b_2 = v_2^j - y_i, \quad c_1 = w_1^k - z_i, \quad c_2 = w_2^k - z_i,$$

$g_z(x, y, z)$ is the vertical gravitational attraction at (x, y, z) resulting from the homogeneous body, γ is the universal gravitational constant, and ρ is the density of an elementary mass within the body. The location of the mass is (u, v, w) . We evaluated Eq. 4 for a right rectangular prism. The integral can be expressed in elementary form as

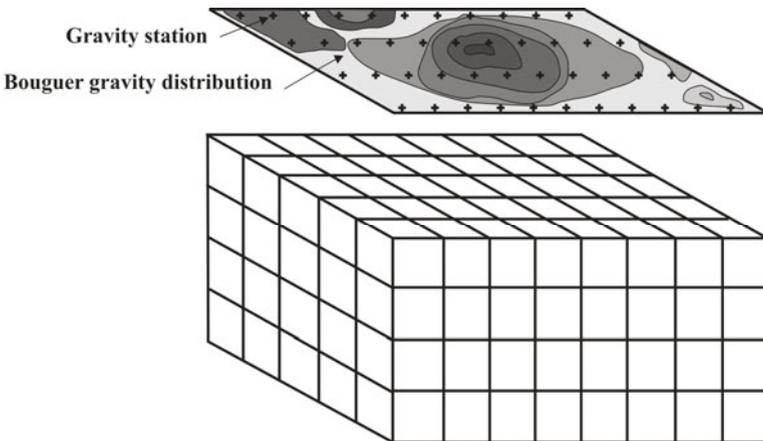


Fig. 1. Parameterization of the Earth model as rectangular blocks.

$$\begin{aligned}
g_z = & -\gamma \rho \left[x \ln \left(y + (x^2 + y^2 + z^2)^{\frac{1}{2}} \right) + y \ln \left(x + (x^2 + y^2 + z^2)^{\frac{1}{2}} \right) \right. \\
& \left. + z \arctan \times z \left((x^2 + y^2 + z^2)^{\frac{1}{2}} x^{-1} y^{-1} \right) \right] \\
& \left[u_2^j - x_i \right] \left[v_2^j - y_i \right] \left[w_2^j - z_i \right] \\
& \left[u_1^j - x_i \right] \left[v_1^j - y_i \right] \left[w_1^j - z_i \right]
\end{aligned} \tag{5}$$

and

$$\begin{aligned}
A_{ij} = & -\gamma \left[x \ln \left(y + (x^2 + y^2 + z^2)^{\frac{1}{2}} \right) + y \ln \left(x + (x^2 + y^2 + z^2)^{\frac{1}{2}} \right) \right. \\
& \left. + z \arctan \times z \left((x^2 + y^2 + z^2)^{\frac{1}{2}} x^{-1} y^{-1} \right) \right] \\
& \left[u_2^j - x_i \right] \left[v_2^j - y_i \right] \left[w_2^j - z_i \right] \\
& \left[u_1^j - x_i \right] \left[v_1^j - y_i \right] \left[w_1^j - z_i \right]
\end{aligned} \tag{6}$$

The edges of the j prism are parallel to the reference axes, and the limiting coordinates for its volume are u_1^j , u_2^j for the x coordinate, v_1^j , v_2^j for the y coordinate and w_1^j , w_2^j for the z coordinate (negative downward). The anomalous bodies responsible for observed gravity anomaly will be determined as a composition or accretion of prismatic cells with an assigned density contrast. Equation 5 is the product of the density of the block and a constant A_{ij} (the term in square brackets in Eq. 5), whose value is determined by the geometry of the block and its relation to the observation point. It is written as:

$$g_i = \sum_{j=1}^m A_{ij} P_j . \tag{7}$$

The difference between the theoretical values calculated \mathbf{g}_i^{th} from Eq. 5 and the observed gravity values $\mathbf{g}_i^{\text{obs}}$ at any observation point can be calculated. However, before this work:

□ The gravity values calculated for initial density structure at points ρ_i are:

$$g_i^o = \sum_{j=1}^m A_{ij} P_j^o . \tag{8}$$

□ In second step, a better model will be obtained by adding an additional density contrast $\Delta\rho_j$ (maybe positive value $\Delta\rho^+$ or maybe negative value $\Delta\rho^-$) to the initial values \mathbf{g}_i^o .

□ To obtain the final model, a linear regional trend is then adjusted

$$\delta g_i = g_{\text{obs}} - (g_i^o + A_{ij}\Delta\rho_j) - (p_o + p_x x_i + p_y y_i), \quad (9)$$

where p_o , p_x , and p_y are unknown simple regional trend (linear for instance) coefficients. In the next step,

$$\delta g_i = \sum_{j=1}^m A_{ij} \delta P_j. \quad (10)$$

The geometry of the sources remains fixed, so the geometric gravitational coefficients are constant (A_{ij}) or n map observations and m prisms in the earth model. A system of linear equations can be formed as:

$$\begin{bmatrix} \delta_{g1} \\ \delta_{gn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{1m} \\ a_{n1} & a_{nm} \end{bmatrix} * \begin{bmatrix} \delta_{p1} \\ \delta_{pm} \end{bmatrix}. \quad (11)$$

The differential density is the only unknown in Eq. 11. The goal of the inversion is to estimate the differential density for each block (prism). Some researchers assumed a known density contrast and designed nonlinear operators to determine the geometry of the source. The wave number domain methods have proven effective to determine the depth to a particular density interface, despite the difficulties associated with wave number domain transformations. These types of methods are less effective in determining full 3D solutions with multiple sources and varying density contrasts. Other researchers used an approach similar to the one that is presented in this paper. This approach solves the unknown density distribution by applying a linear operator to the data. Linear space domain operators are designed by fixing the geometry of the sources, usually by dividing the source volume into a number of elementary shapes and allowing the density of the cells to vary.

The linear approach was effective in determining the nature of small-scale geological features when given a reasonable starting model. Recent advances in computational power have also made it feasible to use on regional problems. Methods are typically based on combining linear, spatial, and wave number domain calculations. These methods are strongly dependent on the variance of the initial model. Users are strongly advised to run the program with different variances and initial models before making any geologic interpretation (Bear *et al.* 1995).

Equation 11 can be rewritten in matrix form as:

$$\mathbf{g} = \mathbf{A} \boldsymbol{\rho}, \quad (12)$$

where \mathbf{A} is an $n \times m$ matrix of geometrical coefficients, \mathbf{g} is an n -length differential gravity vector, and $\boldsymbol{\rho}$ is an m -length vector of unknown perturbations.

$$\begin{aligned}
E &= \mathbf{g}_{\text{obs}} - \mathbf{g}_{\text{cal}} = \mathbf{g}_{\text{obs}} - A\mathbf{p} \Rightarrow e^2 = E^T E \\
\Rightarrow e^2 &= [\mathbf{g}_{\text{obs}} - A\mathbf{p}]^T [\mathbf{g}_{\text{obs}} - A\mathbf{p}] \\
\Rightarrow e^2 &= \mathbf{g}_{\text{obs}}^T \mathbf{g}_{\text{obs}} - 2\mathbf{p}^T A^T \mathbf{g}_{\text{obs}} + \mathbf{p}^T A^T A \mathbf{p} \\
\text{minimization} &\Rightarrow \frac{\partial e}{\partial \mathbf{p}} = -2A^T \mathbf{g}_{\text{obs}} + 2A^T A \mathbf{p} \Rightarrow \Delta \mathbf{p} = (A^T A)^{-1} A \Delta \mathbf{g} \\
\Rightarrow e^2 &= E^T E + \lambda^2 \mathbf{p}^T \mathbf{p} \Rightarrow \Delta \mathbf{p} = [A^T A + \lambda I]^{-1} A^T E
\end{aligned} \tag{13}$$

When performing Tikhonov or similar types of regularization of ill-conditioned linear systems, a free parameter λ must be determined. Common techniques, such as using L-curves, are somewhat daunting for non-experts, and automated methods of choosing λ have not been widely applied. A given choice of λ implies a corresponding residual for the resulting regularized system. An appropriate residual is easily determined from the error level on the right side if that is known.

Software solutions available in various forms at www.rejonesconsulting.com solve this general linear problem:

$$\begin{aligned}
A\mathbf{x} &= \mathbf{b} \quad (\text{ordinary equations}), \\
E\mathbf{x} &= \mathbf{f} \quad (\text{equality constraints}), \\
G\mathbf{x} &\geq \mathbf{h} \quad (\text{inequality constraints}),
\end{aligned}$$

where A has m rows and n columns; $A\mathbf{x} = \mathbf{b}$ may be underdetermined, squared or over-determined (that is, $m < n$, $m = n$, or $m > n$); $A\mathbf{x} = \mathbf{b}$ may be ill-conditioned and/or singular.

The SVD of A is $A = USV^T$. In the absence of the small elements in S , the natural solution is $\mathbf{x} = VS^{-1}U^T\mathbf{b}$.

Applying the equality and inequality constraints are necessary because of the NNLS algorithm in Lawson, which is well known and will not be presented here.

The interesting part of the solution process is determining a regularization parameter to apply to the ordinary equations when they are ill-conditioned and no good error estimates are available for \mathbf{b}_i . We obtained the parameter λ to use when we add the following equations to the system $A\mathbf{x} = \mathbf{b}$:

$$\lambda I\mathbf{x} = 0,$$

where I is an identity matrix of size n by n .

Adding these equations is typically referred to as Tikhonov regularization. Determining a value for λ automatically (or even manually) is difficult. We found that this can be done by examining the Picard Condition Vector, or \mathbf{p} , where $V^T\mathbf{x} = S^+U^T\mathbf{b} = \mathbf{p}$ (S^+ is the same as S^{-1} , except zero is used when S_i is zero).

The Discrete Picard Condition says that for a good solution, one should expect the coefficients in \mathbf{p} to decline toward zero. In ill-conditioned problems, the Discrete Picard Condition is violated. We developed heuristics to detect where to pick the index in \mathbf{p} to declare that the condition was violated. The number of elements of \mathbf{p} before the rise was called the “usable rank” and was labeled \mathbf{u} . We then estimated the average error in \mathbf{b} as follows: if \mathbf{S}^U was the same as \mathbf{S}^{-1} , except for the rows beyond $i = \mathbf{u}$, it was set to zero. We then computed \mathbf{x}^U , the more truncated SVD solution, for \mathbf{x} :

$$\mathbf{x}^U = \mathbf{V}\mathbf{S}^U\mathbf{U}^T\mathbf{b} .$$

The residual vector \mathbf{r} is then:

$$\mathbf{r} = \mathbf{A}\mathbf{x}^U - \mathbf{b}$$

and

$$\sigma = \sqrt{\frac{1}{m-u} \sum_{i=1}^{i-m} r_i^2} . \quad (14)$$

We estimated that the average error or standard deviation (σ) in the original \mathbf{b}_i to be the root-mean-square of the elements of \mathbf{r} . It is, however, divided by the number of unused rows, $m-r$, rather than by the number of elements, m . Next, λ is determined by applying the “discrepancy” method. λ is then adjusted until the solution to the regularized problem has the same norm of the residual vector as \mathbf{x}^U . The value of λ is then used to apply the equality and inequality constraints (Lawson and Hanson 1995). Geophysics articles and references divide the study area into a set of prisms (Pick *et al.* 1973), which are selected from the Bear article, but with a difference. In the majority of articles (possibly in all of the articles), the minimum and maximum depths are not clear. In this article, however, they were calculated using the Euler deconvolution method.

A code was written in MATLAB that receives data and divides it into a set of prisms, meaning that the desired area, X_{\min} to X_{\max} , Y_{\min} to Y_{\max} , and Z_{\min} to Z_{\max} , is divided into a set of prisms with same dimensions. The number and dimensions of the prisms can be changed as desired. According to the MATLAB programming language, the area can be simply and quickly divided into 100 000 prisms. The linear operation in $\mathbf{g} = \mathbf{A}\boldsymbol{\rho}$ was used according to Bear *et al.* (1995). The solution of the equation is completed according to the Tikhonov and Lovenberg-Marquardt methods, which determines the amount of λ . In this paper, the Autoreg method was applied to determine the most appropriate amount of λ . This was the first time this method was applied to solve a geophysics and gravity problem. By achieving the best λ and solving the previously discussed equation, the density contrast of all prisms was obtained.

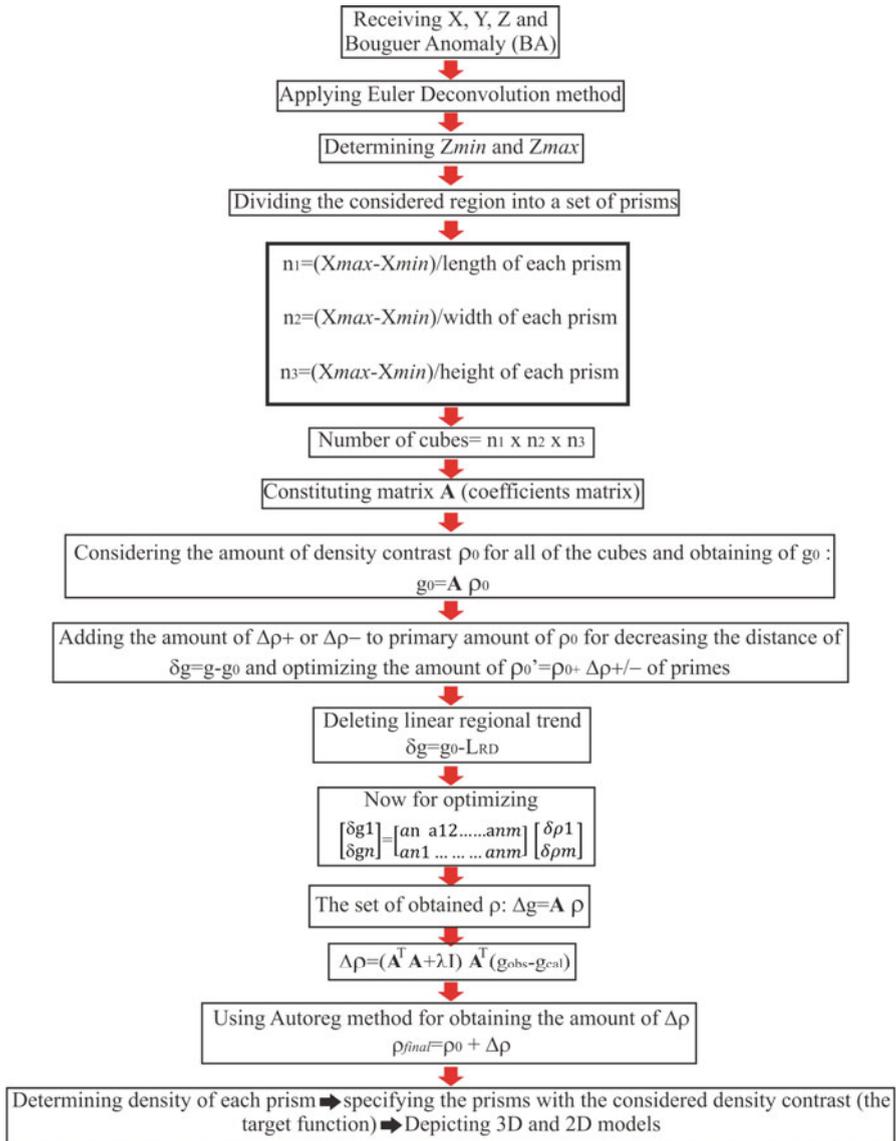


Fig. 2. The main flow chart of a 3D inversion algorithm of gravity data.

The next code section investigated the densities. After solving the equation using the previously mentioned methods for the code, the amount of a density contrast was obtained. The author looked for prisms with 400 density contrasts. After running the MATLAB code, the range of the cubes for the

study were specified. By devoting a trivial percent of error, 395 (minimum) and 405 (maximum), the differences can be determined. The code investigated all the prisms. The prisms with the discrepancy were then determined. They are depicted first in 3D diagram, and then in different 2D forms. Cubes of prisms with density discrepancies of -0.2 g/cm^3 were identified because the density of the region is between 2.3 g/cm^3 , and the salt density is 2.1 g/cm^3 . In this case, prisms with this density discrepancy are depicted. The flow chart of the methodology is shown in Fig. 2.

3. APPLICATION TO SYNTHETIC DATA

A buried anomalous structure is composed of two bodies that are characterized by a density contrast of 0.4 g/cm^3 with respect to the non-anomalous subsurface matter and, more or less, aligned along the X axis (Camacho *et al.* 2000, 2002, 2011). Figure 3 shows this structure. The anomalous masses and mass centers of these bodies are: body 1 with a mass of $264 \times 10^{11} \text{ kg}$ and a depth to its center $Z_1 = -134 \text{ m}$ and body 2 with a mass of $360 \times 10^{11} \text{ kg}$ and a depth to its center $Z_2 = -360 \text{ m}$.

First, three-dimensional Euler deconvolution was applied to the model data. Most of the Euler solutions occurred in the vicinity of models (depth range between 0 to 500 m). The inversion constructed model spaces within a certain distance of the Euler solutions, and then searched systematically in those spaces.

The two L geometrical bodies with a positive density contrast of 0.4 g/cm^3 appear at different depths (Figs. 4-6).

First, we applied the 3D Euler deconvolution method to the synthetic model data. Most of the Euler solutions occurred in the vicinity of models with depths ranging between 0 and 500 m. Then, the model was divided into 960 prisms with dimensions of $200 \times 200 \times 50 \text{ m}$ in the x , y , and z directions, respectively.



Fig. 3. Representation of original anomalous structures for first simulation test (Camacho *et al.* 2002).

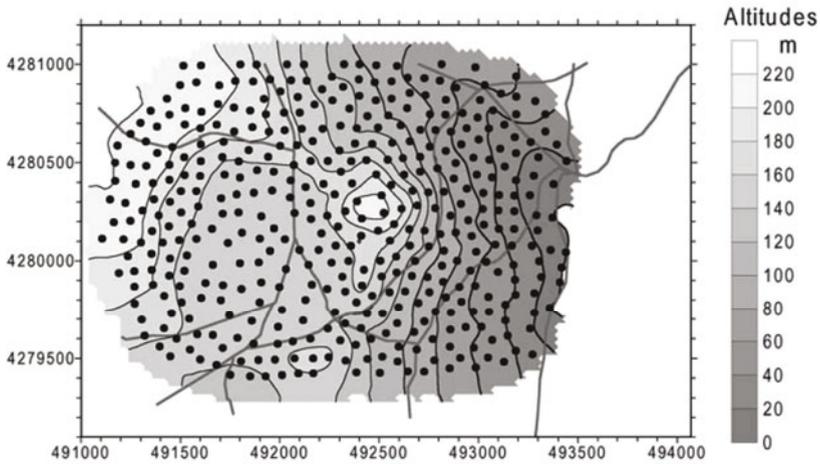


Fig. 4. Simulation example: contour map of altitudes ranging from 0 to 229 m and locations of 420 stations covering an area with a 2000 m diameter with a step of approximately 100 m (Camacho *et al.* 2002).

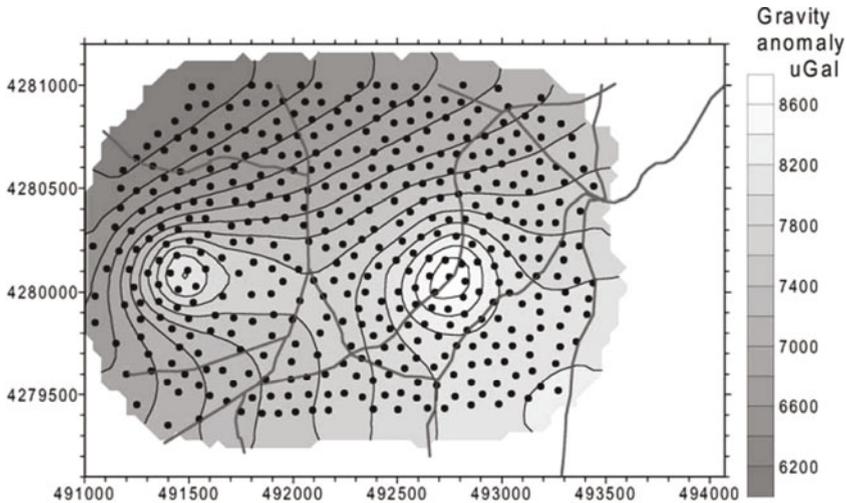


Fig. 5. Gravity anomaly due to anomalous bodies in Fig. 3 with anomalous density contrast of 0.4 g/cm^3 at the stations in Fig. 4 (Camacho *et al.* 2002).

Using the formula of Nagy *et al.* (2000), a coefficient matrix (prism gravity effect) was calculated, resulting in a 420×540 array matrix (being composed of 420 Bouguer anomaly measuring points and 540 prisms). Using the developed algorithm, all prisms with density contrasts of 0.4 g/cm^3 were obtained and are shown in Figs. 7 and 8. It is noticeable that if the in-

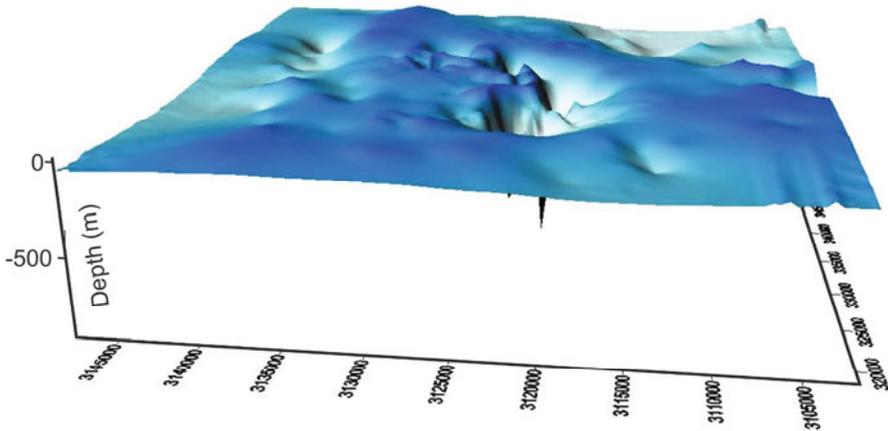


Fig. 6. 3D depth estimation of model with Euler deconvolution.

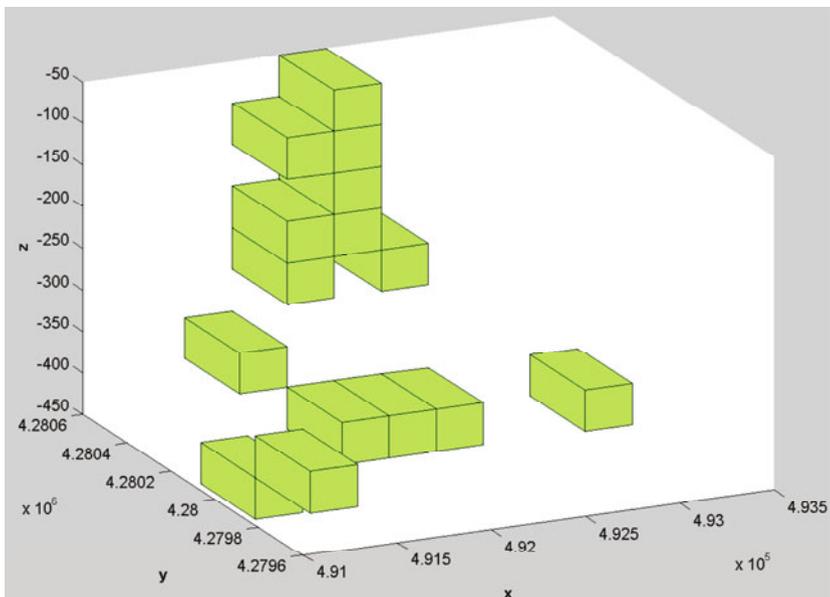


Fig. 7. 3D results of the inversion of gravity data from the synthetic model (green cubes represent prisms with a density contrast of 0.4 g/cm^3).

tended model is a combination of two or more objects with different density contrasts, we can do the search as many times as we wish using this method and depict the results each time in separate forms. Considering the intended model and the results achieved, this method can forecast the intended model with an acceptable rate; such a forecast includes the real depth and form of the anomaly source.

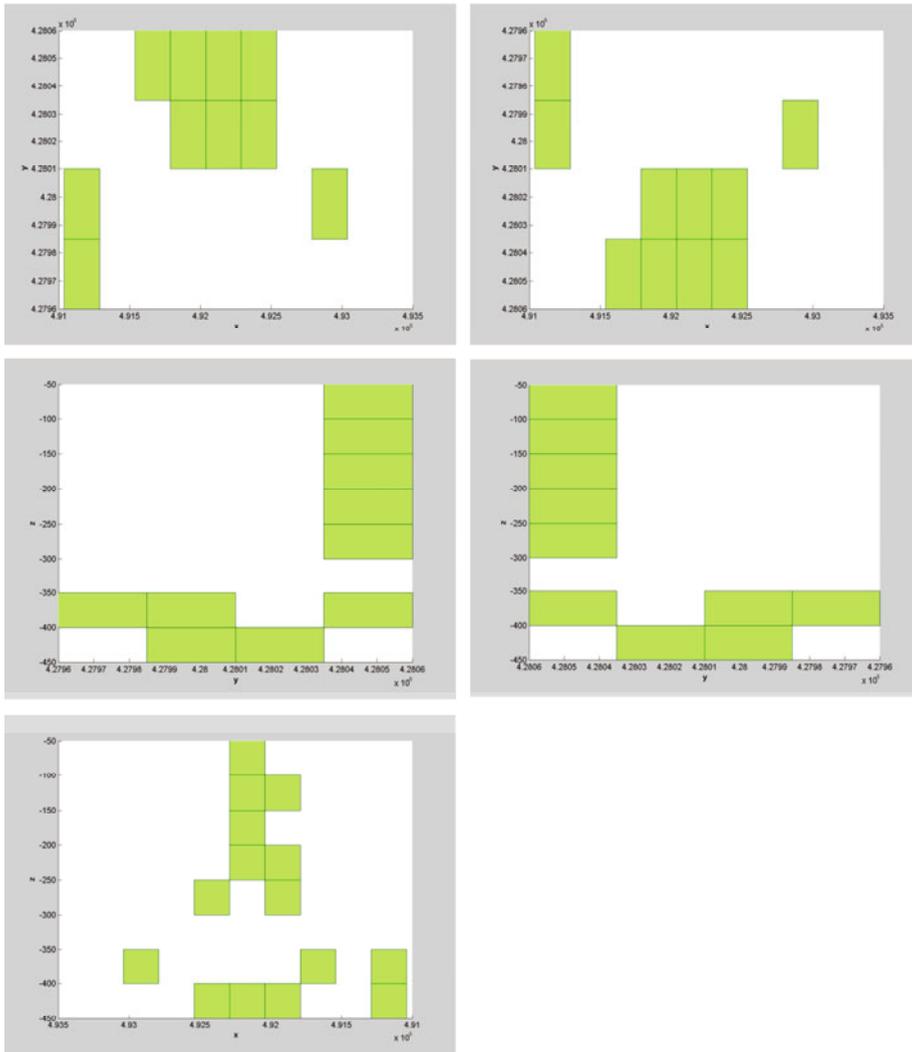


Fig. 8. All 2D views of the gravity inversion results from the synthetic model (green cubes represent prisms with a density contrast of 0.4 g/cm^3).

4. APPLICATION TO REAL GRAVITY DATA

The case study is the Bandar Charak (also known as Dehnow) region, located in southern Iran (Fig. 9a). A common method to determine the Bouguer density value as a random variable independent of topographic alternations was introduced by the Nettleton algorithm. During the correction processes, unexpected errors may occur. For example, the folded region of

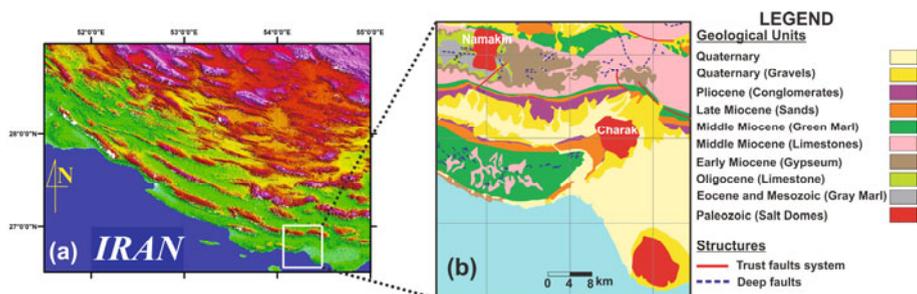


Fig. 9: (a) Elevation map from Shuttle Radar Topographic Mission (SRTM) data showing the location of the study area in Iran; and (b) geologic map of the study area.

Zagros is a result of crustal thickening processes in the southern regions of Iran. Sedimentary sequences in the Charak-Namakin anticlines are important geological units that have been selected for oil prospecting by the National Iranian Oil Company.

Gravimetric data sets were acquired at 776 stations by the National Iranian Oil Company through systematic land surveys. The main target areas along the Charak-Namakin salt domes are geographically located between 54.00-54.30 degrees longitude and 27.00-28.45 degrees latitude. Both the Asmari (Oligocene) and Pabdeh-Gurpi formations (early Cenozoic) contain limestones with gray marl intercalations, which is a potentially valuable facies for hosting hydrocarbon reservoirs under ascending movements of the Paleozoic formations in diapiric systems. The Hormoz series (Cambrian), containing gypsum and other related evaporates, traps oil volatiles that migrated to the nearby permeable brecciated structures. The optimum Bouguer density is determined to only relate to the Charak geological impressions, which associates a number of gravimetric anomalies with the probable oil trap locations. This means that Bouguer anomalies are appropriate for density estimations using statistical techniques. The author applied this technique in the Bandar Charak (Hormozgan-Iran) with various geological/geophysical properties. These inversion results are comparable to both values obtained from density logs in the mentioned area and other methods, such as fractal methods.

The study area is surrounded by the cities of Ashkenan, Ahal, Boochir, Hamiran, Hashniz, and Kemeschck. The Tabnack gas structure is located to the west of this district. The area can be accessed through Asalouie (Bandar Lengeh, Lamard, and Ashkenan-Gaybandy roads) and has a very harsh topography with mountains and valleys. The climate is very hot and wet during the summer and average during the winter.

Geologically, Dehnow is a part of the Fars sedimentary basin in south-east Iran. Salt outcrops can be recognized at two points in the Dehnow anticline. The Khamy formation and Bangestan group are the oldest geological structures in the area that have outcrops. Younger structures are Aghajary, Mokhtari, Mishan, Gachsaran, and Asmary. Dominant structures trend northwest-southeast. The Dehnow anticline is located between the Hendurabi and Razak faults. These faults are almost perpendicular to the Dehnow anticline. Taking the combined geological-residual gravity contour map into account, the Dehnow anticline trends northwest-southeast.

A low gravity anomaly is located in the southeast of the anticline in the salt outcrop. A basic study of the geology of the area, a detailed investigation of the structural features (such as faults associated with the Dehnow anticline) and the application of geophysical techniques and other exploration methods is necessary to investigate the subsurface extension of this anticline and to identify the salt plug intrusion into the anticline. Gravity anomalies are the result of the interference among geological sources with different shapes, densities, and depths. Linear anomalies in geophysical maps, which may correspond to buried faults, contacts, and other tectonic and geological features, are particularly interesting for geologists. Most short-wavelength anomalies are caused by near-surface contacts of rocks that have density contrasts (Esmail Zadeh *et al.* 2010). Table 1 shows the density determination by sampling and system measurements in the Charak region, and Fig. 9b shows the geological map of Charak area.

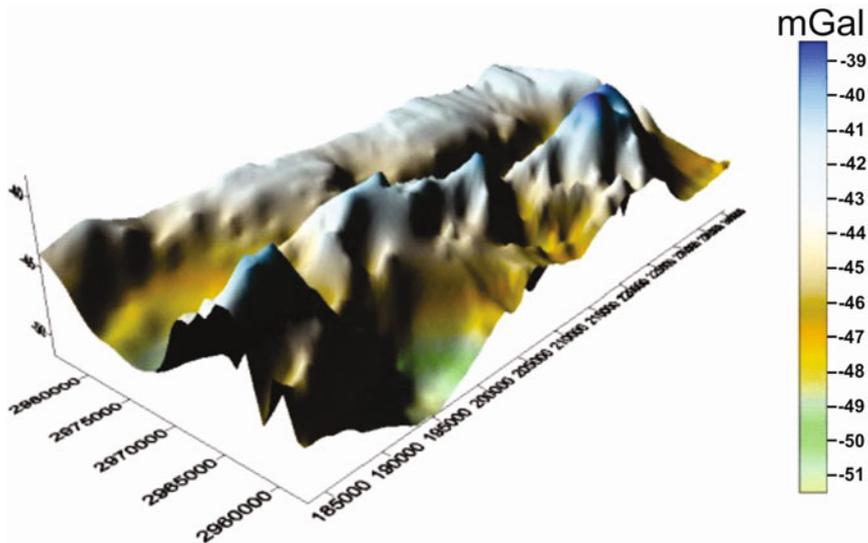


Fig. 10. Bouguer gravity map of the study area.

Table 1

Density determination by sampling and system measurements in the Charak region
(National Oil Company of Iran, reported by Farmani 2003)

Sampling number	Coordinate [deg.]		Stratum	Lithology	Density [g/cm ³]
	Longitude [deg-min-s]	Latitude [deg-min-s]			
L18	54-36-48.2	26-31-48.6	Bakhtiari formation	Conglomerates and sandstone	1.87
L18					1.90
L18					1.90
L19					1.89
L19					1.86
L20	54-17-13.1	26-47-46	Mishan formation	Green marl	2.14
L20					2.12
L21					2.13
L22					2.07
L22					2.14
L23	54-16-57.4	26-48-5.9	Aghajari formation	Sandstone and marl	2.03
L24					2.02
L25					2.04
L26	53-38-17.8	27-5-2	Bangestan group	Limestone	2.45
					2.39
L26					2.41
L27					2.45
					2.39
L28					2.44
L29					2.43
L30	53-38-18.6	27-4-57.3			2.43
L31					2.45
L32					2.44
L35	53-37-43.3	27-4-18.3	Asmari – Gurpi formation	Limestone – gray marl	2.36
L36					2.32
L37					2.32

Note: The mean density of the region is 2.3 g/cm³.

The Bouguer gravity map of Bandar Charak region is presented in Fig. 10. The gravity anomalies range from -51.5 to -38.5 mGal. The 3D depth model of the study area is shown in Fig. 11. The 3D and 2D inversion results of the measured gravity data from the study region case using the developed algorithm are presented in Figs. 12 and 13, respectively.

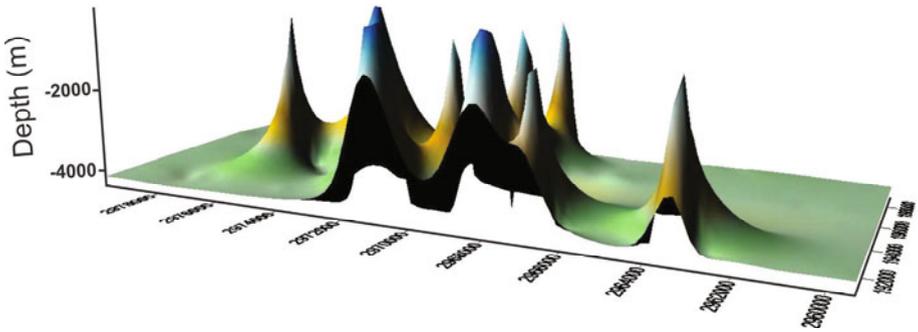


Fig. 11. 3D depth estimation with Euler deconvolution from the study area.

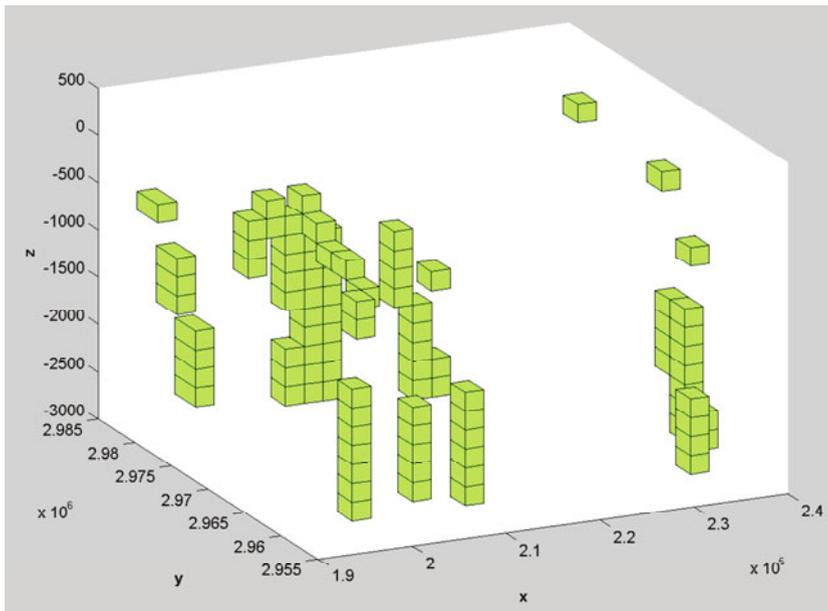


Fig. 12. 3D gravity inversion result of the case study area (green cubes represent prisms with a density contrast of -0.2 g/cm^3).

For the real case study, the minimum and maximum depths were calculated between zero and 4000 m using the Euler deconvolution method. Afterwards, the study area was divided into prisms with dimensions of $2000 \times 2000 \times 200 \text{ m}$ in the x , y , and z directions, respectively.

Using the developed 3D gravity inversion algorithm, all prisms with density contrasts of -0.2 g/cm^3 were detected and are shown in Figs. 12 and 13. In fact, prisms with such a density contrast are indicative of the figure of

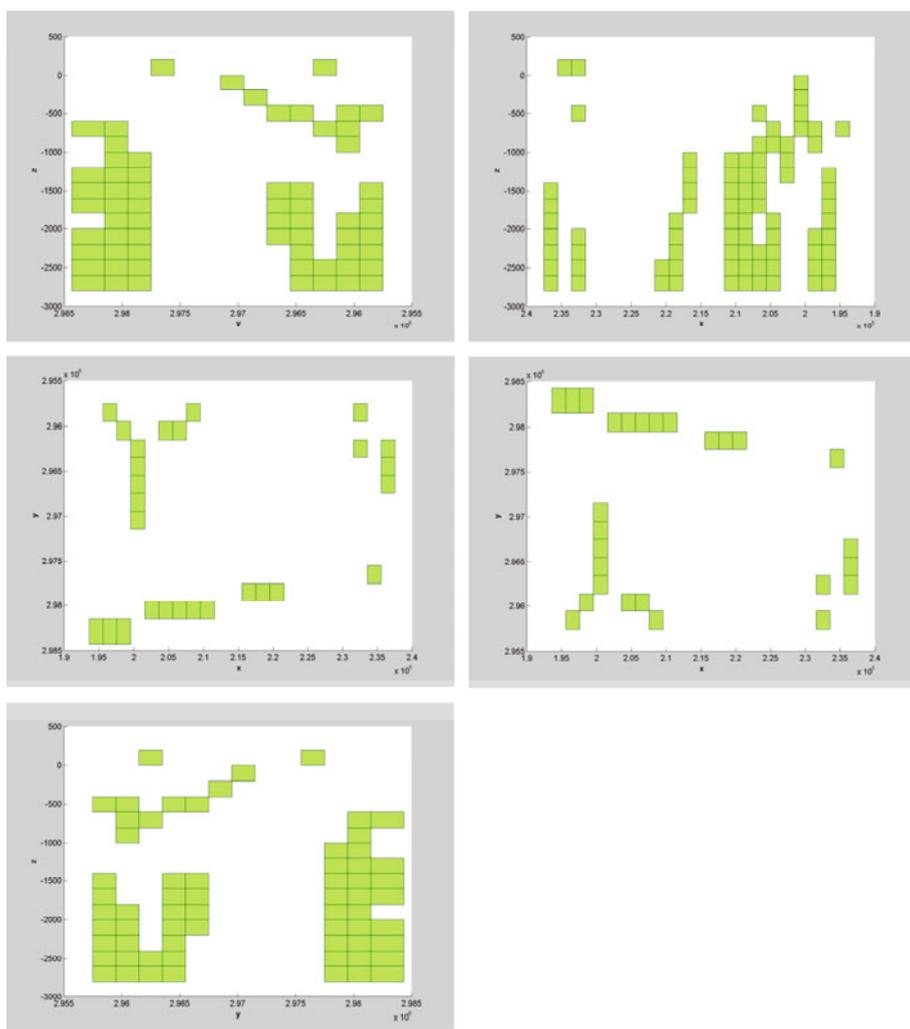


Fig. 13. All 2D results of the gravity inversion of the case study area (green cubes represent prisms with a density contrast of -0.2 g/cm^3).

the salt dome situated within the study area, making these results compatible with the research findings from geological (Bosak *et al.* 1998) and geophysical investigations (Esmacil Zadeh *et al.* 2010) in the study area.

5. CONCLUSIONS

When performing Tikhonov's or similar types of regularization of ill-conditioned linear systems, a free parameter λ must be determined. A given choice of λ implies a corresponding residual for the resulting regularized sys-

tem. An appropriate residual is easily determined from the error level on the right side if it is known. This paper uses an estimate for the error on the right side that can be determined automatically by projecting the rows of the matrix onto the “usable” rows of an orthogonalized version of the system. Thus, the problem of picking λ is transformed into a problem of picking a “usable rank” at which to split the orthogonalized system. In this paper, the top of the salt and a part of the base of salt are not assumed to be accurately imaged. The densities of the background sediments, and thus the density contrast of the salt body, are assumed to be known. To estimate the depth of the top and bottom of the base of the salt, we used Euler deconvolution. We used this depth as *a priori* information in the inversion gravity procedure. Our 3D model shows the shape of the salt dome in this area.

References

- Bear, G.W., H.J. Al-Shukri, and A.J. Rudman (1995), Linear inversion of gravity data for 3-D density distributions, *Geophysics* **60**, 5, 1354-1364, DOI: 10.1190/1.1443871.
- Bosák, P., J. Jaroš, J. Spudil, and P. Sulovský (1998), Salt plugs in the eastern Zagros, Iran: Results of regional geological reconnaissance, *Geolines* **7**, 1, 3-180.
- Camacho, A.G., F.G. Montesinos, and R. Vieira (2000), Gravity inversion by means of growing bodies, *Geophysics* **65**, 1, 95-101, DOI: 10.1190/1.1444729.
- Camacho, A.G., F.G. Montesinos, and R. Vieira (2002), A 3-D gravity inversion tool based on exploration of model possibilities, *Comput. Geosci.* **28**, 2, 191-204, DOI: 10.1016/S0098-3004(01)00039-5.
- Camacho, A.G., J. Fernández, and J. Gottsmann (2011), The 3-D gravity inversion package GROWTH2.0 and its application to Tenerife Island, Spain, *Comput. Geosci.* **37**, 4, 621-633, DOI: 10.1016/j.cageo.2010.12.003.
- Cheng, D. (2003), Inversion of gravity data for base salt, M.Sc. Thesis, Colorado School of Mines, Golden, USA, 95 pp.
- El Dawi, M.G., T. Liu, H. Shi, and D. Luo (2004), Depth estimation of 2-D magnetic anomalous sources by using Euler deconvolution method, *Am. J. Appl. Sci.* **1**, 3, 209-214, DOI: 10.3844/ajassp.2004.209.214.
- Esmail Zadeh, A., F. Doulati Ardejani, M. Ziaii, and M. Mohammado Khorasani (2010), Investigation of salt plugs intrusion into Dehnow anticline using image processing and geophysical magnetotelluric methods, *Russ. J. Earth Sci.* **11**, ES3008, DOI: 10.2205/2009ES000375.
- Farmani, F. (2003), *Gravity Explorations Report in Namakin-Charak Region*, NIOC Press, 129 pp. (in Persian).

- Hansen, R.O., and L. Suciú (2002), Multiple-source Euler deconvolution, *Geophysics* **67**, 2, 525-535, DOI: 10.1190/1.1468613.
- Hsu, S-K. (2002), Imaging magnetic sources using Euler's equation, *Geophys. Prospect.* **50**, 1, 15-25, DOI: 10.1046/j.1365-2478.2001.00282.x.
- Jones, R.E. (2006), Automatically regularized nonnegative solutions for ill-conditioned linear systems. **In:** *Proc. 2006 Inverse Problems in Engineering and Science Conf.*, 13 pp.
- Lawson, C.L., and R.J. Hanson (1995), *Solving Least Squares Problems*, Classics in Applied Mathematics, Vol. 15, SIAM, Philadelphia, 337 pp., DOI: 10.1137/1.9781611971217.fm.
- Li, Y. (2012), Recent advances in 3D generalized inversion of potential-field data. **In:** *Proc. 82nd SEG Annual Meeting*, 4-9 November 2012, Las Vegas, USA, Vol. 2, 878-884.
- Li, Y., and D.W. Oldenburg (1998), 3-D inversion of gravity data, *Geophysics* **63**, 1, 109-119, DOI: 10.1190/1.1444302.
- Mushayandevu, M.F., P. van Driel, A.B. Reid, and J.D. Fairhead (2001), Magnetic source parameters of two-dimensional structures using extended Euler deconvolution, *Geophysics* **66**, 3, 814-823, DOI: 10.1190/1.1444971.
- Nabighian, M.N., V.J.S. Grauch, R.O. Hansen, T.R. LaFehr, Y. Li, J.W. Peirce, J.D. Phillips, and M.E. Ruder (2005), The historical development of the magnetic method in exploration, *Geophysics* **70**, 6, 33ND-61ND, DOI: 10.1190/1.2133784.
- Nagy, D., G. Papp, and J. Benedek (2000), The gravitational potential and its derivatives for the prism, *J. Geodesy* **74**, 7-8, 552-560, DOI: 10.1007/s001900000116.
- Pick, M., J. Picha, and V. Vyskočil (1973), *Theory of the Earth's Gravity Field*, Elsevier, Amsterdam.
- 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.
- Rim, H., Y.S. Park, M. Lim, S.B. Koo, and B.D. Kwon (2007), 3D gravity inversion with Euler deconvolution as a priori information, *Explor. Geophys.* **38**, 1, 44-49, DOI: 10.1071/EG07010.
- Saibi, H., J. Nishijima, T. Hirano, Y. Fujimitsu, and S. Ehara (2008), Relation between structure and low-temperature geothermal systems in Fukuoka city, southwestern Japan, *Earth Planets Space* **60**, 8, 821-826, DOI: 10.1186/BF03352833.
- 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.
- Toushmalani, R. (2011), Application of fuzzy c-means algorithm in Euler deconvolution depth optimization, *Int. J. Acad. Res.* **3**, 2, 44-48.

-
- Toushmalani, R., and M. Hemati (2013), Euler deconvolution of 3D gravity data interpretation: New approach, *J. Appl. Sci. Agric.* **8**, 5, 696-700.
- Toushmalani, R., and H. Saibi (2015), Fast 3D inversion of gravity data using Lanczos bidiagonalization method, *Arab. J. Geosci.* **8**, 7, 4969-4981, DOI: 10.1007/s12517-014-1534-4.

Received 3 February 2014

Received in revised form 1 September 2014

Accepted 26 September 2014



Determination of Sedimentary Basin Basement Depth: A Space Domain Based Gravity Inversion using Exponential Density Function

Vishnubhotla CHAKRAVARTHI and Batta RAMAMMA

Centre for Earth and Space Sciences,
University of Hyderabad, Hyderabad, India;
e-mails: vcvarthi@rediffmail.com (corresponding author),
ramageophd@gmail.com

Abstract

An automatic inversion using ridge regression algorithm is developed in the space domain to analyze the gravity anomalies of sedimentary basins, among which the density contrast decreases with depth following a prescribed exponential function. A stack of vertical prisms having equal widths, whose depths become the unknown parameters to be estimated, describes the geometry of a sedimentary basin above the basement complex. Because no closed form analytical equation can be derivable in the space domain using the exponential density-depth function, a combination of analytical and numerical approaches is used to realize forward gravity modeling. The depth estimates of sediment-basement interface are initiated and subsequently improved iteratively by minimizing the objective function between the observed and modeled gravity anomalies within the specified convergence criteria. Two gravity anomaly profiles, one synthetic and a real, are interpreted using the proposed technique to demonstrate its applicability.

Key words: gravity anomaly, exponential density contrast, automatic inversion, space domain, sedimentary basin.

1. INTRODUCTION

Gravity method plays an important role in the studies related to sedimentary basin modeling because detectable gravity anomalies can be observable on the surface of the Earth due to the presence of significant density contrast between sediment infill and the underlying basement. These observed gravity anomalies, considered to have been made on topographic elevations, can be modeled to decipher the geometry of the basement structure below the sedimentary load. Due to deficit in density of sedimentary rocks with the underlying basement complex, negative gravity anomalies are usually observed over sedimentary basins. It is well-known that the interpretation of gravity anomalies for subsurface density structure(s) is a non-unique problem, because the surface gravity anomalies can be explained by a variety of mass distributions at different depths (Blakely 1995). Such an ambiguity in gravity interpretation is often tackled by assigning a mathematical geometry to the anomalous mass with a known density and then to invert the anomalies for the unknown parameters such that the estimated structure is geologically sensible. Using *a priori* information in model space (derived from drilling/ other geophysical data) would further reduce the degree of uncertainty in interpretation.

Many 2D indirect methods (forward modeling) are available to compute the gravity anomalies of geological sources with uniform density (*e.g.*, Won and Bevis 1987, Singh 2002). However, these forward modeling techniques find limited application in analyzing gravity anomalies of sedimentary basins because the parameters describing the structure are not known in advance. Bott (1960) and Murthy and Rao (1989) proposed direct methods of interpretation to solve the structures of sedimentary basins from the observed gravity anomalies, where the cross-section of a sedimentary basin was viewed as a collage of vertical prisms, all having equal widths and uniform density. However, the assumption of uniform density for sedimentary rocks is seldom valid in reality (Abdoh *et al.* 1990, Abbott and Louie 2000, Chakravarthi 2003, Gómez-Ortiz *et al.* 2005, Mantlík and Matias 2010, Kadima *et al.* 2011). Cowie and Karner (1990) demonstrated that the density-depth curves constructed for different stratigraphic units in sedimentary basins from measured well density logs exhibit a range of densities but the mean sediment density clearly increases with depth with the highest rate of increase in the top few kilometers. Such a variation of density with depth can be effectively simulated by an exponential density function if differential compaction is assumed to be the most important diagenetic process in the evolution of sedimentary basins (Cordell 1973, García-Abdeslem 1992). Hence, the use of exponential density-depth function in the analysis of gravity anomalies of sedimentary basins often paves the way for more reliable interpreta-

tions. However, it becomes a strenuous task to accommodate exponential density variation in the existing forward modeling schemes, including some of the available commercial software such as GM-SYS (Northwest Geophysical Associates 2004), because large numbers of constant density bodies are required to adequately explain the same density structure (Zhou 2013).

Direct modeling methods are being developed to analyze the gravity anomalies of sedimentary basins using the exponential density-depth function. Because of the fact that no closed form analytical equations can be derivable in the space domain for the gravity anomalies with an exponential density function, many algorithms perform forward modeling in the frequency domain and then transform the anomalies back to the space domain for further analysis. For instance, Cordell (1973) developed a recursive method made use of both the gravity field and its vertical derivative (determined by convolution in discrete Fourier series) to solve the structure of a sedimentary basin from observed gravity anomalies. Granser (1987) had calculated the gravity effect of a structure based on series expansion, the numerical evaluation of which was performed by fast Fourier transform. Chai and Hinze (1988) proposed methods to analyze both profile and two-dimensional gravity data, where the forward modeling was realized in the wave number domain followed by its conversion to the space domain by a shift-sampling technique. Rao *et al.* (1993) derived Fourier transforms of gravity anomalies of some simple geometric models with exponential density contrast and used them in the analysis of the gravity anomalies of sedimentary basins; however, these strategies likely to yield unreliable interpretations when the sediment-basement interface has major undulations as in the case of the San Jacinto graben, California. The method developed by Rao and Rao (1999) also involved the calculation of the gravity effect in the frequency domain and its subsequent transformation to the space domain by Filon's method (Filon 1928). In recent past, Chappell and Kusznir (2008) extended the method of Granser (1987) to calculate the gravity anomaly as a function of the Fourier transforms of the bounding surfaces of a basin with irregular top and bottom surfaces. Nonetheless, the enlisted methods incur truncation errors when the modeled anomalies transform from the frequency domain to the space domain (Chakravarthi and Sundararajan 2007).

On the other hand, a few space domain based algorithms are available to model the gravity anomalies, where the exponential density variation was accommodated in the interpretation by alternative means. For instance, Murthy and Rao (1979) proposed the subdivision of each side of a 2D polygon into a number of segments, along which the density contrast was assumed to vary linearly with depth, whereas Guspi (1990) described the exponential density-depth function by a series approximation to compute the gravity response. The method of Murthy and Rao (1979) consumes signifi-

cant amount of time even for forward modeling (Rao *et al.* 1994), whereas the method of Guspi (1990) requires the knowledge of the degree of the polynomial, which is generally not known *a priori*.

In this paper, we develop a space domain based inversion technique to analyze the gravity anomalies of sedimentary basins, among which the density contrast varies exponentially with depth. The present technique uses ridge regression algorithm to analyze the gravity anomalies. The applicability of the method is exemplified with both synthetic and real field examples.

2. FORWARD MODELING – THEORETICAL CONSIDERATIONS

Figure 1 shows the cross-section of a sedimentary basin. Let the profile, AA', run along the x -axis transverse to the strike of the basin. The structure of the basin is approximated by a series of outcropping vertical prisms put in juxtaposition and having equal widths. In a Cartesian coordinate system, let $2T$ be the width of one such prism along the x -axis and z_B represent the thickness of the corresponding prism along the z -axis, positive vertically downwards (Fig. 1). Let the density contrast along the prism vary vertically with depth following the exponential equation of the form (Cordell 1973)

$$\Delta\rho(z) = \Delta\rho_0 e^{-\lambda z}, \quad (1)$$

where $\Delta\rho(z)$ is the density contrast at any depth z , $\Delta\rho_0$ is the density contrast observed at the ground surface, λ is a constant expressed in inverse length units. The gravity anomaly of one such prism, $\Delta g_{\text{prm}}(x_j, z_j)$, at any observation, $P(x_j, z_j)$, on the profile, AA', outside the source region can be expressed as

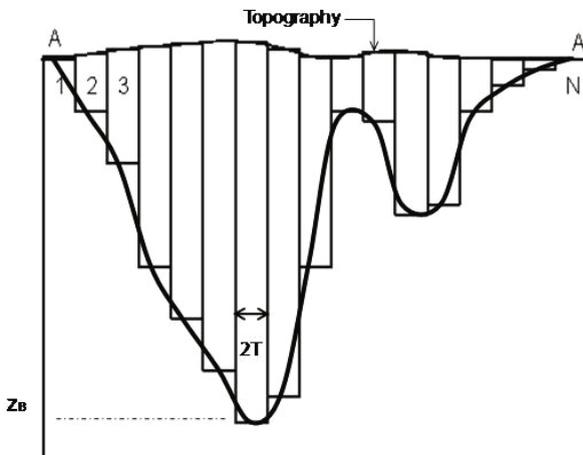


Fig. 1. Cross-section of a sedimentary basin (curved line) and its approximation by an ensemble of vertical prisms (step line).

$$\Delta g_{\text{prm}}(x_j, z_j) = 2G \int_s \Delta \rho(z) \frac{(z - z_j) dx dz}{\left[x - x_j - z + z_j \right]} \quad (2)$$

Here, G is universal gravitational constant and (x, z) are source coordinates. Substituting Eq. 1 for $\Delta \rho(z)$ in Eq. 2 and upon integration within the limits of x , Eq. 2 takes the form

$$\Delta g_{\text{prm}}(x_j, z_j) = 2G \Delta \rho_0 \int_0^{z_b} e^{-\lambda z} \left[\tan^{-1} \frac{(x - x_j + T)}{z - z_j} - \tan^{-1} \frac{(x - x_j - T)}{z - z_j} \right] dz \quad (3)$$

Equation 3 has to be solved numerically because no closed form solution exists for it in the space domain. Further, it is to be realized that Eq. 3 is strictly valid for the profile, AB, which runs transverse to the strike of the prism. In case the profile runs at an angle, α , with the x -axis then x_j in Eq. 3 needs to be replaced by $x_j \cos \alpha$ (Chakravarthi and Ramamma 2013).

The total gravity anomaly produced by a basin at any observation can be obtained as

$$\Delta g_{\text{basin}}(x_j, z_j) = \sum_{i=1}^N \Delta g_{\text{prm}}(x_i, z_i) \quad (4)$$

where N stands for the number of prisms/observations on the profile.

3. INVERSION OF GRAVITY ANOMALIES

The objective of gravity inversion is to fit the modeled gravity anomalies to the observed ones by adjusting the thickness parameters of the basin in a least square sense such that the modeled gravity response of the optimum depth structure mimics the observed anomaly. The present algorithm consists of two modules, one for initialization of the model space and the other for refinement. The algorithm performs both modules automatically in the sense that it initiates the depth structure of a sedimentary basin from observed gravity anomalies and improves the structure iteratively, based on the differences between the observed and modeled gravity anomalies within the specified convergence criteria.

For initialization, it is presumed that the observed gravity anomaly at each station on the profile is being generated by an infinite horizontal subsurface slab in which the density contrast decreases exponentially with depth following Eq. 1. Accordingly, initial or approximate depths to basement interface are estimated using the Bouguer slab formula of Cordell (1973)

$$z_{\text{Bou}} = \frac{-1}{\lambda} \log \left(1 - \frac{\lambda \Delta g_{\text{obs}}(x_i, z_i)}{2\pi G \Delta \rho_0} \right) \quad (5)$$

where $\Delta g_{\text{obs}}(x_i, z_i)$ is the observed residual gravity anomaly at any station.

The modeled gravity response of the basin, $\Delta g_{\text{basin}}(x_j, z_j)$, calculated at each observation on the profile using Eq. 4 obviously deviates from the observed anomaly, $\Delta g_{\text{obs}}(x_j, z_j)$, because the initial depths obtained from Eq. 5 are only approximate. The difference between these two anomalies at any observation at the end of the k -th iteration can be expressed as the cumulative of the products of vertical gradients of the anomaly and corresponding depth improvements of the prisms as

$$\Delta g_{\text{obs}}(x_j, z_j) - \Delta g_{\text{basin}}(x_j, z_j) = \sum_{i=1}^N \left[\frac{\partial \Delta g_{\text{prism}}(x_i, z_i)}{\partial z} \right]_{z_k} dz_i . \quad (6)$$

Linear equation similar to Eq. 6 is constructed for each observation and N normal equations are framed and solved for the improvements in N depth parameters of the prisms by minimizing the objective function J , defined as the sum of the squares of the differences between the observed and modeled gravity anomalies,

$$J = \sum_{j=1}^N \left[\Delta g_{\text{obs}}(x_j, z_j) - \Delta g_{\text{basin}}(x_j, z_j) \right]^2 . \quad (7)$$

The system of normal equations can be expressed in a matrix form as

$$(A + \delta I)X = B , \quad (8)$$

where A is a nxn matrix whose elements A_{nj} are given by

$$A_{nj'} = \sum_{n=1}^N \sum_{m=1}^N \frac{\partial \Delta g_{\text{prism}}(x_m, z_m)}{\partial a_{j'}} \frac{\partial \Delta g_{\text{prism}}(x_m, z_m)}{\partial a_n} , \quad j' = 1, 2, \dots, N, \quad (9)$$

$$X = da_n , \quad (10)$$

$$B = \sum_{m=1}^N \left[\Delta g_{\text{obs}}(x_m, z_m) - \Delta g_{\text{basin}}(x_m, z_m) \right] \frac{\partial \Delta g_{\text{prism}}(x_m, z_m)}{\partial a_{j'}} , \quad j' = 1, 2, \dots, N, \quad (11)$$

Here, a_n , $n = 1, 2, \dots, N$ are depth parameters of prisms and da_n represents corresponding depth improvements. δ is the damping factor, and I is a diagonal matrix containing the diagonal elements of the matrix A . The partial derivatives required in Eqs. 9-10 are evaluated numerically, which involves the calculation of the rate of change of the gravity anomaly with respect to the thickness of each prism. Initially the value of δ is set to an arbitrary value of 0.5 and Eq. 8 is solved for the increments/decrements da_n and subsequently added to/subtracted from existing parameters of a_n to obtain the improved parameters, a'_n . If the current value of the objective function, J_{mod} , obtained with the improved parameters, a'_n , is less than its previous value, J , then J_{mod} is assigned to J and a'_n to a_n , and the damping factor, δ , is de-

creased by a factor of $1/2$. In case J_{mod} is greater than J , then the value of δ is doubled and Eq. 8 is solved for da_n and are added to/subtracted from a_n and this process repeats until J_{mod} becomes less than or equal to J . The algorithm terminates in case:

- the specified number of iterations are completed or,
- the current value of the objective function J falls below user specified predefined allowable error or,
- the current value of δ attains a large value.

4. APPLICATIONS

The applicability of the proposed inversion technique is demonstrated on two examples: one synthetic and a real. In case of real field example, the observed gravity anomalies across the San Jacinto graben, California, are analyzed and compared with those previously reported. In both cases the density contrast varies exponentially with depth and observer locations are at the top of the topography at $z_j = 0$.

4.1 Synthetic example

Figure 2a shows 40 equispaced noisy gravity anomalies (solid circles) in the interval $x \in [0 \text{ km}, 40 \text{ km}]$ produced by a synthetic model, whose structure resembles a typical block faulted intracratonic rift basin filled with thick sectioned sediments (Fig. 2b). In this case, the pseudorandom noise was Gaussian with zero mean and a standard deviation of 0.12 mGal. We assume that the density contrast of sediments within the basin varies according to Eq. 1 with $\Delta\rho_0 = -0.45 \text{ gm/cm}^3$, and $\lambda = 0.39 \text{ km}^{-1}$. The anomaly shows asymmetric nature across the strike of the structure with large gradient observed over the western margin of the basin (Fig. 2a). The gravity profile along which the interpretation is intended covers the lateral dimensions of the basin and extends farther away to stations resting on the basement. We interpret the gravity anomalies using the proposed inversion technique described in the text to recover the basin structure. For such an inversion, the algorithm had performed 13 iterations before it got terminated as the misfit fell below a predefined allowable error of 10^{-5} mGal. The initial structure of the basin estimated by the algorithm is shown in Fig. 2b and the corresponding anomaly produced by it in Fig. 2a, respectively. The objective function J , reduced drastically from its initial value of 86.8 to 0.0007 at the end of the 3rd iteration and then gradually reaches to almost 0.0 at the end of the 13th iteration (Fig. 2c). The variation of the damping factor, δ , with iteration during the process of inversion is shown in Fig. 2d. The fit between the observed and modeled gravity anomalies (shown in Fig. 2a as a solid line) at the end of the 13th iteration is satisfactory. A maximum error of 0.006 mGal between these

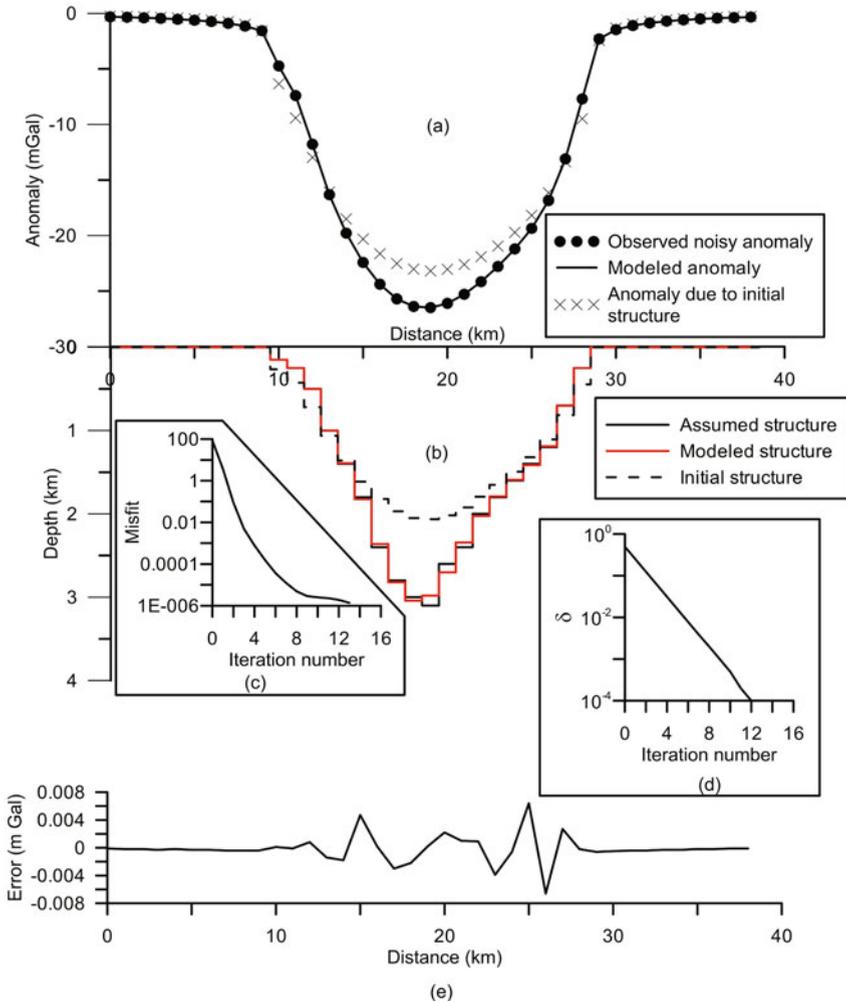


Fig. 2: (a) Observed and modeled gravity anomalies for initial and estimated depth structures of a synthetic model of a sedimentary basin using exponential density function; (b) assumed, initial and estimated depth structures; (c) variation of objective function with iteration; (d) changes in the damping factor against the iteration number; and (e) error between the observed and modeled gravity anomalies after inversion.

two anomalies is observed at the 25th km on the profile (Fig. 2e). The estimated structure subsequent to inversion is shown in Fig. 2b along with the assumed structure. No significant changes in the estimated depth are noticed beyond the concluding iteration.

By and large, the algorithm successfully recovered the structure with a few deviations in and around the depocentre (Fig. 2b). A maximum error of 3.9% between the assumed and estimated depths is observed at the 20th km on the profile (Fig. 2b). Such an error is insignificant considering the presence of noise in the residual signal of the structure.

4.2 Field example – San Jacinto graben, California

The San Jacinto graben is bounded by two parallel branches of the San Jacinto fault, and has a northwesterly trend (Cordell 1973). The observed negative gravity anomalies over the graben (Fig. 3a) are attributed to the low density basin fill consisting of Pliocene and Pleistocene detrital sedimentary rocks and Pleistocene and Holocene alluvium. On the basis of seismic refraction data, Fett (1968) had determined the depth to the basement in the center of the graben as 2.4 km. Cordell (1973) had used two exponential density depth models based on Eq. 1, one with $\Delta\rho_0 = -0.55 \text{ gm/cm}^3$ and $\lambda = 0.5 \text{ km}^{-1}$ (shown as solid line in Fig. 3c) and the other with $\Delta\rho_0 = -0.55 \text{ gm/cm}^3$ and $\lambda = 1.0 \text{ km}^{-1}$ (shown as dashed line in Fig. 3c) to describe the density contrast-depth data of the graben derived from seismic refraction surveys (shown as step line in Fig. 3c). He demonstrated that the use of former density model in the gravity analysis of the graben had yielded structural solution (shown as dashed line in Fig. 3b) that was consistent with the seismically derived information by Fett (1968).

For the present case, we have interpreted the same gravity data using our inversion technique to decipher its basement configuration. The exponential density model defined with $\Delta\rho_0 = -0.55 \text{ gm/cm}^3$ and $\lambda = 0.5 \text{ km}^{-1}$ (Cordell 1973) has been used in the present inversion. We have digitized the anomaly at an interval of 0.322 km (Fig. 3a) covering a profile length of about 10.0 km across the graben and subjected the anomalies for inversion. In this case, the algorithm performed 114 iterations (Fig. 3d) before it got terminated. The approximate structure of the graben shown in Fig. 3b was based on the initial depths calculated using Eq. 5. The corresponding gravity response of the initial structure is shown in Fig. 3a. The objective function corresponding to the initial structure was 252.8 (Fig. 3d). The changes in the damping factor, δ , (dashed line) and objective function (solid line) against the iteration number were shown in Fig. 3d. The modeled gravity anomalies at the end of the 114th iteration are shown in Fig. 3a and the inferred structure of the graben in Fig. 3b, respectively. No significant changes are found in estimated depths of the graben subsequent to the concluding iteration. The modeled gravity anomalies and the inferred structure of the graben interpreted by Cordell (1973) are also shown in Fig. 3a and b for comparison. The maximum depth to the basement estimated from present inversion is 2.59 km,

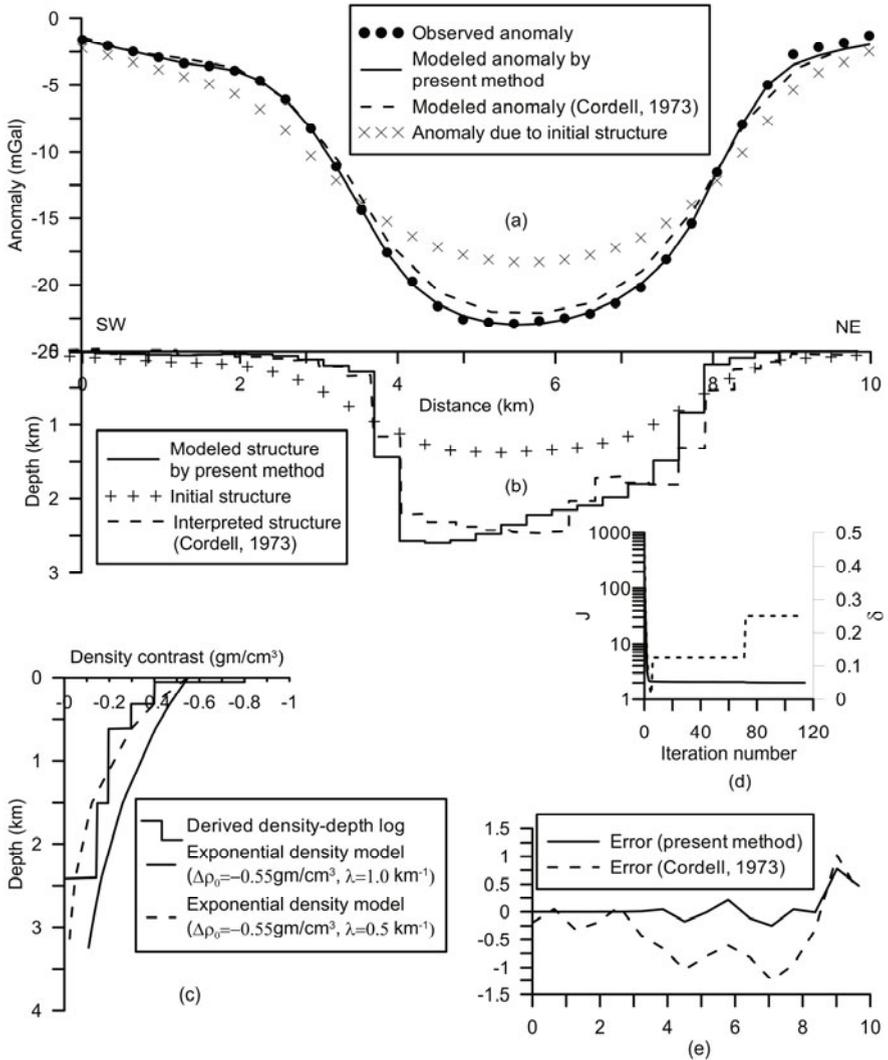


Fig. 3: (a) Observed and modeled gravity anomalies for initial and estimated depth structures using exponential density model, San Jacinto graben, California; modeled anomalies by Cordell (1973) are also shown; (b) initial and estimated depth structure by the present method; inferred depth model by Cordell (1973) is also shown; (c) predicted density contrast-depth data by Fett (1968) (solid step line) and fitted exponential density models by Cordell (1973); solid line corresponds to $\Delta\rho_0 = -0.55 \text{ gm/cm}^3$ and $\lambda = 0.5 \text{ km}^{-1}$ and dashed line corresponds to $\Delta\rho_0 = -0.55 \text{ gm/cm}^3$ and $\lambda = 1.0 \text{ km}^{-1}$; (d) variation of objective function and damping factor with iteration number; and (e) error between the observed and modeled gravity anomalies after inversion.

whereas Cordell (1973) concluded that the maximum thickness of sediments within the graben was at least 2.44 km. The error between the observed and modeled gravity anomalies in each case is shown in Fig. 3e.

By and large, the structural models inferred from the present technique and the one by Cordell (1973) show more or less similar morphological features over the shoulders of the graben. For instance, both models divulge that the graben is bounded by steeply dipping fault system towards southwest and by gently dipping fault system towards northeast. However, the structural model deciphered from our technique reveals a progressive deepening of the basement towards southwest, which however was not repeated in Cordell's (1973) model (Fig. 3b). The basement configuration derived from our analysis appears to be more reliable than the one inferred by Cordell (1973) because the modeled gravity anomalies of the structure from the present inversion more closely fit the observed gravity anomalies (Fig. 3a) than the ones realized by Cordell (1973). Further, the error between the observed and modeled gravity anomaly in our case hardly exceeds 0.25 mGal in the central part of the basin, whereas it was more than -1.1 mGal in Cordell's (1973) interpretation.

5. CONCLUSIONS

We develop a space domain based inversion technique using the ridge regression algorithm to analyze the gravity anomalies of sedimentary basins among which the density contrast obeys exponential decrease with depth. We realize forward gravity modeling of sedimentary basins in the space domain using a combination of analytical and numerical approaches. The present technique initiates the structure of a sedimentary basin with an assumption that the observed gravity anomaly at each observation is being generated by an infinite subsurface horizontal slab in which the density contrast decreases exponentially with depth. The proposed inversion generates the initial structure of a sedimentary basin from the observed gravity anomalies and improves the structure in an iterative approach based on the differences between the observed and modeled gravity anomalies until the modeled anomalies closely mimic the observed ones.

The applicability and validity of this inversion technique is demonstrated on both synthetic and real field gravity anomalies. In case of synthetic example pseudorandom noise is added to the residual gravity signal produced by a sedimentary basin prior to inversion. The algorithm almost recovered the basin structure even in the presence of random noise; however insignificant error in the estimated depth is found in around the depocentre of the assumed structure. Further, the observed gravity anomalies over the San Jacinto graben, California have been analyzed using our technique and the estimated structure is compared with the structural model derived originally

by Cordell (1973). The estimated maximum thickness of the graben from the proposed inversion more or less coincides with the maximum thickness inferred by Fett (1968) and Cordell (1973). However, our deciphered model of the graben modestly deviates from the one by Cordell (1973) in the sense that our model shows progressive deepening of the basement towards the southwest, which was not repeated in later case. The fact that the present inversion technique yielded information that is consistent with the assumed parameters in case of synthetic structure and with available information in case of field example testifies the applicability of the technique.

Acknowledgements. The authors sincerely thank the anonymous reviewers for their critical reviews and feedback to revise the manuscript as presented.

References

- Abbott, R.E., and J.N. Louie (2000), Depth to bedrock using gravimetry in the Reno and Carson City, Nevada, area basins, *Geophysics* **65**, 2, 340-350, DOI: 10.1190/1.1444730.
- Abdoh, A., D. Cowan, and M. Pilkington (1990), 3D gravity inversion of the Cheshire basin, *Geophys. Prospect.* **38**, 8, 999-1011, DOI: 10.1111/j.1365-2478.1990.tb01887.x.
- Blakely, R.J. (1995), *Potential Theory in Gravity and Magnetic Applications*, Cambridge University Press, Cambridge.
- Bott, M.H.P. (1960), The use of rapid digital computing methods for direct gravity interpretation of sedimentary basins, *Geophys. J. Int.* **3**, 1, 63-67, DOI: 10.1111/j.1365-246X.1960.tb00065.x.
- Chai, Y., and W.J. Hinze (1988), Gravity inversion of an interface above, which the density contrast varies exponentially with depth, *Geophysics* **53**, 6, 837-845, DOI: 10.1190/1.1442518.
- Chakravarthi, V. (2003), Digitally implemented method for automatic optimization of gravity fields obtained from three-dimensional density interfaces using depth dependent density, US Patent 6615139 B1.
- Chakravarthi, V., and B. Ramamma (2013), Gravity anomaly modeling of multiple geological sources having different strike lengths and arbitrary density contrast variations, *Near Surf. Geophys.* **11**, 4, 363-370, DOI: 10.3997/1873-0604.2013001.
- Chakravarthi, V., and N. Sundararajan (2007), Marquardt optimization of gravity anomalies of anticlinal and synclinal structures with prescribed depth-dependent density, *Geophys. Prospect.* **55**, 4, 571-587, DOI: 10.1111/j.1365-2478.2007.00625.x.

- Chappell, A.R., and N.J. Kusznir (2008), Three-dimensional gravity inversion for Moho depth at rifted continental margins incorporating a lithosphere thermal gravity anomaly correction, *Geophys. J. Int.* **174**, 1, 1-13, DOI: 10.1111/j.1365-246X.2008.03803.x.
- Cordell, L. (1973), Gravity analysis using an exponential density-depth function – San Jacinto Graben, California, *Geophysics* **38**, 4, 684-690, DOI: 10.1190/1.1440367.
- Cowie, P.A., and G.D. Karner (1990), Gravity effect of sediment compaction: examples from the North Sea and the Rhine Graben, *Earth Planet. Sci. Lett.* **99**, 1-2, 141-153, DOI: 10.1016/0012-821X(90)90078-C.
- Fett, J.D. (1968), Geophysical investigation of the San Jacinto Valley, Riverside County, California, M.Sc. Thesis, University of California, Riverside, USA, 87 pp.
- Filon, L.N.G. (1928), On a quadrature formula for trigonometric integrals, *Proc. Roy. Soc. Edinburgh* **49**, 1, 38-47.
- García-Abdeslem, J. (1992), Gravitational attraction of a rectangular prism with depth-dependent density, *Geophysics* **57**, 3, 470-473, DOI: 10.1190/1.1443261.
- Gómez-Ortiz, D., R. Tejero-López, R. Babín-Vich, and A. Rivas-Ponce (2005), Crustal density structure in the Spanish Central System derived from gravity data analysis (Central Spain), *Tectonophysics* **403**, 1-4, 131-149, DOI: 10.1016/j.tecto.2005.04.006.
- Granser, H. (1987), Three-dimensional interpretation of gravity data from sedimentary basins using an exponential density-depth function, *Geophys. Prospect.* **35**, 9, 1030-1041, DOI: 10.1111/j.1365-2478.1987.tb00858.x.
- Guspi, F. (1990), General 2D gravity inversion with density contrast varying with depth, *Geoexploration* **26**, 4, 253-265, DOI: 10.1016/0016-7142(90)90007-F.
- Kadima, E., D. Delvaux, S.N. Sebagenzi, L. Tack, and S.M. Kabeya (2011), Structure and geological history of the Congo Basin: an integrated interpretation of gravity, magnetic and reflection seismic data, *Basin Res.* **23**, 5, 499-527, DOI: 10.1111/j.1365-2117.2011.00500.x.
- Mantlik, F., and M.J.S. Matias (2010), Interpretation and modeling of regional gravity data of the Aveiro Basin (Northwest Portugal), *C. R. Geosci.* **342**, 11, 823-836, DOI: 10.1016/j.crte.2010.06.005.
- Murthy, I.V.R., and D.B. Rao (1979), Gravity anomalies of two-dimensional bodies of irregular cross-section with density contrast varying with depth, *Geophysics* **44**, 9, 1525-1530, DOI: 10.1190/1.1441023.
- Murthy, I.V.R., and S.J. Rao (1989), A FORTRAN 77 program for inverting gravity anomalies of two-dimensional basement structures, *Comput. Geosci.* **15**, 7, 1149-1156, DOI: 10.1016/0098-3004(89)90126-X.

- Northwest Geophysical Associates (2004), GM-SYS gravity/magnetic modeling software. User's guide, Version 4.9, Northwest Geophysical Associates, Inc., Corvallis, USA, 101 pp.
- Rao, C.V., V. Chakravarthi, and M.L. Raju (1994), Forward modelling: Gravity anomalies of two-dimensional bodies of arbitrary shape with hyperbolic and parabolic density functions, *Comput. Geosci.* **20**, 5, 873-880, DOI: 10.1016/0098-3004(94)90118-X.
- Rao, D.B., and C.P.V.N.J.M. Rao (1999), Two-dimensional interpretation of gravity anomalies over sedimentary basins with an exponential decrease of density contrast with depth, *Proc. Indian Acad. Sci. (Earth Planet. Sci.)* **108**, 2, 99-106, DOI: 10.1007/BF02840488.
- Rao, D.B., M.J. Prakash, and N.R. Babu (1993), Gravity interpretation using Fourier transforms and simple geometrical models with exponential density contrast, *Geophysics* **58**, 8, 1074-1083, DOI: 10.1190/1.1443491.
- Singh, B. (2002), Simultaneous computation of gravity and magnetic anomalies resulting from a 2-D object, *Geophysics* **67**, 3, 801-806, DOI: 10.1190/1.1484524.
- Won, I.J., and M. Bevis (1987), Computing the gravitational and magnetic anomalies due to a polygon: Algorithms and Fortran subroutines, *Geophysics* **52**, 2, 232-238, DOI: 10.1190/1.1442298.
- Zhou, X. (2013), Gravity inversion of 2D bedrock topography for heterogeneous sedimentary basins based on line integral and maximum difference reduction methods, *Geophys. Prospect.* **61**, 1, 220-234, DOI: 10.1111/j.1365-2478.2011.01046.x.

Received 2 October 2013

Received in revised form 13 June 2014

Accepted 22 July 2014



Seismicity of Czorsztyn Lake Region: A Case of Reservoir Triggered Seismic Process?

Wojciech BIAŁOŃ, Ewa ZARZYCKA, and Stanisław LASOCKI

Institute of Geophysics, Polish Academy of Sciences, Warszawa, Poland
e-mails: w.bialon@igf.edu.pl (corresponding author), ezarzycka@igf.edu.pl,
lasocki@igf.edu.pl

Abstract

Czorsztyn Lake is an artificial water reservoir backed up by the hydropower plant Niedzica earth dam on Dunajec River in south Poland. Its filling began in 1995 and ended in 1997. The reservoir of 234.5 million m³ capacity is shallow, between 20 to 50 m of water column, on average. Until 2011 the seismic activity in this region was sparse, some 1 event trimonthly. However, in November 2011 more than 60 events occurred. Such bursts of activity, separated by low activity periods, continue to appear. Since August 2013 the area is monitored by a local seismic network. The setup allows to accurately locate the epicenters and to determine source mechanisms for stronger events. The events are clustered and aligned along NE-SW direction and their mechanisms are very similar, indicating N-S strike slip faulting. This and the irregular pattern of activity suggest that this seismicity is triggered by the reservoir impoundment.

Key words: triggered and induced seismicity, Pieniny Klippen Belt.

1. INTRODUCTION

The system of two artificial lakes, Czorsztyn–Niedzica and Sromowce Wyżne, locates in the edge of Nowy Targ Valley in close neighborhood of

the Pieniny National Park and the Middle Ages castle in Niedzica. The history of building an artificial lake in the upper course of Dunajec River started at the beginning of 20th century. According to Bajorek and Zielińska (2010), first works concerning a use of energy of Carpathian rivers were due to engineers K. Pomianowski and T. Becker in 1904; another work was due to Prof. G. Narutowicz in 1911. The plan of building dams in Rożnów and Czorsztyn–Niedzica was developed in 1934. Because of economic reasons building of only the dam in Rożnów was undertaken and started in 1935. After the 2nd World War, the idea of building the dam in Czorsztyn–Niedzica was resumed in 1967. In 1971, preparatory works began, new roads were made, and environmental works started. The main construction works started in 1975. Impoundment of the reservoir began in 1995 and the hydropower plant Niedzica was made operational in 1997.

The dam in Niedzica on Dunajec River (Fig. 1) is the biggest earth dam in Poland. The dam length is 404 m, the width at the base is 300 m and the maximum height from the control tunnel foundation is 56 m. The maximum capacity of Czorsztyn reservoir is 234.5 million m³ with the average depth from about 20 m to about 50 m at the front of the dam. Sromowce Wyżne Lake is an auxiliary reservoir for Czorsztyn Lake (Fig. 2). The maximum capacity of Sromowce Lake is 6.7 million m³, and the maximum height of its embankment is 11 m. Both dams hold electric power generators. In Niedzica dam there are 2 reversible generators with the total power up to 92.75 MW. In Sromowce there are 4 small generators with total power of about 2 MW. The detailed technical information can be found on the website: www.zew-niedzica.com.pl. Main tasks of the reservoir system are: flood control, water supply and irrigation of surrounding areas, and electricity production (Bajorek and Zielińska 2010).



Fig. 1. The dam in Niedzica on Dunajec River. View from the weir of Sromowce Lake.

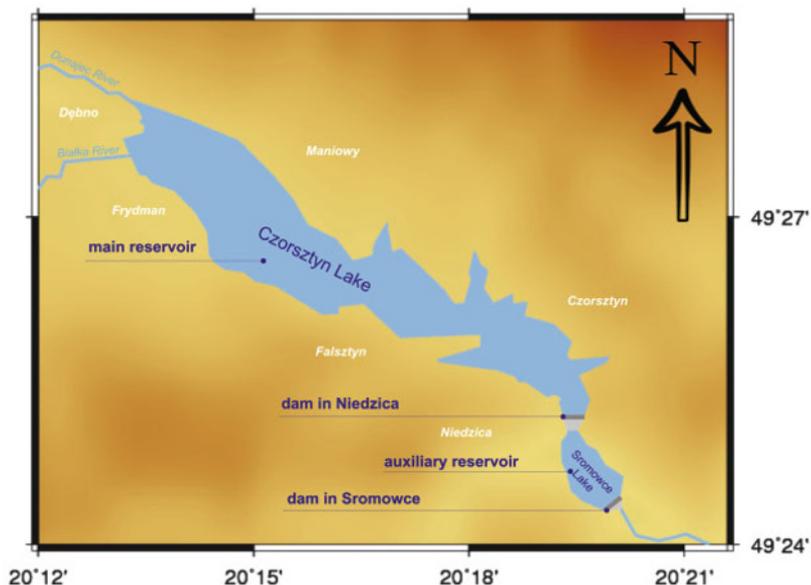


Fig. 2. Plan view of Czorsztyn Lake and Sromowce Lake.

Figure 3 shows a geological-tectonic map of the Czorsztyn Lake area with major tectonic borders and Miocene faults (Birkenmajer 2003). Czorsztyn Lake is situated in a border zone of two major Carpathian tectonic units: Inner and Outer Carpathians. The border zone is formed by Magura unit from the north and Pieniny Klippen Belt (PKB) from the south. Magura unit is a part of the Outer Carpathians. PKB separates the Inner Carpathians from the Outer Carpathians. According to Birkenmajer (2003), PKB extends for over 600 km as a north curved arc from the Tertiary Vienna depression up to Romania. PKB width varies from several kilometers, sometimes decreasing to just a few hundred meters. PKB has an extremely complicated geological-tectonic structure; it is cut by a series of young faults mostly of N-S strikes. PKB is mostly built of carbonate rocks (limestone and marlstone), shale, and claystone. Its narrow width and the significant length suggest correlation with a dislocation or a system of dislocations extending along the tectonic boundary of Western Carpathians.

The Niedzica dam locates at the eastern border of Orawa – Nowy Targ Basin. Compared to a negligible seismicity of Poland, the Orawa – Nowy Targ Basin area is known as the most seismically active zone in the Polish Carpathians, despite the fact that earthquakes in this region occur sporadically. The first information about earthquakes in this region dates back to 1717. Other earthquakes were observed in 1840, 1901, 1935, and 1966. Stronger

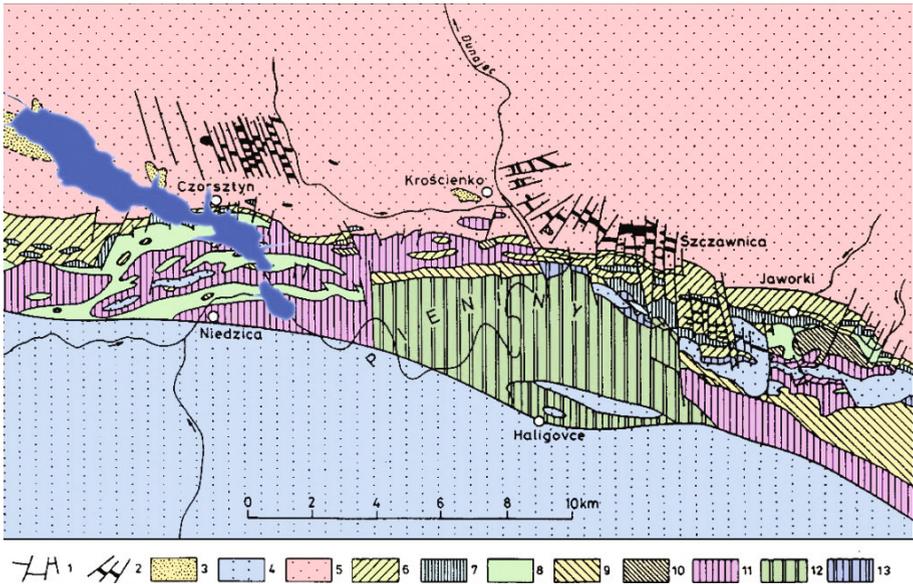


Fig. 3. Geological-tectonic map of the Czorsztyn Lake area (Birkenmajer 2003, map modified): 1 – Miocene fault system; 2 – Sarmatian andesite intrusions; 3 – fresh-water Neogene deposits; 4 – Inner-Carpathian Paleogene; 5 – Magura Paleogene; 6, 7 – Laramian Grajcarek Unit; 8–13 – tectonic units of Pieniny Klippen Belt: 8 – Czorsztyn, 9 – Czertezik, 10 – Niedzica, 11 – Branisko, 12 – Pieniny, and 13 – Haligowce.

seismic events occurred in the area of Krynica in 1992-1993 and in Podhale region in 1995 and in 30 November 2004 (Guterch *et al.* 2005). From 1998 to 2010 only 54 earthquakes were recorded in a direct area of Czorsztyn Lake, which makes, on average, about one event in three months. Unexpectedly, a significant increase of seismic activity was noticed in 2011. More than 60 seismic events were recorded close to Czorsztyn Lake in November 2011. Most of them were weak. The strongest earthquake, which occurred on 5 November 2011, had magnitude M_w 2.6. After November 2011, the activity returned to its previous low level. However, since January 2013 a gradual increase of activity was noted. In March 2013, 177 earthquakes occurred. The strongest event had magnitude M_w 3.5.

2. SENTINELS

The first instrumental measurements of the seismicity in Orawa–Nowy Targ Basin began in the 1960-ies with installing in Niedzica a seismological station of the Institute of Geophysics PAS. In a response to the increasing seismicity in the Czorsztyn Lake vicinity, observed since 2011, the Institute of

Geophysics PAS in cooperation with Niedzica Hydro Power Plants Company (Zespół Elektrowni Wodnych Niedzica S.A., ZEW Niedzica S.A.) launched local seismic network SENTINELS (SEismic Network for TRiggered and Natural Earthquake Location and Source determination) on 24 August 2013. The current SENTINELS network consists of seven short-period seismic stations and one broadband station. The short period stations comprise Lennartz 3D Lite seismometers and NDL recorders, the broadband station has STS-2 Kinematics seismometer and MK-6 recorder. NDL and MK6 recorders have been made by Institute of Geophysics PAS. More technical information can be found on the website: www.igf.edu.pl. The SENTINELS stations are emplaced around Czorsztyn Lake in facilities of ZEW Niedzica S.A., Krościenko administration of forestland and Nowy Targ municipality. Present locations of the SENTINELS stations are shown in Fig. 4. Before the SENTINELS network was set, locations of epicenters of

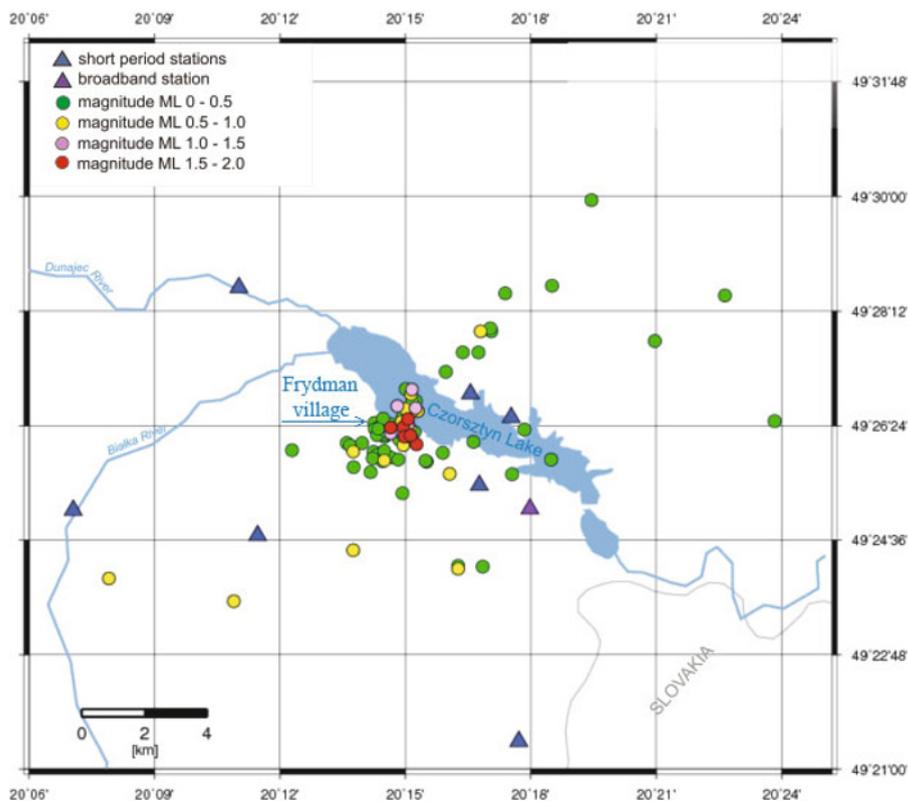


Fig. 4. Locations of earthquake recorded by SENTINELS network in the Czorsztyn Lake region from August 2013 to June 2014. Circles represent epicenters, triangles mark seismic stations.

only the strongest earthquakes had been determined. For this purpose, waveforms from the broadband station in Niedzica, from local Podhale stations which belong to the Institute of Geophysics PAS, and from far Czech and Slovakian stations were used. Due to an irregular distribution of all these stations as well as significant distances of some of them from the study area, the accuracy of epicenter location was not high. The SENTINELS network enables recording and more precise parameterization of weaker events. Earthquakes from local magnitude $M_L0.1$ are recorded, and the catalog is complete from $M_L0.5$. The SENTINELS network will be modernized in the framework of the project IS-EPOS “Digital Research Space of Induced Seismicity for EPOS Purpose”, co-financed from the funds of the European Regional Development Fund (ERDF) as part of the Operational Program Innovative Economy (www.is-epos.eu). The number of seismic stations will be increased to 10 and the data downloading interface will be upgraded. In order to reduce ambient noise, some of the stations will be relocated. Additionally, seismic profiling studies are foreseen in order to develop the first detailed velocity model of the subsurface in PKB region and to get an insight into local geotectonics from the viewpoint of the seismogenesis.

2.1 Seismicity of Czorsztyn Lake area

Digital records from broadband Niedzica station are available since May 1998. The numbers of earthquakes daily in the period from May 1998 to June 2014 are shown in Fig. 5. As mentioned in the introduction, until November 2011 earthquakes in the Czorsztyn Lake region were noticed only occasionally. From 1998 until November 2011, 60 seismic events were recorded; the strongest one occurred on 4 January 2005 and had magnitude $M_w2.0$.

The first period of increased seismic activity took place from November to December 2011. 82 seismic events were recorded in this period (63 in November and 19 in December). The strongest earthquake had magnitude $M_w2.9$. The second period of increased seismic activity occurred from July to September 2012. 29 earthquakes were recorded (12 in July, 14 in August, and 3 in September). The strongest event had magnitude $M_w2.4$ and took place in September. The strongest activity was observed in the third period between December 2012 and March 2013. 290 earthquakes were recorded, 13, 26, 77, and 174 in the consecutive months, respectively. The strongest earthquake so far in this region occurred on 1 March 2013. Its magnitude was $M_w3.5$. This event was felt by inhabitants of the area.

From the beginning of observations by means of SENTINELS network (24 August 2013) 101 earthquakes were recorded in the vicinity of Czorsztyn Lake. The greatest number of seismic events was noted in September

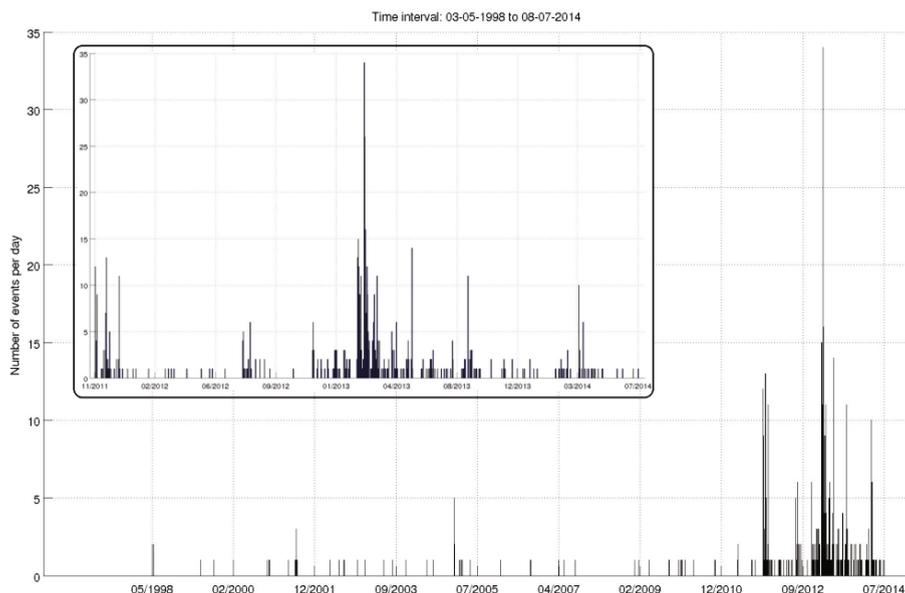


Fig. 5. Numbers of earthquakes daily from May 1998 to June 2014, with zoomed period from November 2011 to June 2014.

2013 (29 events) and in March 2014 (24 events). The strongest seismic event recorded by SENTINELS network so far took place on 1 April 2014 and had magnitude M_w 3.1. Epicenters of earthquakes are grouped at around Frydman village, which locates on the south-west shore of the lake. The epicenters are arranged in a belt extending from the north-east to the south-west (Fig. 4).

The average depth of earthquakes is approximately 6.5 km. The hypocenter of the deepest earthquake was at about 9 km below the surface and the hypocenter of the shallowest one was at about 2 km below the surface. For all recorded events, local magnitude M_L is determined; for stronger events, also moment magnitude M_w is calculated. Magnitude completeness level for SENTINELS observations is 1.4 for moment magnitude M_w and 0.5 for local magnitude M_L .

It was possible to evaluate focal mechanism for 16 strongest events. The calculations were done by means of the program FOCI (Kwiatek 2011), which implements a seismic moment tensor inversion in time domain (Wiejacz 1992). The resultant mechanisms are shown in Fig. 6. All mechanisms are very similar: strike slip faulting with a very small dip slip component. The result suggests a common genesis of these earthquakes. The mechanism of the earthquake located by the south-west edge of the area is different however, the result is uncertain because the event was located nearly outside the network.

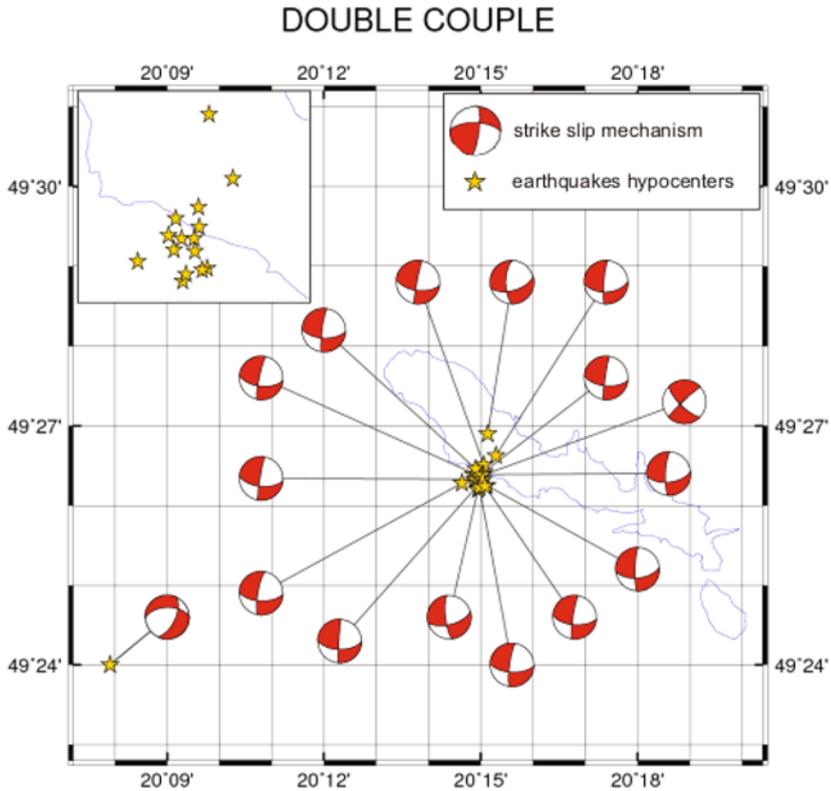


Fig. 6. Mechanisms of earthquake from Czorsztyn Lake region.

3. DISCUSSION AND CONCLUSIONS

The significantly increased seismic activity since 2011 is highly irregular. In some periods, the earthquake productivity is much higher than in the others. Both, the increase of seismic activity and its time irregularity suggest a connection of this seismicity with the impoundment of Czorsztyn Lake.

The phenomenon of seismicity induced or triggered by impoundments of water reservoirs is well known worldwide. Triggering is caused by a combined effect of two factors: (i) additional load of water accumulated in the reservoir, imposed on rock formations, which changes an initial stress field, and (ii) an increase of pore pressure in fault zones nearby the reservoir as a result of contraction of pores due to the changed stress field or by a fluid diffusion. In general, the second mechanism, which weakens the faults, is more important. The first factor acts immediately, results of the second can

appear after some delay time. However, a necessary condition for earthquake triggering by water reservoirs is the existence of a tectonically loaded fault in the reservoir zone.

The best known example of the reservoir triggered seismicity is the case of Koyna reservoir in India, when seismicity started shortly after the impoundment in 1962 (Gupta 2002). Since the impoundment, more than 100 thousand earthquakes from magnitude $M \geq 0$ occurred. Ten earthquakes had magnitude $M \geq 5.0$. The strongest event took place on 10 December 1967 and had magnitude $M = 6.3$. Koyna is not an isolated example. For instance, the Institute of Geophysics PAS in cooperation with the Institute of Geophysics, Vietnam Academy of Science and Technology, is monitoring earthquakes triggered by the impoundment of Song Tranh 2 dam reservoir in central Vietnam (Wiszniewski *et al.* 2014). Project details are available on the website www.is-epos.eu.

Impoundment of Czorsztyn Lake was completed in 1997, and the increase of the seismic activity has appeared 14 years later. Delayed seismic responses to reservoir impoundment were noticed on many instances around the world. 15 years after the impoundment of Aswan reservoir in Egypt an earthquake $M5.4$ was recorded in 1981 (Mekkawi *et al.* 2004). Around the reservoir Carno do Cajurain, Brasil, the delay was 18 years.

Czorsztyn Lake is shallow and the effect of the water load at depth is negligible. It is likely that the seismicity is a response to fault weakening due to diffusion effects. Regardless its cause, the earthquake triggering in this region would indicate existence of a fault or fault system being close to failure, which is quite surprising taking into account the rare tectonic seismicity occurrence in the area in both historical and instrumental periods.

Mutually close locations and similar focal mechanisms of earthquakes from Czorsztyn Lake region suggest that they are fracturing of the same segment of a strike slip fault. Seismic mechanism indicates that this fault strikes NS (Fig. 6.) This is not consistent with an alignment of epicenter locations (Fig. 4), which suggests NE-SW strike of the fault. The above-mentioned detailed seismic profiling studies of the area, planned in the framework of IS-EPOS project, will help to clarify this problem.

Acknowledgments. This work was done in the framework of the project IS-EPOS “Digital Research Space of Induced Seismicity for EPOS Purposes” (POIG.02.03.00-14-090/13-00), funded by the National Centre for Research and Development in the Operational Program Innovative Economy in the years 2013-2015.

References

- Birkenmajer, K. (2003), Geological problems under study in the Pieniny Mountains, *Pieniny Przyroda Człowiek* **8**, 33-40 (in Polish with English abstract).
- Bajorek, L., and T. Zielińska (2010), Niedzica Hydro Powerplants joint stock company– reservoir creation and Company activity. **In:** R. Soja, S. Knutelski, J. Bodziarczyk (eds.), *Pieniny – Zapora – Zmiany*, Monografie Pienińskie, Vol. 2, 23-35 (in Polish with English abstract).
- Gupta, H.K. (2002), A review of recent studies of triggered earthquakes by artificial water reservoirs with special emphasis on earthquakes in Koyna, India, *Earth Sci. Rev.* **58**, 3-4, 279-310, DOI: 10.1016/S0012-8252(02)00063-6.
- Guterch, B., H. Lewandowska-Marciniak, and J. Niewiadomski (2005), Earthquakes recorded in Poland along the Pieniny Klippen Belt, Western Carpathians, *Acta Geophys. Pol.* **53**, 1, 27-45.
- Kwiatek, G. (2011), FOCI—software for calculation of the seismic moment tensor and source parameters in mining environment, <http://www.induced.pl/projekty/foci>.
- Mekkawi, M., J.-R. Grasso, and P.-A. Schnegg (2004), A long-lasting relaxation of seismicity at Aswan reservoir, Egypt, 1982-2001, *Bull. Seismol. Soc. Am.* **94**, 2, 479-492, DOI: 10.1785/0120030067.
- Wiejacz, P. (1992), Calculation of seismic moment tensor for mine tremors from the Legnica–Głogów Copper Basin, *Acta Geophys. Pol.* **40**, 2, 103-122.
- Wiszniowski, J., N. Van Giang, B. Plesiewicz, G. Lizurek, D.Q. Van, L.Q. Khoi, and S. Lasocki (2014), Preliminary results of anthropogenic seismicity monitoring in the region of Song Tranh 2 reservoir, Central Vietnam, *Acta Geophys.* **63**, 3, 843-862, DOI: 10.1515/acgeo-2015-0021.

Received 24 November 2014

Received in revised form 24 March 2015

Accepted 26 March 2015



Shallow Water Turbulent Surface Wave Striking an Adverse Slope

Sujit K. BOSE

S.N. Bose National Centre for Basic Sciences, Kolkata 700064, India
e-mail: sujitkbose@yahoo.com

Abstract

The problem of a sinusoidal wave crest striking an adverse slope due to gradual elevation of the bed is relevant for coastal sea waves. Turbulence based RANS equations are used here under turbulence closure assumptions. Depth-averaging the equations of continuity and momentum, yield two differential equations for the surface elevation and the average forward velocity. After nondimensionalization, the two equations are converted in terms of elevation over the inclined bed and the discharge, where the latter is a function of the former satisfying a first order differential equation, while the elevation is given by a first order evolution equation which is treated by Lax-Wendroff discretization. Starting initially with a single sinusoidal crest, it is shown that as time progresses, the crest leans forwards, causing a jump in the crest upfront resulting in its roll over as a jet. Three cases show that jump becomes more prominent with increasing bed inclination.

Key words: surface waves, shallow water waves, inclined bed, turbulence, unsteady flow, breaking waves.

1. INTRODUCTION

Free surface waves on a layer of water, incident upon an adverse slope, has been of theoretical interest for several decades due to recurring occurrence of

tsunamis and other coastal water wave phenomena. Assuming inviscid irrotational motion governed by Laplace's equation, detailed theoretical analyses are given in Stoker (1957). Inclusion of viscosity and bed friction using Navier-Stokes equations together with the laws of Chezy, Manning or Strickler for bed friction are treated in Mader (2004). Following the devastating Indonesian tsunami in the year 2004, Kundu (2007), has presented different aspects of water wave propagation. Unsteady waves climbing an adversely sloping bed are essentially turbulent in nature. In the hydraulic engineering literature, steady and unsteady open channel flows are extensively studied modelled by an equation of continuity and a St. Venant momentum equation (Chow 1959). Strelkoff (1969), Yen (1973), and Basco (1987) also deal with integration of these equations. Bose and Dey (2007) on the other hand, give a systematic investigation of the two dimensional unsteady curvilinear free surface flows, based on the Reynolds Averaged Navier Stokes (RANS) equations, using reasonable turbulence closure assumptions. By this procedure, they obtained explicit equations for the depth-averaged equation of continuity and a nonlinear PDE for the momentum equation that generalises the St. Venant equation. Subsequently, Bose and Dey (2009) generalised the method to treat the case of undulating erodible bed to provide a theory of dune and antidune propagation.

The shallow water equation and the St. Venant equations require numerical treatment. Such treatments are described in Abbott (1979), Cunge *et al.* (1980), Benque *et al.* (1982), and Mader (2004). Fennema and Chauhry (1990) adopted the McCormack scheme for the solution of the two dimensional shallow water equations. Garcia-Navarro *et al.* (1992, 1995), on the other hand, developed upwind TVD scheme for the one and two dimensional shallow water equations. Finite difference semi-implicit scheme was developed by Casulli and Cheng (1992) and Casulli and Stelling (1998), while an implicit scheme was put forward by Namin *et al.* (2001). Chen (2003) developed a novel free-surface correction method for two dimensional flows. Xing and Shu (2005) have designed a new high order finite difference WENO scheme for these equations. In finite element methods, Katopodes (1984) developed a dissipative Galerkin scheme, while Quecedo and Pastor (2003) treated the case of one dimensional equations over inclined and curved beds.

With the objective of studying the propagation of tidal bores upstream of estuarine rivers, Bose and Dey (2013) presented a theory of a surging flow over an adverse slope. The theory is based on the RANS equations following Bose and Dey (2007, 2009) for the predominantly turbulent motion. Integration over the depth leads to continuity and momentum equations for the surface elevation η and average forward velocity U . The system of differential equations is numerically solved by replacing the time derivatives by second order finite difference formulae and integrating the ordinary differential equations in the forward space variable by the Runge-Kutta method. In this paper, the evolution

of a travelling sinusoidal wave crest striking an adverse slope is studied by the same methodology. Here though the full equations can be developed as in Bose and Dey (2013), for numerical computations of the evolving wave form, the contribution of the instantaneous vertical acceleration is neglected in comparison to the convective vertical acceleration. This results in reducing the size of the new momentum equation. The bed friction term is also neglected, because of its smallness in the problem. A new numerical technique is adopted here. Introducing the nondimensional elevation of the free surface above the inclined bed to be ζ and q to be the nondimensional discharge, the continuity and the momentum equations are expressed in terms of ζ and q . It could then be shown that q is a function of ζ , *i.e.* $q = F(\zeta)$ where F satisfies a first order ODE obtained from the momentum equation. Solution of this equation coupled with the continuity equation, which becomes a one dimensional evolution equation that can be treated by a Lax-Wendroff type scheme. The numerical solution is graphically presented at certain time steps, which show the crest to be leaning forward, resulting in sharp rise of the next crest. The rising crest rolls over as a jet due to forward velocity.

2. BASIC EQUATIONS OF TURBULENT SHALLOW WATER WAVE MOTION OVER INCLINED BED

A free sinusoidal wave propagating on a layer of water of uniform depth η_0 approaches the base of a rising plane bed at an angle β , striking it normally. In the definition sketch, Fig. 1, the origin of coordinates is taken at the base of the inclined bed and the x-axis is taken horizontally, while the y-axis is taken vertically upwards. The equation of the bed is therefore $h = x \tan \beta$, $x \geq 0$ and $h = 0$, $x < 0$.

The quantities of primary interest in the evolving waves are the surface elevation η above the x-axis and the depth-averaged streamwise flow velocity U .

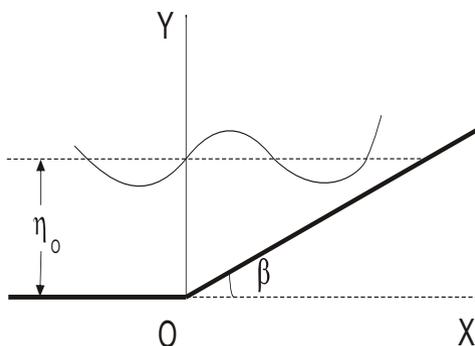


Fig. 1. Definition sketch of a wave striking an inclined bed.

Since the generated wave motion is turbulent, the appropriate procedure is to base the theory on RANS equations as in Bose and Dey (2009) for flows on undulating bed. In the treatment of such curvilinear turbulent flow, the normal acceleration at a point is assumed to be that due to the convective part only, neglecting the contribution from the instantaneous vertical acceleration. Though the contribution of the term can be retained (Bose and Dey 2013), its negligence also leads to considerable simplification in the governing momentum equation. The slopes $|\partial\eta/\partial x|$ and $|\partial h/\partial x|$ are however considered finite, because the smallness of the former slope is likely to be broken during the heaving motion and formation of breaking waves as the incident wave encounters the adverse slope. Thus, if (u, v) be the instantaneous velocity components in the two dimensional flow, the two components can be split by Reynolds decomposition into the time-averaged part (\bar{u}, \bar{v}) and the fluctuating part (u', v') in the forms

$$u(x, y, t) = \bar{u}(x, y, t) + u'(x, y, t), \quad v(x, y, t) = \bar{v}(x, y, t) + v'(x, y, t) \quad (1)$$

where t is the time. The time-averaging of the equation of continuity yields the equations

$$\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} = 0, \quad \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} = 0 \quad (2)$$

while the time-averaging of the two-dimensional Navier-Stokes equations yield the boundary layer approximated RANS equations

$$\frac{\partial \bar{u}}{\partial t} + \bar{u} \frac{\partial \bar{u}}{\partial x} + \bar{v} \frac{\partial \bar{u}}{\partial y} = -\frac{1}{\rho} \frac{\partial \bar{p}}{\partial x} + \frac{1}{\rho} \frac{\partial \tau}{\partial y} + \nu \frac{\partial^2 \bar{u}}{\partial y^2} - \frac{\partial}{\partial x}(\overline{u'^2}) \quad (3a)$$

$$\frac{\partial \bar{v}}{\partial t} + \bar{u} \frac{\partial \bar{v}}{\partial x} + \bar{v} \frac{\partial \bar{v}}{\partial y} = -\frac{1}{\rho} \frac{\partial \bar{p}}{\partial y} + \frac{1}{\rho} \frac{\partial \tau}{\partial x} + \nu \frac{\partial^2 \bar{v}}{\partial y^2} - \frac{\partial}{\partial y}(\overline{v'^2}) - g \quad (3b)$$

where $\bar{p}(x, y, t)$ is the time-averaged hydrostatic pressure, $\tau(x, y, t)$ the Reynolds shear stress, that is $-\rho \overline{u'v'}$, ν the kinematic coefficient of viscosity, and g the acceleration due to gravity. Equations (2) and (3a,b) form an underdetermined system in five variables, namely, \bar{u}, \bar{v}, u', v' and \bar{p} , even if we assume that the time averaged Reynolds stress components $\overline{u'^2}, \overline{v'^2}$ and τ can somehow be separately computed. Consequently, additional assumptions are required, on the basis of the characteristics of the flow.

Firstly, the free surface flow is of shear type, in which it is assumed that the streamwise gradients of the Reynolds stress are negligible, that is,

$$\frac{\partial \tau}{\partial x} \approx 0, \quad \frac{\partial}{\partial x}(\overline{u'^2}) \approx 0, \quad \frac{\partial}{\partial x}(\overline{v'^2}) \approx 0 \quad (4)$$

secondly, it is assumed that the contribution of the magnitude of the viscous stress is much smaller than that of the Reynolds stress in the momentum equa-

tions (4a,b) in terms of a slowly varying function $(y - h)^{1/p}$, $p \gg 1$. Consequently from Eq. 4a, it follows that (Bose and Dey 2007)

$$\bar{u} = \frac{1+p}{p} U(x, t) \left(\frac{y-h}{\eta-h} \right)^{1/p} \tag{5}$$

where $\eta(x, t)$ is the free surface elevation at $(x, 0)$ and $U(x, t) = (\eta - h)^{-1} \int_h^\eta \bar{u} dy =$ depth-averaged velocity at $(x, 0)$. In Eq. 3b, on the other hand, the turbulent and the viscous shear stress terms drop out on account of Eq. 4. By a similar argument, an approximate expression for $\partial\bar{u}/\partial x$ can be constructed (Bose and Dey 2013), and Eq. 2 then yields

$$\bar{v} = -(\eta - h) \frac{\partial U}{\partial x} \left(\frac{y-h}{\eta-h} \right)^{(1+p)/p} \tag{6}$$

The above equation implies that if U diminishes with x , \bar{v} becomes positive (upwards), and if U increases with x , \bar{v} becomes negative (downwards).

The continuity and the momentum equations 2 and 3a,b with \bar{u} and \bar{v} given by Eqs. 5 and 7 are then averaged over the depth from h to η . The averaging of Eq. 2 leads to the depth-averaged equation of continuity

$$\frac{\partial}{\partial t}(\eta - h) + \frac{\partial}{\partial x}[(\eta - h)U] = 0 \tag{7}$$

Bose and Dey 2009), and for the depth-average of the forward momentum equation 4a, it can be shown that (Bose and Dey 2009)

$$\int_h^\eta \left(\frac{\partial \bar{u}}{\partial t} + \bar{u} \frac{\partial \bar{u}}{\partial x} + \bar{v} \frac{\partial \bar{u}}{\partial y} \right) dy = \frac{\partial}{\partial t}[(\eta - h)U] + \alpha \frac{\partial}{\partial x}[(\eta - h)U^2] \tag{8}$$

where $\alpha = (1 + p)^2/[p(2 + p)]$. The contribution of the forward pressure gradient term in Eq. 3a is evaluated by considering Eq. 3b. The convective term in the latter equation yields by using Eq. 2

$$\bar{u} \frac{\partial \bar{v}}{\partial x} + \bar{v} \frac{\partial \bar{v}}{\partial y} = (\bar{u}^2 \sec^3 \psi) \kappa = \bar{u}^2 \frac{\partial^2 y}{\partial x^2} \tag{9}$$

where $\tan \psi = \bar{v}/\bar{u}$ is the slope of the streamline of the time-averaged flow through the point $P(x, y)$ and $\partial^2 y/\partial x^2$ is proportional to the curvature κ of the stream line at this point. Following Boussinesq, it is assumed that the second derivative varies linearly from the bed level to the free surface, that is

$$\frac{\partial^2 y}{\partial x^2} = \frac{y-h}{\eta-h} \frac{\partial^2 \eta}{\partial x^2} \tag{10}$$

Inserting Eqs. 10 with 11 in Eq. 3b and neglecting the instantaneous vertical acceleration $\partial \bar{v} / \partial t$ of the curvilinear flow, an integration with respect to y from η to y yields

$$\frac{\bar{p}}{\rho} = \frac{\bar{p}_0}{\rho} - g(y - \eta) - \frac{1}{2} \left(\frac{1 + p}{p} \right) U^2(\eta - h) \frac{\partial^2 \eta}{\partial x^2} \left[\left(\frac{y - h}{\eta - h} \right)^{2(1+p)/p} - 1 \right] - \bar{v}^2 \quad (11)$$

where $\bar{p} = \bar{p}_0$ at the free surface $y = \eta$. It therefore follows that

$$\frac{1}{\rho} \int_h^\eta \frac{\partial \bar{p}}{\partial x} dy = g(\eta - h) \frac{\partial \eta}{\partial x} + \gamma \frac{\partial}{\partial x} \left[U^2(\eta - h)^2 \frac{\partial^2 \eta}{\partial x^2} \right] + \frac{1}{2} \left(\frac{1 + p}{p} \right) U(\eta - h) \frac{\partial^2 \eta}{\partial x^2} \frac{dh}{dx} \quad (12)$$

where $\gamma = (1 + p)^2 / [p(2 + 3p)]$. The contribution of the viscous and Reynolds shear stresses in Eq. 3a is

$$\frac{1}{\rho} \int_h^\eta \frac{\partial \tau}{\partial y} dy + \nu \int_h^\eta \frac{\partial^2 \bar{u}}{\partial y^2} dy = -\nu \frac{\partial \bar{u}}{\partial y} \Big|_{y=h} = -\frac{\tau_0}{\rho} \quad (13)$$

as the viscous and the Reynolds shear stresses vanish at the free surface and the latter vanishes at the bed level as well. τ_0 in the above equation represents the bed shear stress. Being an additional unknown parameter, it can only be represented by a convenient empirical formula, such as that of Manning (Chow 1959). The effect of this stress, however contributes insignificantly to the surface waves at the top, as was shown by Bose and Dey (2014) in the case of surface gravity waves. Consequently it is dropped in the present analysis. The depth-averaged forward momentum equation, under the approximations 4 therefore becomes

$$(\eta - h) \frac{\partial U}{\partial t} + (2\alpha - 1)(\eta - h)U \frac{\partial U}{\partial x} + (\alpha - 1)U^2 \left(\frac{\partial \eta}{\partial x} - \frac{dh}{dx} \right) + \gamma \frac{\partial}{\partial x} \left[U^2(\eta - h)^2 \frac{\partial^2 \eta}{\partial x^2} \right] + \frac{1}{2} \left(\frac{1 + p}{p} \right) U^2(\eta - h) \frac{\partial^2 \eta}{\partial x^2} \frac{dh}{dx} + g(\eta - h) \frac{\partial \eta}{\partial x} = 0. \quad (14)$$

In the above equation, p is approximately taken as 7 – a value that holds exactly for flows on plane bed (Schlichting and Gersten 2000). This yields $\alpha \approx 1$ and $\gamma \approx 2/5$. Also for motion over the inclined bed $dh/dx = \tan \beta$ for $x \geq 0$. Eqs. 8 and 15 constitute the required continuity and momentum equations that generalise the well known St. Venant equations.

3. WAVE MOTION OVER THE UPWARD INCLINED BED

When a horizontally travelling free surface wave strikes an adversely sloping bed, a heaving motion with the formation of breaking waves follow as a result

of tapering of the flow section. In order to numerically study the motion, the following nondimensional variables are introduced:

$$\hat{\eta} = \frac{\eta}{\eta_0}, \quad \hat{x} = \frac{x}{\eta_0}, \quad \hat{h} = \frac{h}{\eta_0}, \quad \hat{U} = \frac{U}{\sqrt{g\eta_0}}, \quad \hat{t} = t\sqrt{\frac{g}{\eta_0}}. \tag{15}$$

It is assumed that initially a sinusoidal wave form with a trough above the origin of the inclination is incident on the inclined bed:

$$\hat{\eta} = 1 + \hat{a} \sin \hat{k}(\hat{x} - \hat{c}\hat{t}) \tag{16}$$

where \hat{a} , \hat{k} , \hat{c} are, respectively, the nondimensional amplitude, wave number and velocity of propagation of the wave. For the progress of the wave, the equation of continuity 8 is then written in the form

$$\frac{\partial \zeta}{\partial \hat{t}} + \frac{\partial q}{\partial \hat{x}} = 0 \tag{17}$$

where, $\zeta = \hat{\eta} - \hat{h}$ and $q = \zeta \hat{U}$, respectively, represent the elevation above the inclined bed and the discharge at the cross-section at \hat{x} . The momentum equation 15, with $\alpha = 1$, $\gamma = 2/5$ then becomes

$$\frac{\partial \hat{U}}{\partial \hat{t}} + \hat{U} \frac{\partial \hat{U}}{\partial \hat{x}} + \frac{2}{5} \zeta \hat{U}^2 \frac{\partial^3 \zeta}{\partial \hat{x}^3} + \frac{4}{5} \hat{U} \frac{\partial}{\partial \hat{x}} (\zeta \hat{U}) \frac{\partial^2 \zeta}{\partial \hat{x}^2} + \hat{U}^2 \tan \beta \frac{\partial^2 \zeta}{\partial \hat{x}^2} + \frac{\partial \zeta}{\partial \hat{x}} + \tan \beta = 0. \tag{18}$$

From the above equation, \hat{U} can be eliminated by setting $\hat{U} = q/\zeta$.

The solution of Eqs. 17 and 18 is of the form $\zeta = \zeta(\hat{x}, \hat{t})$, $q = q(\hat{x}, \hat{t})$. Assuming the existence of the inverse of this function pair, $\hat{x} = \hat{x}(\zeta, q)$, $\hat{t} = \hat{t}(\zeta, q)$. Hence, $q = q[\hat{x}(\zeta, q), \hat{t}(\zeta, q)]$, whose solution if it exists is of the form

$$q = F(\zeta) \tag{19}$$

which means that the discharge at a position \hat{x} at given time \hat{t} depends solely on the elevation at that point and at that time. This physically plausible assumption is sometimes assumed in surface wave propagation theory (Stoker 1957). Substitution of Eq. 19 in Eq. 17 yields

$$\frac{\partial \zeta}{\partial \hat{t}} + F'(\zeta) \frac{\partial \zeta}{\partial \hat{x}} = 0. \tag{20}$$

Similarly since

$$\frac{\partial \hat{U}}{\partial \hat{t}} = - \left(-F'^2 + \frac{FF'}{\zeta} \right) \frac{\zeta_{\hat{x}}}{\zeta} \tag{21a}$$

$$\frac{\partial \hat{U}}{\partial \hat{x}} = \left(F' - \frac{F}{\zeta} \right) \frac{\zeta_{\hat{x}}}{\zeta} \tag{21b}$$

where the prime denotes differentiation with respect to ζ and the subscript \hat{x} denotes partial differentiation with respect to \hat{x} , Eq. 18 then becomes

$$\zeta^2 F'^2 - 2\zeta \left(1 + \frac{2}{5} \zeta \phi_1 \right) F F' + \left(1 - \frac{2}{5} \zeta^2 \phi_3 - \frac{4}{7} \tan \beta \zeta \phi_2 \right) F^2 - \zeta^3 \left(1 + \frac{\tan \beta}{\phi} \right) = 0 \quad (22)$$

where $\zeta_{\hat{x}} =: \phi$, $\zeta_{\hat{x}\hat{x}} =: \phi_1$, $\phi_1/\phi =: \phi_2$, and $\zeta_{\hat{x}\hat{x}\hat{x}}/\zeta_{\hat{x}} =: \phi_3$. As in the case of q , ϕ , ϕ_1 , ϕ_2 , and ϕ_3 can be argued to be functions of ζ . The solution of the quadratic equation yields the differential equation for F as

$$\zeta F' = \left(1 + \frac{2}{5} \zeta \phi_1 \right) F + \left[4\zeta \left\{ \left(\frac{1}{5} \phi_1 + \frac{1}{7} \tan \beta \phi_2 \right) + \frac{\zeta}{10} \left(\phi_3 + \frac{2}{5} \phi_1^2 \right) \right\} F^2 + \zeta^3 \left(1 + \frac{\tan \beta}{\phi} \right) \right]^{1/2} \quad (23)$$

where the positive sign of the square root is taken to ensure increasing values of the discharge q for increasing values of the elevation ζ . Integration of the ordinary differential equation 23 yields the discharge function F for different elevation profiles ζ above the inclined bed. The elevation ζ in turn is governed by the evolution equation 20.

The evolution of the wave form 16 with time, as solution of Eqs. 20 and 23 is sought numerically for $\hat{a} = 0.2$, $\hat{c} = 0.2$, and $\hat{k} = \pi$. The computation is started at time $\hat{t} = 0$ with the profile 16, discretised by points at subinterval length $h = 0.1$ over a wave cycle. The derivative functions ϕ , ϕ_1 , ϕ_2 , and ϕ_3 are computed by second order finite difference formulae. Equation 23 is then integrated by the fourth order Runge-Kutta formula. The latter procedure requires values of the elevation ζ at intermediate points other than the points of discretization. This is accomplished by spline interpolation (Bose 2009). The integration is initialised by taking $F = c = 0.2$ for $\zeta = 1$, as required by Eq. 20. The values of F are thus computed for the discrete data set of ζ at subintervals of 0.01 to cover all the elevations over the wave cycle. The derivatives of F , viz. F' and F'' , are then computed by second order finite difference formulae over the discrete values of ζ .

The temporal development of motion is given by the quasi-linear equation 20. For this equation, a second order Lax-Wendroff type scheme can be developed:

$$\zeta_m^{n+1} = \zeta_m^n - \frac{r}{2} F'(\zeta_m^n) (\zeta_{m+1}^n - \zeta_{m-1}^n) + \frac{r^2}{2} F''(\zeta_m^n) \left[\frac{1}{2} F''(\zeta_m^n) (\zeta_{m+1}^n - \zeta_m^n)^2 + F'(\zeta_m^n) (\zeta_{m+1}^n - 2\zeta_m^n + \zeta_{m-1}^n) \right] \quad (24)$$

where ζ_m^n represents the elevation above the inclined bed at a point $x = mh$, ($m = 0, 1, 2, 3, \dots$) at time $t = nk$, ($n = 0, 1, 2, 3, \dots$) such that $r = k/h$. A value of $r = 1/2$ is chosen as is the usual case with the Lax-Wendroff method. Equation 24 yields the elevations at the next time step that requires the values of $F'(\zeta_m^n)$ and $F''(\zeta_m^n)$. These are obtained by spline interpolation of the earlier computed values of F' and F'' . The iterations are carried over a number of cycles until a breakdown is indicated by negative argument of the square root in Eq. 23.

In the above computation, the values of the bed inclination β are typically taken as 1° , 3° , and 5° . The results are shown in Figs. 2, 3, and 4 as the wave progresses at different times. In these figures the abscissa represents \hat{x} and the ordinate represents $\hat{\eta}$. Typically the crest of the wave progressively leans forward, affecting the crest in front by a sharp rise in the elevation. Evidently, such sharply rising crests – that possesses a forward velocity – leads to breaking waves. The phenomenon becomes more pronounced as the bed elevation β increases. The case of $\beta = 0^\circ$ (level bed) was also tried and a trend similar to

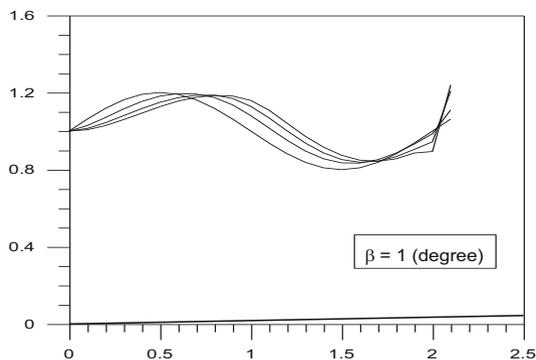


Fig. 2. Surface elevation at times 0.00, 0.05, 0.10, and 0.15.

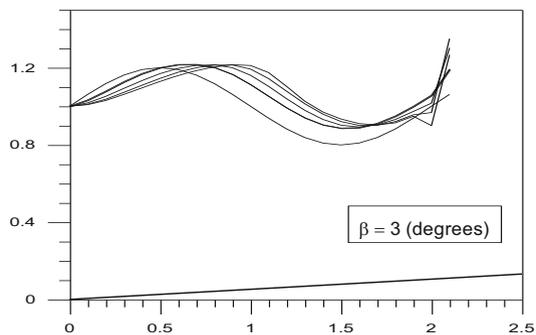


Fig. 3. Surface elevation at times 0.00, 0.05, 0.10, 0.15, and 0.20.

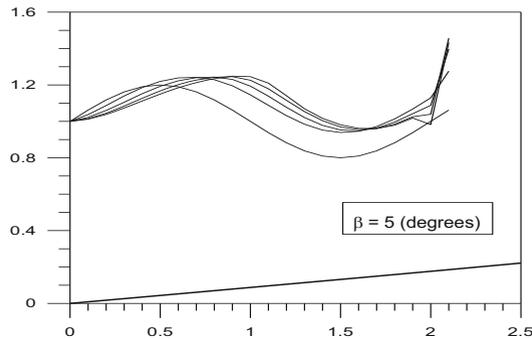


Fig. 4. Surface elevation at times 0.00, 0.05, 0.10, 0.15, and 0.20.

that of $\beta = 1^\circ$ was noticed. This shows that sinusoidal wave on a horizontal bed is essentially unstable. This feature is in accordance with the recent finding of Bose and Dey (2014) that periodic waves have a somewhat different form that are akin to “waves of permanent shape” (Lamb 1932).

4. CONCLUSIONS

The purpose of this theoretical study is to examine the progression of a sinusoidal wave crest on a sheet of water as it meets an adversely inclined bed. The angle of inclination of the bed is supposed to be small, like that in the case of sea coasts. The motion in practice is turbulent in such cases, and as such the theory is based on the two dimensional Reynolds Averaged Navier-Stokes (RANS) equations. The RANS equations of conservation of mass and momentum are closed together with some assumptions appropriate for the flow. The equations thus lead to the $1/p$ th power law of variation in the vertical y direction for the forward time-averaged velocity \bar{u} and an expression for the time-averaged pressure \bar{p} containing the gravitational hydrostatic part and an additional term due to vertical convective acceleration under the power law. The instantaneous vertical acceleration is neglected in comparison to the dominant convective acceleration in the progressing heaving motion. This results in considerable simplification of the final momentum equation as well. The two conservation equations of mass and forward momentum are depth-averaged to yield the required two equations in terms of the surface elevation η and the depth-averaged forward velocity U . Converting the two equations in nondimensional form, the equation pair is numerically treated for a sinusoidal wave crested trough, by eliminating the nondimensional U in favour of the nondimensional discharge q . Arguing that q can be considered a function of the nondimensional elevation ζ , viz. $q = F(\zeta)$, the nonlinear momentum equation gets converted into a first order ordinary differential equation for $F(\zeta)$.

The equation is treated by the fourth order Runge-Kutta method for appropriate initial conditions of the problem. The mass conservation equation, on the other hand, gets converted into a first order time evolutionary partial differential equation that can be treated by a method of Lax-Wendroff type, using the solution for $F(\zeta)$. Using this coupled technique, the numerical solution shows that as the sinusoidal wave crest advances, it leans forward leading to steep rise of the crest in front after some time, resulting in its breaking due to the forward momentum. This feature becomes more prominent as the angle of elevation of the inclined bed increases. A numerical test for a flat bed was also carried out and it also exhibited this breaking unstable phenomenon for the sinusoidal wave, in agreement with the finding of Bose and Dey (2014) that gravity waves on turbulent channels have shape other than pure sinusoids. The numerical treatment presented in this paper is new, different from the one adopted by Bose and Dey (2013) for the case of surging flow up an incline as in the case of tidal bores.

Acknowledgment. The author is thankful to the S.N. Bose National Center for Basic Sciences, Kolkata for providing necessary facilities for undertaking this research.

References

- Abbott, H.B. (1979), *Computational Hydraulics: Elements of the Theory of Free Surface Flows*, Pitman Publ., London.
- Basco, D.R. (1987), Computation of rapidly varied, unsteady, free-surface flow, Water-Resources Investigations Report 83-4284, U.S. Geological Survey, Reston, Virginia, U.S.A.
- Benqué, J.P., J.A. Cunge, J. Feuillet, A. Hauguel, and F.M. Holly Jr. (1982), New method for tidal current computation, *J. Waterw. Port Coast. Ocean Div. ASCE* **108**, 3, 396-417.
- Bose, S.K. (2009), *Numeric Programming in Fortran*, Narosa, New Delhi.
- Bose, S.K., and S. Dey (2007), Curvilinear flow profiles based on Reynolds averaging, *J. Hydraul. Eng.* **133**, 9, 1074-1079, DOI:10.1061/(ASCE)0733-9429(2007)133:9(1074).
- Bose, S.K., and S. Dey (2009), Reynolds averaged theory of turbulent shear flows over undulating beds and formation of sand waves, *Phys. Rev. E* **80**, 3, 036304, DOI:10.1103/PhysRevE.80.036304.
- Bose, S.K., and S. Dey (2013), Turbulent unsteady flow profiles over an adverse slope, *Acta Geophys.* **61**, 1, 84-97, DOI:10.2478/s11600-012-0080-2.
- Bose, S.K., and S. Dey (2014), Gravity waves on turbulent shear flow: Reynolds averaged approach, *J. Hydraul. Eng.* **140**, 3, 340-346, DOI:10.1061/(ASCE)HY.1943-7900.0000820.

- Casulli, V., and R.T. Cheng (1992), Semi-implicit finite difference methods for three-dimensional shallow water flow, *Int. J. Num. Meth. Fluids* **15**, 6, 629-648, DOI:10.1002/fld.1650150602.
- Casulli, V., and G.S. Stelling (1998), Numerical simulation of 3D quasi-hydrostatic, free-surface flows, *J. Hydraul. Eng.* **124**, 7, 678-686, DOI:10.1061/(ASCE)0733-9429(1998)124:7(678).
- Chen, X.J. (2003), A free-surface correction method for simulating shallow water flows, *J. Comput. Phys.* **189**, 2, 557-578, DOI:10.1016/S0021-9991(03)00234-1.
- Chow, V.T. (1959), *Open Channel Hydraulics*, McGraw-Hill, New York.
- Cunge, J.A., F.M. Holly, and A. Verwey (1980), *Practical Aspects of Computational River Hydraulics*, Pitman Publ., London.
- Fennema, R.J., and M.H. Chaudhry (1990), Explicit methods for 2-D transient free-surface flows, *J. Hydraul. Eng.* **116**, 8, 1013-1034, DOI:10.1061/(ASCE)0733-9429(1990)116:8(1013).
- Garcia-Navarro, P., F. Alcrudo, and J.M. Savirón (1992), 1-D open-channel flow simulation using TVD-McCormack scheme, *J. Hydraul. Eng.* **118**, 10, 1359-1372, DOI: 10.1061/(ASCE)0733-9429(1992)118:10(1359).
- Katopodes, N.D (1984), A dissipative Galerkin scheme for open-channel flow, *J. Hydraul. Eng.* **110**, 4, 450-466, DOI:10.1061/(ASCE)0733-9429(1984)110:4(450) .
- Kundu, A. (ed.) (2007), *Tsunami and Nonlinear Waves*, Springer, Berlin.
- Lamb, H. (1932), *Hydrodynamics*, Cambridge Univ. Press, Cambridge.
- Mader, C.L. (2004), *Numerical Modeling of Water Waves*, CRC Press, Boca Raton.
- Namin, M.M., B. Lin, and R.A. Falconer (2001), An implicit numerical algorithm for solving non-hydrstatic free-surface flow problems, *Int. J. Num. Meth. Fluids* **35**, 341-356.
- Quecedo, M., and M. Pastor (2003), Finite element modelling of free surface flows on inclined and curved beds, *J. Comput. Phys.* **189**, 1, 45-62, DOI: 10.1016/S0021-9991(03)00200-6.
- Radespiel, R., and N. Kroll (1995), Accurate flux vector splitting for shocks and shear layers, *J. Comput. Phys.* **121**, 1, 66-78, DOI:10.1006/jcph.1995.1179.
- Schlichting, H., and K. Gersten (2000), *Boundary Layer Theory*, 8th ed., Springer-Verlag, Berlin.
- Stoker, J.J. (1957), *Water Waves*, Interscience, New York.
- Strelkoff, T. (1969), One-dimensional equations of open-channel flow, *J. Hydraul. Div.* **95**, 3, 861-876.

- Xing, Y., and C.-W. Shu (2005), High order finite difference WENO schemes with the exact conservation property for the shallow water equations, *J. Comput. Phys.* **208**, 1, 206-227, DOI:10.1016/j.jcp.2005.02.006.
- Yen, B.C. (1973), Open-channel flow equations revisited, *J. Eng. Mech. Div.* **99**, 5, 979-1009.

Received 10 July 2014
Accepted 15 October 2014



Investigation of the 16-year and 18-year ZTD Time Series Derived from GPS Data Processing

Zofia BAŁDYSZ, Grzegorz NYKIEL, Mariusz FIGURSKI,
Karolina SZAFRANEK, and Krzysztof KROSZCZYŃSKI

Military University of Technology, Faculty of Civil Engineering and Geodesy,
Warszawa, Poland; e-mail: zbaldysz@wat.edu.pl

Abstract

The GPS system can play an important role in activities related to the monitoring of climate. Long time series, coherent strategy, and very high quality of tropospheric parameter Zenith Tropospheric Delay (ZTD) estimated on the basis of GPS data analysis allows to investigate its usefulness for climate research as a direct GPS product. This paper presents results of analysis of 16-year time series derived from EUREF Permanent Network (EPN) reprocessing performed by the Military University of Technology. For 58 stations Lomb-Scargle periodograms were performed in order to obtain information about the oscillations in ZTD time series. Seasonal components and linear trend were estimated using Least Square Estimation (LSE) and Mann-Kendall trend test was used to confirm the presence of a linear trend designated by LSE method. In order to verify the impact of the length of time series on trend value, comparison between 16 and 18 years were performed.

Key words: GPS, ZTD, time series, troposphere.

1. INTRODUCTION

Water vapour is one of the main gases causing greenhouse effect. Because it is responsible for energetic balance of the Earth (Held and Soden 2006), and in 60-70% for increasing the temperature on the surface of Earth (COST

2012), water vapour plays a key role in the process of climate monitoring. Besides many techniques which allow to measure water vapour content in the atmosphere (Integrated Water Vapour – IWV), GPS system plays a significant role in this task. Advanced analysis of GPS observations provides tropospheric parameter Zenith Tropospheric Delay (ZTD), which represents the impact of weather conditions (troposphere) on the propagation of satellite signals. ZTD consists of two parts: a delay caused by the hydrostatic part of atmosphere (Zenith Hydrostatic Delay – ZHD), and delay related to the wet part of atmosphere (Zenith Wet Delay – ZWD). ZHD accounts for about 90% of ZTD, and it is easy to model due to the low variability of this parameter in time. ZWD, which accounts for about 10% of ZTD, cannot be accurately modeled due to uneven space, temporal, and vertical distribution of water vapour. Because the size of ZTD depends on the state of the troposphere, this parameter shows correlation with changeable-in-time temperature (Guerova 2013) and time-varying content of water vapour (Yong *et al.* 2008). Using selected meteorological data it also allows to estimate IWV (Bevis *et al.* 1992, Hagemann *et al.* 2003, Wang and Zhang 2009) with accuracy at the level of 1-2 kg m⁻² (Bock *et al.* 2007, Byun and Bar-Server 2009). Denser than in the case of radiosonde, network of GPS permanent stations which leads measurements with high temporal resolution, enables monitoring of troposphere variability and detecting trends in water vapour content for large areas, simultaneously ensuring proper spatial distribution. Monitoring of long-term changes is crucial in the context of monitoring climate changes. Longstanding global investigations, conducted using for example DORIS technique, show discrepancy in character and size of IWV trend, both in north and south hemisphere (Bock *et al.* 2014). Analogical situations concerning global trends occur when using GPS system (Ning 2012, Jin *et al.* 2007). In that case, discrepancies were shown not only in each hemisphere, but also in single continents on similar latitudes – for example in Europe. Different characters of trends for Europe were also shown in an analysis performed using radiosonde data conducted only for northern hemisphere (Ross and Elliot 2001). In that case, discrepancies were shown also for stations situated close to each other.

In this paper, the authors focus on ZTD, as a direct GPS product free from uncertainty connected to the interpolation of meteorological measurements (Schüler 2001) and used models, for both determining ZHD and converting ZWD to IWV (van Malderen *et al.* 2014). Thanks to homogenous set of the EUREF Permanent Network (EPN) data and due to coherent strategy for the satellite data processing, GPS ZTD time series can be used in researches related to the monitoring of climate changes and for climate model simulations (Pacione *et al.* 2014). Studies connected to variability of ZTD provide also information about the average local weather conditions, because

size and character of these oscillations are related to such factors as latitude and height of the receiver or distance from big masses of water (Jin *et al.* 2007). To represent changes in the troposphere above Europe, data processed by Military University of Technology Local Analysis Centre (MUT LAC) consisting of 16-year times series determined for 59 EPN stations were used. For each station, the authors determined annual oscillations with their harmonic derivatives, and linear trend by means of Least Square Estimation (LSE). Trend with Mann–Kendall trend test to verify the results of the LSE was also detected. Studies related to climate change monitoring, based on GNSS long time series of homogeneous observations are one of tasks defined in the COST Action ES1206 “Advanced Global Navigation Satellite Systems tropospheric products for monitoring severe weather events and climate (GNSS4SWEC)”.

2. ZTD DATA

Electromagnetic wave going through a neutral atmosphere is delayed, because of the refraction and tropospheric attenuation. The value of this delay is given in zenith direction and is defined as follows (Bevis *et al.* 1992):

$$ZTD = \int_H^{+\infty} [n(z)-1] dz = 10^{-6} \int_H^{+\infty} N(s) dz , \tag{1}$$

where $n(z)$ is the atmospheric refractive index, and $N(z)$ is the atmospheric refractivity as a function of altitude z .

This delay is caused by hydrostatic part of the atmosphere (ZHD) and by wet part of the atmosphere (ZWD). In the case of GPS system, the size of these delays depends on total way along which the signal crosses the atmosphere (Slant Tropospheric Delay – STD) and hence, depends on of the satellite’s zenith distance. To relate the size of delay from zenith direction (ZTD) to satellite direction (STD), mapping functions are being used. They provide *a priori* ZTD value and transfer it to STD, with takes into consideration different methods in mapping hydrostatic (ZHD) and wet (ZWD) part of the atmosphere. Mapping functions are approximately equal to $1/\sin e$, where e is the elevation angle, but in case of precise measurements it is required to use continued function given by Marini (1972), normalized and described by the formula (Herring 1992):

$$m(f) = \frac{1 + \frac{a}{1 + \frac{b}{1+c}}}{\sin e + \frac{a}{\sin e + \frac{b}{\sin e + c}}} , \tag{2}$$

where e is the elevation angle, and a, b, c are coefficients related to the state of the troposphere.

Data used in the analysis comes from the EPN (Bruyninx 2004). This network was built on the basis of global geodetic network International GNSS Service (IGS) and plays a role of IGS densification in Europe. It has been working since 1996, and since that time it has been giving solutions for consecutive realizations of both ITRS and ETRS89 systems. Routine processing of the observations performed by the EPN network has been conducted by using different strategies, models, parameters and software, constantly since 1996, which caused collecting inhomogeneous data (coordinates and troposphere parameters). As a consequence, it was impossible to carry out a proper analysis of long time series obtained from EPN. Taking into account these problems and the fact that new, more precise products (like Earth rotation parameters or GPS satellite orbits) appeared, in 2007 it was decided to recalculate these data. This work, called Repro 1 campaign, was preceded by tests done simultaneously by the Royal Observatory of Belgium and the Military University of Technology. MUT reprocessing (Figurski *et al.* 2009, Söhne *et al.* 2010) included data from all EPN stations from January 1996 to December 2007 and used orbits and Earth Rotation Parameters (ERP) from the Potsdam–Dresden re-processing (Steigenberger *et al.* 2006) with the Bernese GPS software version 5.0 (Dach *et al.* 2007). Daily RINEX files containing less than 50% of possible observations were ignored. Residuals larger than 0.02 m were marked in a preprocessing step and removed in the final estimation. All stations with North East residuals greater than 10 mm and Up residuals greater than 20 mm were investigated and removed. Zenithal Total Delay was determined using Niell's mapping functions (Niell 1996) with 3° elevation mask. After the Repro1 campaign, MUT LAC has been providing solutions only for one of EPN subnets. Consequently, we had a set of homogeneous data (from all EPN network) only to the end of 2007 and after this year, only for those stations that have been allocated in MUT LAC. Therefore, out of the time series available from the whole EPN network, we picked those stations which started operating no later than 1998, and then for stations that have not been attached to MUT LAC (located in different subnets) we determined coordinates and troposphere parameters, according to the MUT LAC Repro 1 strategy. Thanks to consistent calculating strategy used for all observations, we received at least 16-year homogeneous time series of hourly solutions, which are an essential element in research connected with long time climate trend (Bengtsson *et al.* 2004) for 59 EPN stations (Fig. 1 and Table 1 in the Appendix).

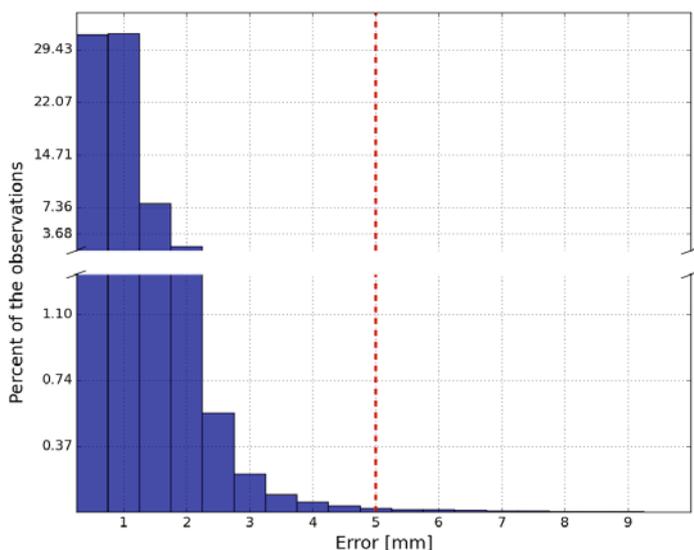


Fig. 2. Histogram of the uncertainty for ZTD solutions (16 years, 59 EPN stations).

longer time series (for stations which have been operating since January 1996) to period of January 1998 – December 2013 to ensure their coherence. For this period, all ZTD solutions for which coordinates were not fixed properly were rejected and all data with error larger than 5 mm were removed. Figure 2 presents histogram of ZTD error for all the solutions. Most of them is in the 1-2 mm range. Red dashed line shows the 5 mm reject criterion and the number of solutions that were removed from further analysis.

The next step in data screening was the rejection of stations whose time series were characterized by a low number of solutions in relation to their theoretical maximum number. The number of available solutions was on an average level of 95.36% of the theoretical number of solutions. Available solutions which fulfill the criterion of 5 mm account for about 95.10% of the theoretical number of solutions. The best quality of the time series has the DELF station. The number of solutions available for this station accounts for 99.58% of the theoretical number of solution and the available number of solutions that fulfill the criterion of 5 mm accounts for 99.51% of the theoretical number of solution. The worst quality has ANKR station, with the number of available solutions and the number of available solutions that meet the 5 mm criterion at the level of 82.82 and 81.42%, respectively. Because the ANKR has less than 90% of the theoretical number of solutions, this station was removed from further analysis.

4. ANALYSIS OF ZTD TIME SERIES

In order to determine the value of linear trend in ZTD time series, seasonal variations, like, *e.g.*, annual and semi-annual oscillations were taken into consideration. Time series of stations refer to exactly the same period (1998-2013) but they have some gaps in data; therefore, to obtain information about oscillation, the Lomb–Scargle periodogram was prepared for every station (Hocke 1998). This method estimates frequency spectrum by fitting linear least-squares of sine and cosine model to the observed time series (Lomb 1976, Press *et al.* 1992):

$$x(t_i) = a \cos(\omega t_i - \Theta) + b \sin(\omega t_i - \Theta) + n_i , \tag{3}$$

where $x(t_i)$ is the observed time series at time t_i , a and b are constant amplitudes, ω is the angular frequency, Θ is additional phase (required for the orthogonalization of the sine and cosine model functions when the data are unevenly spaced), and n_i is the noise at time t_i .

Periodograms show various oscillations and characters of time series, because stations involved in the analysis are spread throughout Europe and different average weather conditions had influence on the ZTD size. All stations have a clear annual oscillation, and most of them have clear semi-annual oscillations. However, some of them have also other variations or different configurations of variations (*e.g.*, without semi-annual oscillations). For every station 2 periodograms were prepared (with and without annual oscillation), due to strong annual oscillations which disturb the character of oscillations with smaller amplitude and thus hinder their investigation. Figure 3 shows Lomb–Scargle periodograms for the ZTD time series for selected stations. Station GRAS (Fig. 3a) has clearly only annual oscillation with amplitude 41.7 mm (a similar character have, *e.g.*, MARS and CAGL stations). Significant annual and semi-annual oscillations has, *e.g.*, the GLSV station (Fig. 3b) with amplitude 53.3 and 10.6 mm, respectively. In some cases, semi-annual oscillations were smaller than oscillations with 1/3 year frequency – like SFER (Fig. 3c) with amplitudes of 24.2, 4.6, and 7.0 mm for annual, semi-annual, and ter-annual oscillations, respectively. MAS1 station (Fig. 3d) has noticeable annual, semi-annual, ter-annual, and even quarto-annual amplitudes with 29.9, 7.9, 5.8, and 2.5 mm, respectively. In case of RAMO station (Fig. 3e) semi-annual and ter-annual oscillations have significant value of amplitude as compared to the value of annual oscillation amplitude.

Based on the results from the periodograms analysis, estimation of the linear trend and seasonal components was carried out using the model:

$$x(t) = x_0 + v_x t + A_A^I \sin 2\pi t + A_A^O \cos 2\pi t + A_{SA}^I \sin 4\pi t + A_{SA}^O \cos 4\pi t + A_{TA}^I \sin 6\pi t + A_{TA}^O \cos 6\pi t + A_{QA}^I \sin 8\pi t + A_{QA}^O \cos 8\pi t , \tag{4}$$

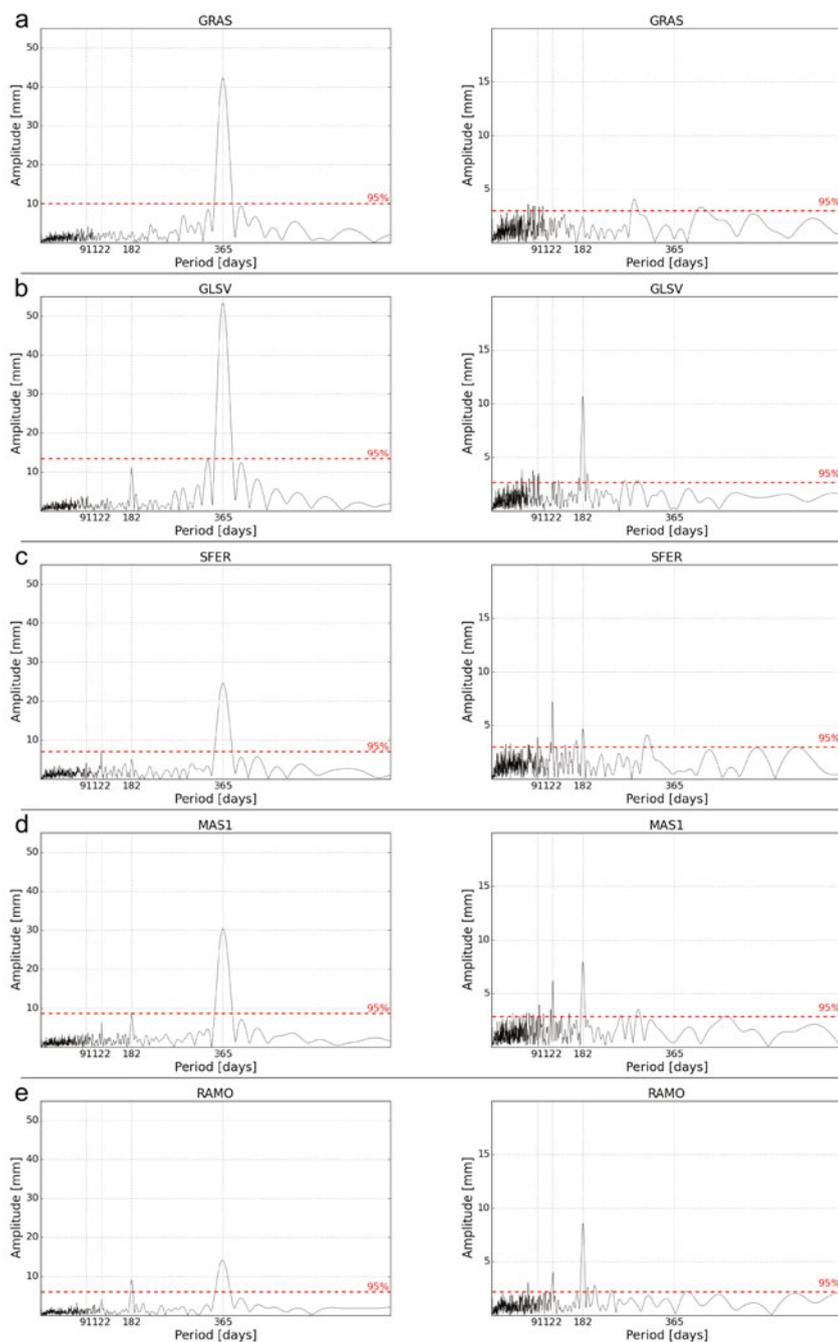


Fig. 3. Lomb–Scargle periodograms for GRAS (a), GLSV (b), SFER (c), MAS1 (d), and RAMO (e) stations.

where $x(t)$ is the observed time series, $v_x t$ is the linear trend, $A_A^I, A_A^O, etc.$ are coefficients for the condition of sine and cosine using the method of least squares. The LSE trend and amplitude of annual, semi-annual, ter-annual, and quarto-annual term (obtained from Lomb–Scarge periodograms) were calculated on the basis of hourly data (results in Section 5).

In order to verify the trend estimated using LSE modified Mann–Kendall trend test (Mann 1945, Kendall and Stuart 1970) for all stations was also prepared. Mann–Kendall trend test is a statistical, non-parametric test which can be used for trend detection in climatologic time series (Goosens and Berger 1986, Mavromatis and Stathis 2011) due to the fact that it does not require time series with normally distributed data and has low sensitive to gaps in data (Karmeshu 2012). The test execution returns information about whether a trend exists (hypothesis True) or whether the data is random and independent of each other, and because of that there is no trend in time series (hypothesis False). In addition, MK trend test returns the value of a statistical factor S , whose character (positive or negative) reflects character of trend in the time series:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(X_j - X_i)$$

$$\text{sign}(X_j - X_i) = \begin{cases} 1 & \text{if } X_j - X_i > 0 \\ 0 & \text{if } X_j - X_i = 0 \\ -1 & \text{if } X_j - X_i < 0 \end{cases} \quad (5)$$

where X_j and X_i are time series, with $i = 1, 2, 3, \dots, n-1$ and $j = i+1, i+2, i+3, \dots, n$.

In case of described analysis, the MK trend test for time series with deleted oscillations (annual, semi-annual, ter-annual, quarto-annual) was performed. The MK trend test does not give result as a value of trend, but it is useful in case of these stations which due to the low value of the trend (obtained by the LSE) require additional verification.

5. RESULTS OF ANALYSIS

5.1 16-year time series

In this paper we focused on 16-year ZTD time series, mostly due to the fact that, assuming the need to analyze exactly the same period of time for all stations, their spatial resolution is better than in the case of 18-year time series (more stations). Based on the results obtained from Lomb–Scarge periodograms and LSE method, annual, semi-annual, ter-annual, and quarto-annual oscillations for every station were estimated. The average size of the annual oscillation amplitude (for all stations) is on the level of

46.8 mm, with maximum value for TORI station and minimum value for RAMO station, which are 63.3 and 13.72 mm, respectively. Average amplitude of semi-annual oscillations is 7.5 mm, with maximum value for JOEN station (11.9 mm). For some stations (*e.g.*, MATE) this oscillation does not exist. These amplitudes are similar, but a little smaller than average global values for annual and semi-annual oscillations given by Jin (Jin *et al.* 2007), which are 50 and 10 mm, respectively. Different characters of time series, determined by different size of seasonal components, are shown in Figs. 4 and 5. Small amplitude of annual variations (22.6 mm) for CASC station (Fig. 4) in combination with one of the highest average ZTD values (2.425 m) reflects climate character in area of this station. It is located in Portugal, on the coast of the Atlantic Ocean in the humid subtropical zone, where high humidity and lack of large temperature variations are probably due to the influence of the Gulf Stream. In contrast, GRAS station (Fig. 5) with the 41.8 mm of annual amplitude and the lowest average ZTD value (2.054 mm) represents the conditions corresponding to large heights (station is located at 1319.3 m a.s.l.).

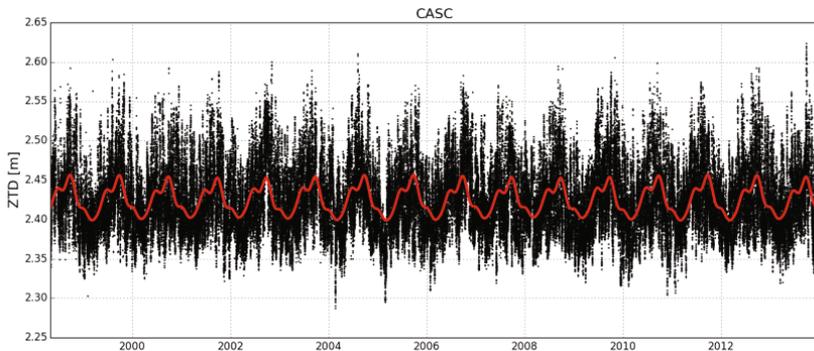


Fig. 4. ZTD time series for CASC (Portugal) station with fitted oscillations.

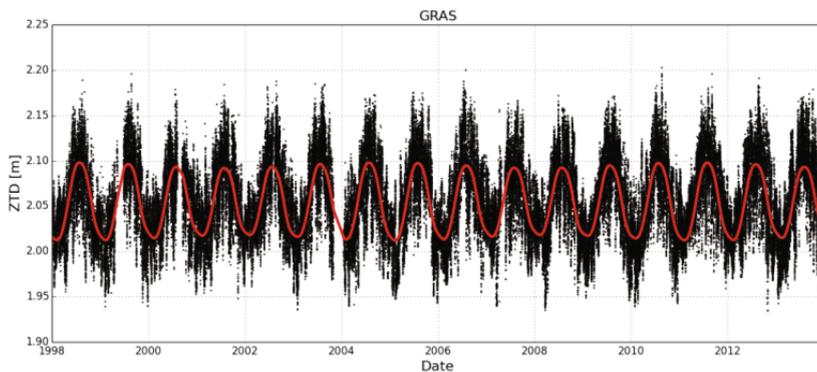


Fig. 5. ZTD time series for GRAS (France) station with fitted oscillations.

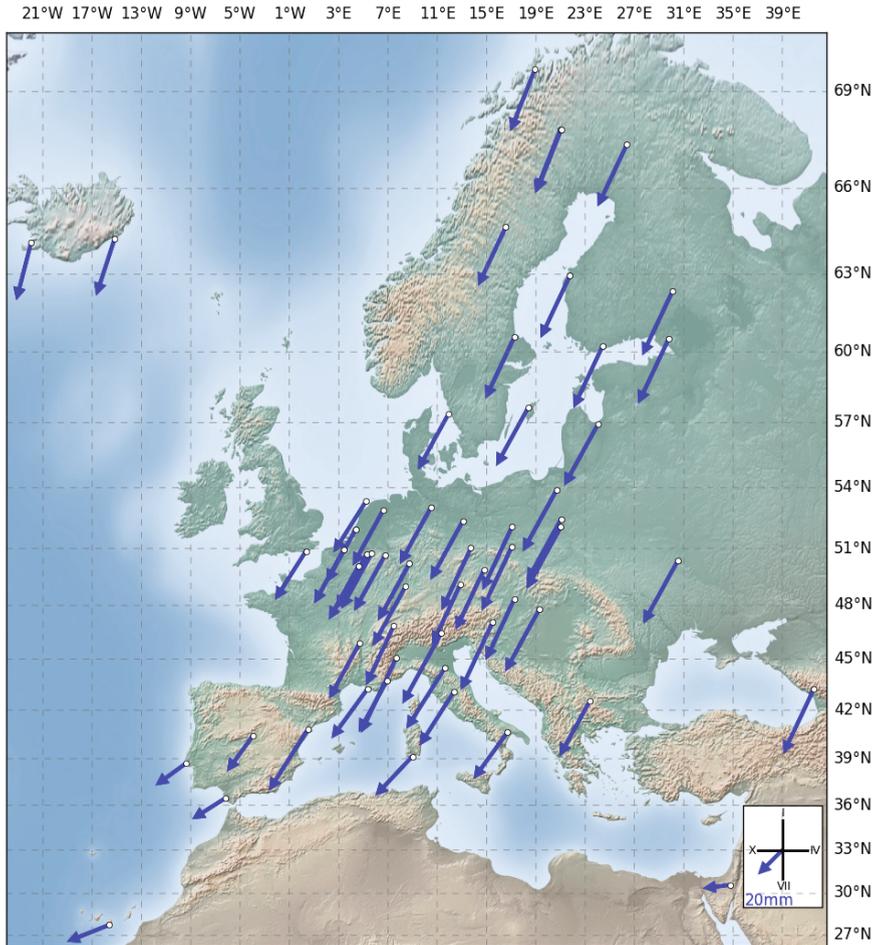


Fig. 6. Amplitude and phase shift for annual ZTD oscillations (January 1998 – December 2013).

Seasonal components of ZTD time series, like annual variations, could be useful for climate applications, especially with taking into consideration size of their amplitude and phase shift (month with maximum value of sinusoid). Variations of ZTD are mostly determined by wet component of troposphere. Their maximum amplitude is usually correlated with maximum temperature and humidity. Figure 6 shows size of amplitude, with an indication of the time when the amplitude reaches a maximum. For most of the station it is in the end of July or in the beginning of August, which coincides with the summer months in which the highest temperatures are observed.

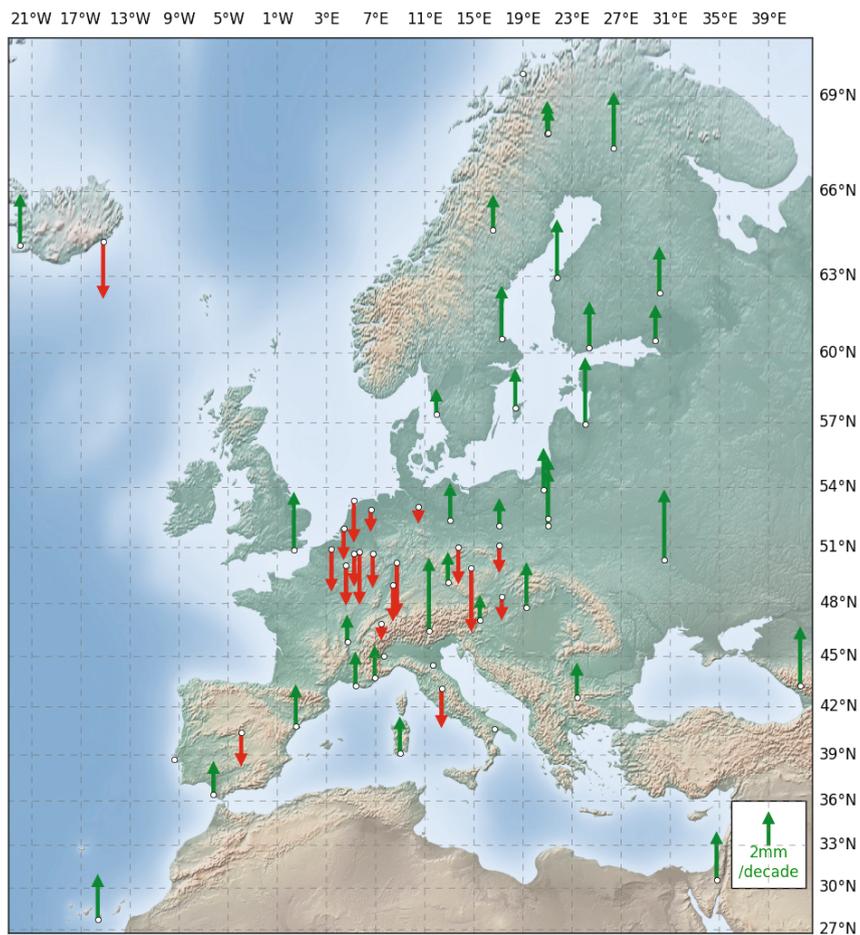


Fig. 7. ZTD trend for 16-year time series (January 1998 – December 2013).

Trend detection in ZTD long time series has very high value for climate change monitoring. Using LSE, the trend value was determined and the existence of trend was confirmed by means of Mann–Kendall method. Only in the case when the Mann–Kendall test confirmed the existence of a trend (True), the station was taken for further analysis. From 58 EPN stations, using both above-mentioned methods, the occurrence of a trend was confirmed for 54 stations. Stations without proven trend are: CASC, MATE, MEDI, TORI. The average value of trend (for the rest of stations) is 1.0 mm/decade, but this value includes both positive and negative trends. 34 stations have positive trend and 19 stations negative. The highest positive trend was found for the BZRG station and its size is 5.5 mm/decade. The most negative trend was found for GOPE station and its size is -4.7 mm/decade. Figure 7 shows

location, size, and type of determined trend. Green arrows indicate positive trends, and red arrows indicate negative. For the north-eastern part of Europe, the trend characters are the same (positive), only with some discrepancies to the size of ZTD trend. The largest discrepancies to the nature of the trend occurred in Central-Western Europe, where in Western Germany, Belgium, Nederland, and Eastern France is the superiority of stations with a negative trend. These negative trends are small in magnitude but occur in almost every station. Their meridional distribution may indicate that the distribution of air masses in the summer months is responsible for the changes in the ZTD values. In July, the circulation of air over continental part of Europe is based largely on the Azores High (high-pressure zone over Atlantic) and low-pressure zone over Eurasian. Maximum range of occurrences of these both atmospheric pressure centers has also a meridian run (which is variable over time), coinciding with the occurrence of these negative trends. Perhaps in the analyzed period of time the average range of one of these atmospheric pressure centers has been changed and that could be reflected in changes of average ZTD values (negative trends). Most of the other stations which are characterized by negative trends (GOPE, DRES, WROC, MOPI, UNPG, VILL) lie near the mountains which very often have a different type of climate, which may cause discrepancies in trends' character.

All results (annual amplitude, semi-annual amplitude, trend, $m\cdot y$, and Mann–Kendall trend test results) for 16-year ZTD time series are presented in Table 2 in the Appendix.

5.2 Comparison of 16- and 18-year ZTD time series

Indicated by Nilsson and Elgered (2008) necessity of data from exactly the same period analysis imposes the need to reduce the time series in order to achieve the most optimal spatial resolution. On the other hand, determination of the trend should be based on the longest possible observations. The two-year difference in the length of the time series may be important not only in relation to the estimated size of the trend, but also as to its character. In EPN network 30 stations have at least 18-year ZTD time series (they have been operating since 1996) and for these stations a comparison between shortened 16- and full 18-year ZTD time series has been prepared.

Annual amplitudes for 18-year ZTD time series are different than in case of 16-year time series for every station. Size of amplitudes for longer period of time are smaller than in case of shorter period of time, but these differences are less than 1 mm. Semi-annual amplitudes for 18-year time series are also different than for 16-year time series (except RAMO station) for every station. In contrast to the annual amplitudes, larger size of semi-annual am-

plitudes occurs in most of the stations for longer period of time. However these differences are also less than 1 mm.

For 4 stations MK trend test gave different result for 16-year time series than in case of 18-year time series. In case of DELF and WARE the result of MK trend test was positive (True) for shorter period of time and negative (False) for longer period of time. In case of MATE and MEDI result of MK trend test was negative (False) for shorter period of time and positive (True) for longer period of time. Differences in the number of stations for which the trend has been confirmed (for both lengths of time series), confirms that in order to analyze the changes in the spatial distribution of ZTD, it is necessary to work on the exact same period of time. This is due to the fact that for small value of the trend and time series not long enough, an additional one or two seasons with stronger or weaker weather conditions than usual, may noticeably affect the size of trend. Average value of trend, for rest 26 stations, is 1.5 mm/decade for 16-year time series and 2.0 mm/decade for 18-year time series. For 4 stations (MAS1, MOPI, RAMO, RIGA) values of trends are exactly the same for two periods of time and are 3.0, -0.1 , 3.3, and 5.0 mm/decade, respectively. For ZIMM station value of trend for 16-year time series is negative (-0.4 mm/decade) and for 18-year time series is positive (1.2 mm/decade), whereas for both time series result of MK trend test is positive (True). In case of this station confidence interval is better for longer period of time, however the size of linear trend for shortened period of time is quite small and probably a slight change in the average weather conditions in additional two years might affect this change. An example of this station shows how sensitive to the changes, depending on the length of time series, the ZTD parameter is. Figure 8 represents all differences between results in trend value obtained from 18- and 16-year time series for 30 EPN stations.

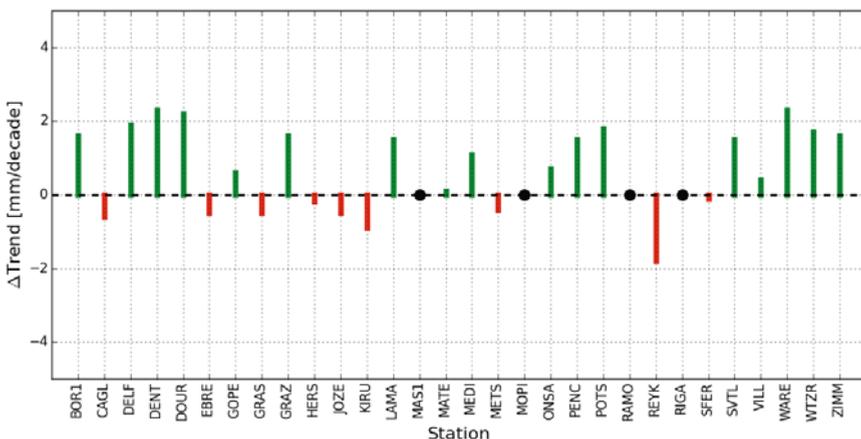


Fig. 8. Differences in trend value for 18- and 16-year time series.

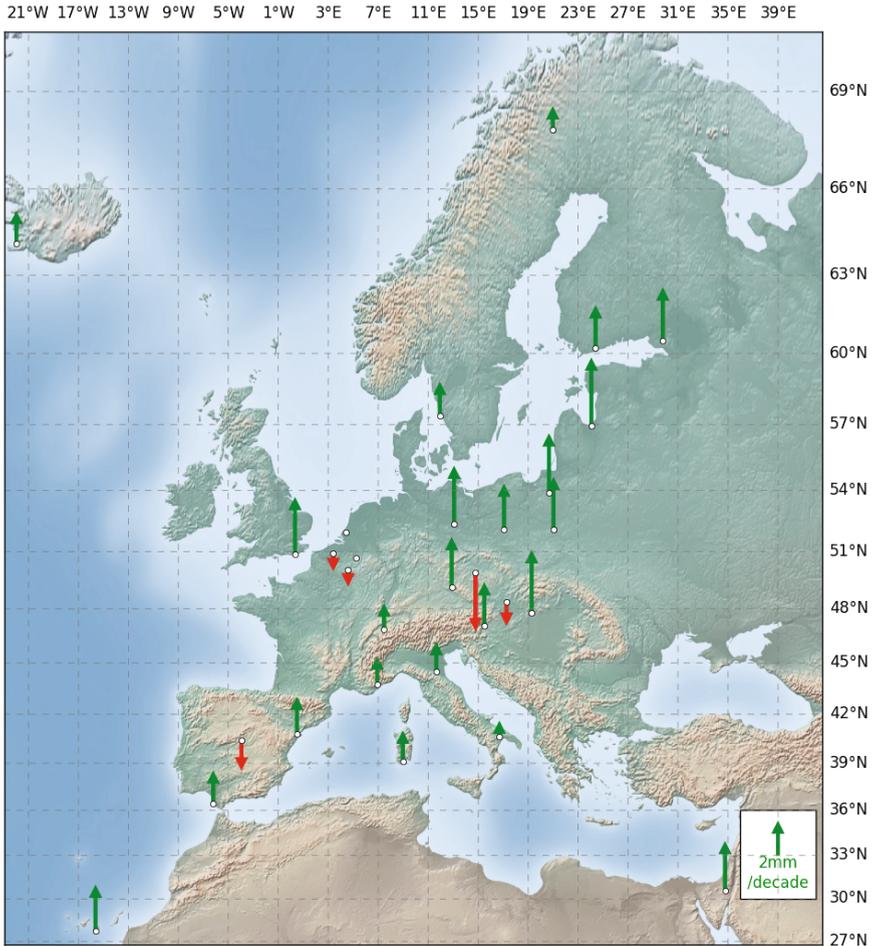


Fig. 9. ZTD trend for 18-year time series (January 1996 – December 2013).

All results for 18-year ZTD time series are presented in Table 2 in the Appendix. In addition the results of linear trends estimated based on 18-years ZTD time series are presented in Fig. 9.

Long time series obtained from 30 EPN stations indicate the advantage of a positive trend in Europe. Only 4 stations (DENT, DOUR, GOPE, VILL) have negative character of trend and 26 stations have positive character of trend. For 2 stations (DELF, WARE), the MK trend test gave a negative result (False). The highest positive trend has RIGA station, which is 5.0 mm/decade. The most negative trend has (as in 16-year time series) GOPE station, which is -4.7 mm/decade. As in the case of 16-year time series, discrepancies exist in Central-Western Europe, although they are less

significant. Primarily, a smaller number of stations in this area could be included in the 18-year analysis (due to too short time series), but on the other hand the size of the trends, at the stations which were subjected to analysis, is lower than in the case of 16-year ZTD time series.

The comparison showed the importance of additional two years of observations for the value of linear trend. The vast majority of stations is represented by a different size of the trend for the 16- and 18-year time series and one of stations has even different nature of trend. This shows how important for climate application is using the longest possible, but at the same time preserving the homogeneity, time series.

6. SUMMARY

In this paper, long ZTD time series were analyzed. Seasonal components (annual, semi-annual, ter-annual, quarto-annual) and linear trend were found using LSE. In order to obtain information about the oscillations, Lomb–Scargle periodograms were performed for every station. Removed annual amplitude from time series allowed to reduce impact that it has on the oscillations with smaller amplitude and thus their correct verification. This approach is relevant for the analysis of a station with large spatial distribution, where the characters of the time series are significantly different from each other mainly due to different weather conditions.

The highest annual amplitude was found for TORI station (63.3 mm) and the lowest for RAMO station (13.7 mm). The highest semi-annual amplitude was found for JOEN station (11.9 mm) and for some stations this oscillation does not exist (*e.g.*, 0.6 mm for MATE station). The nature of the occurring oscillations represents the average weather conditions prevailing in the area. High amplitude of annual oscillations is related to the continental or mountain type of climate, where the moderating influence of the warm mass of water is limited. It is probable that changes in the size of annual oscillations (increasing or decreasing from year to year) may reflect changes in the range of impact on the continent different type of climate, as well as linear trend. The decreasing size of amplitude may be caused by the increasing influence of the warm water masses on average weather conditions. However, the study of such dependence requires other tools and methods than LSE approach.

In the case of a linear trend, results should be considered in two categories: 16-year time series and 18-year time series to assess the impact of additional two years. For 16-year ZTD time series the highest positive and the most negative trends have BZRG and GOPE stations, with 5.5 mm/decade and -4.7 mm/decade, respectively. For 18-year ZTD time series the highest positive and the most negative trends have RIGA and GOPE stations, with 5.0 mm/decade and -4.1 mm/decade, respectively. For ZIMM station the

character of trend for 16-year ZTD data is against the character of the trend for 18-year ZTD data. There are also considerable differences in the size of the trend between two periods of time, which in case of linear trend spatial distribution analysis (important for climate applications), indicates the necessity of use data from exactly the same period of time.

To verify the existence of the trend obtained by LSE method, MK trend test was also performed. This test is particularly useful for stations with a small value of trend designated with LSE. Because MK trend test gave positive results for the verification of the existence of the trends, despite the divergence in their values and characters, the way of interpreting the trend becomes particularly important. Due to the fact that two years may have a significant effect on the character and magnitude of the changes, they should be seen not only in the category of linear change, but also in the category of successive short-term changes (year after year). As in the case of observation of amplitudes changes, this task requires the use of more tools than just LSE method.

Interpretation of the results is ambiguous because of the very large variety of climate types found in Europe, as well as harshness of the terrain, but taking into account the basic aspects arising from locations of the stations, many discrepancies in a trend character can be explained. It should be emphasized that the length of time series is still not sufficient for full climate research. However, on the one hand the development of appropriate methods of obtaining and interpreting results is important from the point of view of the possibility of adjusting constantly developing GNSS technology for the purpose of examining the troposphere. On the other hand, almost 20-year time series can be a sufficient source of data for supplementing and validating the climate models.

Acknowledgments. This research was financed by the Faculty of Civil Engineering and Geodesy of the MUT statutory research funds. The authors would like to acknowledge the contribution of the COST Action ES1206.

Appendix

The list of EPN stations and the results

Table 1

The list of EPN stations taken for analysis

Station	Latitude [deg.]	Longitude [deg.]	H [m]	Station	Latitude [deg.]	Longitude [deg.]	H [m]
ANKR	39.89	32.76	974.8	MATE	40.65	16.70	535.6
BOGO	52.48	21.04	149.6	MEDI	44.52	11.65	50.0
BOR1	52.10	17.07	124.0	METS	60.22	24.40	94.6
BZRG	46.50	11.34	328.8	MOPI	48.37	17.27	579.0
CAGL	39.14	8.97	238.4	ONSA	57.40	11.93	45.5
CASC	38.69	-9.42	77.1	PENC	47.79	19.28	291.7
DELF	51.99	4.39	74.4	POTS	52.38	13.07	174.0
DENT	50.93	3.40	63.9	RAMO	30.60	34.76	893.1
DOUR	50.09	4.59	283.0	REYK	64.14	-21.96	93.1
DRES	51.03	13.73	202.9	RIGA	56.95	24.06	34.7
EBRE	40.82	0.49	107.9	SFER	36.46	-6.21	85.8
EIJS	50.76	5.68	103.8	SJDV	45.89	4.68	336.0
EUSK	50.67	6.76	245.3	SODA	67.42	26.39	299.7
GLSV	50.36	30.50	226.8	SOFI	42.56	23.39	1119.6
GOPE	49.91	14.79	592.6	SVTL	60.53	29.78	77.1
GRAS	43.75	6.92	1319.3	TERS	53.36	5.22	56.1
GRAZ	47.07	15.49	538.3	TORI	45.06	7.66	310.4
HERS	50.87	0.34	76.5	TRO1	69.66	18.94	138.0
HOBU	53.05	10.48	152.3	UNPG	43.12	12.36	351.2
HOFN	64.27	-15.19	82.5	VAAS	62.96	21.77	58.0
JOEN	62.39	30.10	113.7	VIL0	64.70	16.56	449.9
JOZE	52.10	21.03	141.4	VILL	40.44	-3.95	647.5
KARL	49.01	8.41	182.9	VISO	57.65	18.37	79.8
KIR0	67.88	21.06	497.9	WARE	50.69	5.25	188.0
KIRU	67.86	20.97	391.1	WROC	51.11	17.06	181.0
KLOP	50.22	8.73	222.4	WSRT	52.91	6.60	86.0
LAMA	53.89	20.67	187.0	WTZR	49.14	12.88	666.0
MAR6	60.60	17.26	75.4	ZECK	43.29	41.57	1166.8
MARS	43.28	5.35	61.8	ZIMM	46.88	7.47	956.7
MAS1	27.76	-15.63	197.3				

Table 2

Results of analysis of 16- and 18-year ZTD time series.
 Bolded trend values represent positive result of Mann–Kendall trend test.

Station	Mean ZTD [m]	Annual amplitude [mm]	Semi-annual amplitude [mm]	Trend value ± error [mm/year]	Mean ZTD [m]	Annual amplitude [mm]	Semi-annual amplitude [mm]	Trend value ± error [mm/year]
	16-year ZTD time series				18-year ZTD time series			
BOGO	2.374	52.65	9.44	0.45±0.02				
BOR1	2.385	51.85	9.42	0.14±0.02	2.385	51.80	9.62	0.30±0.02
BZRG	2.347	62.04	6.36	0.55±0.02				
CAGL	2.376	37.63	3.42	0.22±0.02	2.376	37.88	2.83	0.16±0.02
CASC	2.425	22.57	1.69	-0.03±0.02				
DELF	2.405	43.89	7.94	-0.18±0.02	2.405	43.76	8.06	0.01±0.02
DENT	2.412	43.98	7.63	-0.28±0.02	2.411	43.78	7.89	-0.05±0.02
DOUR	2.346	43.92	7.68	-0.26±0.02	2.346	43.50	7.96	-0.04±0.02
DRES	2.370	50.89	8.42	-0.23±0.02				
EBRE	2.425	54.21	8.94	0.27±0.02	2.424	54.28	9.13	0.22±0.02
EIJS	2.402	46.00	8.12	-0.37±0.02				
EUSK	2.358	46.36	7.91	-0.20±0.02				
GLSV	2.352	53.26	10.60	0.53±0.02				
GOPE	2.255	48.76	8.40	-0.47±0.02	2.255	48.45	8.54	-0.41±0.02
GRAS	2.055	41.77	2.46	0.19±0.02	2.054	41.61	3.13	0.14±0.02
GRAZ	2.281	56.33	8.06	0.12±0.02	2.281	55.47	8.39	0.28±0.02
HERS	2.406	40.95	7.85	0.42±0.02	2.405	40.95	8.28	0.40±0.02
HOBU	2.377	47.10	8.11	-0.04±0.02				
HOFN	2.36	42.48	6.67	-0.41±0.02				
JOEN	2.349	53.53	11.91	0.31±0.02				
JOZE	2.380	52.85	10.48	0.41±0.02	2.380	52.77	10.65	0.36±0.02
KARL	2.386	49.62	8.43	-0.22±0.02				
KIRO	2.224	49.49	9.46	0.12±0.02				
KIRU	2.253	50.33	10.05	0.19±0.02	2.253	49.94	11.14	0.10±0.02
KLOP	2.367	47.28	7.85	-0.38±0.02				
LAMA	2.359	52.18	9.73	0.27±0.02	2.358	51.76	9.84	0.42±0.02
MAR6	2.370	50.61	9.34	0.37±0.02				
MARS	2.419	44.30	1.21	0.20±0.02				
MAS1	2.392	29.87	7.92	0.30±0.02	2.392	29.81	7.47	0.30±0.02
MATE	2.283	41.1	0.56	0.02±0.02	2.283	40.97	0.62	0.03±0.01
MEDI	2.43	54.48	3.52	0.04±0.02	2.430	54.09	4.37	0.15±0.02
METS	2.36	50.41	9.11	0.31±0.02	2.360	49.92	9.76	0.27±0.02

to be continued

Table 2 (continuation)

Station	Mean ZTD [m]	Annual amplitude [mm]	Semi-annual amplitude [mm]	Trend value \pm error [mm/year]	Mean ZTD [m]	Annual amplitude [mm]	Semi-annual amplitude [mm]	Trend value \pm error [mm/year]
	16-year ZTD time series				18-year ZTD time series			
MOPI	2.256	49.92	8.48	-0.10±0.02	2.257	49.14	8.67	-0.10±0.02
ONSA	2.391	46.63	8.36	0.12±0.02	2.390	46.29	8.91	0.19±0.02
PENC	2.351	52.85	8.38	0.30±0.02	2.350	52.10	8.74	0.45±0.02
POTS	2.378	49.15	8.59	0.23±0.02	2.377	48.92	8.77	0.41±0.02
RAMO	2.153	13.72	8.49	0.33±0.01	2.153	13.72	8.49	0.33±0.01
REYK	2.352	42.43	7.02	0.36±0.02	2.352	42.45	6.59	0.18±0.02
RIGA	2.391	52.79	10.02	0.50±0.02	2.390	52.72	10.39	0.50±0.02
SFER	2.423	24.23	4.58	0.19±0.02	2.423	24.21	4.12	0.18±0.02
SJDV	2.313	45.73	6.45	0.13±0.02				
SODA	2.283	50.99	10.57	0.40±0.02				
SOFI	2.118	46.02	3.73	0.21±0.02				
SVTL	2.368	53.56	11.35	0.21±0.02	2.367	53.47	11.73	0.36±0.02
TERS	2.404	43.67	8.23	-0.27±0.02				
TORI	2.354	63.34	5.94	-0.03±0.02				
TRO1	2.328	49.49	8.80	0.00±0.02				
UNPG	2.342	46.92	1.49	-0.25±0.02				
VAAS	2.367	51.25	9.48	0.42±0.02				
VIL0	2.249	47.87	8.80	0.21±0.02				
VILL	2.255	28.91	0.78	-0.20±0.02	2.255	29.40	1.08	-0.16±0.02
VIS0	2.374	48.88	8.58	0.25±0.02				
WARE	2.374	44.54	8.05	-0.19±0.02	2.374	44.20	8.20	0.04±0.02
WROC	2.375	52.82	9.77	-0.14±0.02				
WSRT	2.399	45.72	8.36	-0.09±0.02				
WTZR	2.237	48.58	8.96	0.17±0.02	2.236	48.17	9.00	0.34±0.02
ZECK	2.095	53.85	6.75	0.43±0.02				
ZIMM	2.162	47.85	6.74	-0.04±0.02	2.162	47.37	7.01	0.12±0.02

References

- Bengtsson, L., S. Hagemann, and K.I. Hodges (2004), Can climate trends be calculated from reanalysis data? *J. Geophys. Res.* **109**, D11, D1111, DOI: 10.1029/2004JD004536.
- Bevis, M., S. Businger, T.A. Herring, C. Rocken, R.A. Anthes, and R.H. Ware (1992), GPS meteorology: Remote sensing of atmospheric water vapor us-

- ing the global positioning system, *J. Geophys. Res.* **97**, D14, 15787-15801, DOI: 10.1029/92JD01517.
- Bock, O., M.-N. Bouin, A. Walpersdorf, J.-P. Lafore, S. Janicot, F. Guichard, and A. Agustí-Panareda (2007), Comparison of ground-based GPS precipitable water vapour to independent observations and numerical weather prediction model reanalyses over Africa, *Q. J. Roy. Meteor. Soc.* **133**, 629, 2011-2027, DOI: 10.1002/qj.185.
- Bock, O., P. Willis, J. Wang, and C. Mears (2014), A high-quality, homogenized, global, long-term (1993-2008) DORIS precipitable water data set for climate monitoring and model verification, *J. Geophys. Res. – Atmos.* **119**, 12, 7209-7230, DOI: 10.1002/2013JD021124.
- Bruyninx, C. (2004), The EUREF Permanent Network: a multi-disciplinary network serving surveyors as well as scientists, *GeoInformatics* **7**, 5, 32-35.
- Byun, S.H., and Y.E. Bar-Server (2009), A new type of troposphere zenith path delay product of the international GNSS service, *J. Geodesy* **83**, 3-4, 367-373, DOI: 10.1007/s00190-008-0288-8.
- COST (2012), Memorandum of understanding for the implementation of a European Concerted Research Action, COST Action ES1206, Advanced Global Navigation Satellite Systems tropospheric products for monitoring severe weather events and climate (GNSS4SWEC), European Cooperation in Science and Technology.
- Dach, R., U. Hugentobler, P. Fridez, and M. Meindl (eds.) (2007), Bernese GPS software version 5.0, User manual, Astronomical Institute, University of Bern, Bern, Switzerland.
- Figurski, M., P. Kamiński, and A. Kenyeres (2009), Preliminary results of the complete EPN reprocessing computed by the MUT EPN Local Analysis Centre, *Bull. Geod. Geomatics* **1**, 163-174.
- Goosens, C., and A. Berger (1986), Annual and seasonal climatic variations over the northern hemisphere and Europe during the last century, *Ann. Geophys.* **4**, 4, 385-400.
- Guerova, G. (2013), Ground-based GNSS meteorology, Gfg² Summer School, 2 July 2013, Potsdam, Germany.
- Hagemann, S., L. Bengtsson, and G. Gendt (2003), On the determination of atmospheric water vapor from GPS measurements, *J. Geophys. Res.* **108**, D21, 4678, DOI: 10.1029/2002JD003235.
- Held, I.M., and B.J. Soden (2006), Robust responses of the hydrological cycle to global warming, *J. Climate* **19**, 21, 5686-5699, DOI: 10.1175/JCLI3990.1.
- Herring, T.A. (1992), Modeling atmospheric delays in the analysis of space geodetic data. In: J.C. de Munck, and T.A.T. Spoelstra (eds.), *Proc. Symp. Refraction of Transatmospheric Signals in Geodesy, 19-22 May 1992, Hague, The Netherlands*, 157-164.

- Hocke, K. (1998), Phase estimation with Lomb–Scargle periodogram method, *Ann. Geophys.* **16**, 3, 356-358.
- Jin, S., J.-U. Park, J.-H. Cho, and P.-H. Park (2007), Seasonal variability of GPS-derived zenith tropospheric delay (1994-2006) and climate implications, *J. Geophys. Res.* **112**, D9, D09110, DOI: 10.1029/2006JD007772.
- Karmeshu, N. (2012), Trend detection in annual temperature and precipitation using Mann Kendall test – A case study to assess climate change on select states in the Northeastern United States, M.Sc. Thesis, University of Pennsylvania, Philadelphia, USA.
- Kendall, M.G., and A. Stuart (1970), *The Advanced Theory of Statistics. Vol. 2: Interference and Relationship*, 3rd ed., Hafner, New York.
- Lomb, N.R. (1976), Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci.* **39**, 2, 447-462, DOI: 10.1007/BF006483.
- Mann, H.B. (1945), Nonparametric tests against trend, *Econometrica* **13**, 3, 245-259, DOI: 10.2307/1907187.
- Marini, J.W. (1972), Correction of satellite tracking data for an arbitrary tropospheric profile, *Radio Sci.* **7**, 2, 223-231, DOI: 10.1029/RS007i002p00223.
- 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.
- Niell, A.E. (1996), Global mapping functions for the atmospheric delay at radio wavelengths, *J. Geophys. Res.* **101**, B2, 3227-3246, DOI: 10.1029/95JB03048.
- Nilsson, T., and G. Elgered (2008), Long-term trends in the atmospheric water vapor content estimated from ground-based GPS data, *J. Geophys. Res.* **113**, D19, D19101, DOI: 10.1029/2008JD010110.
- Ning, T. (2012), GPS meteorology: with focus on climate application, Ph.D. Thesis, Department of Earth and Space Sciences, Chalmers University of Technology, Göteborg, Sweden.
- Pacione, R., B. Pace, and G. Bianco (2014), An homogeneously reprocessed Zenith Total Delay long-term time series over Europe. **In:** *EGU General Assembly, 27 April – 2 May 2014, Vienna, Austria*, id. 2945.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (1992), *Numerical recipes in Fortran*, 2nd ed., Cambridge University Press, Cambridge.
- Ross, R.J., and W.P. Elliott (2001), Radiosonde-based northern hemisphere tropospheric water vapor trends, *J. Climate* **14**, 7, 1602-1612, DOI: 10.1175/1520-0442(2001)014<1602:RBNHTW>2.0.CO;2.
- Schüler, T. (2001), On ground-based GPS tropospheric delay estimation, Ph.D. Thesis, Universität der Bundeswehr, München, Germany, 364 pp.
- Söhne, W., M. Figurski, and K. Szafranek (2010), Homogeneous Zenith Total Delay parameter estimation from European permanent GNSS sites, *Bull. Geod. Geomatics* **69**, 1, 11-22.

- Steigenberger, P., M. Rothacher, R. Dietrich, M. Fritsche, A. Rülke, and S. Vey (2006), Reprocessing of a global GPS network, *J. Geophys. Res.* **111**, B5, B05402, DOI: 10.1029/2005JB003747.
- van Malderen, R., H. Brenot, E. Pottiaux, S. Beirle, C. Hermans, M. de Mazière, T. Wagner, H. de Backer, and C. Bruyninx (2014), A multi-site intercomparison of integrated water vapour observations for climate change analysis, *Atmos. Meas. Tech.* **7**, 8, 2487-2512, DOI: 10.5194/amt-7-2487-2014.
- Wang, J., and L. Zhang (2009), Climate applications of a global, 2-hourly atmospheric precipitable water dataset derived from IGS tropospheric products, *J. Geodesy* **83**, 3-4, 209-217, DOI: 10.1007/s00190-008-0238-5.
- Yong, W., Y. Binyun, W. Debao, and L. Yanping (2008), Zenith Tropospheric Delay from GPS monitoring climate change of Chinese Mainland. **In:** *Int. Workshop on Education Technology and Training and on Geoscience and Remote Sensing, 21-22 December 2008, Shanghai, China*, Vol. 1, 365-368, DOI: 10.1109/ETTandGRS.2008.43.

Received 20 January 2015

Received in revised form 9 April 2015

Accepted 10 April 2015



Angular Distributions of Discrete Mesoscale Mapping Functions

Krzysztof KROSZCZYŃSKI

Military University of Technology, Faculty of Civil Engineering and Geodesy,
Warszawa, Poland; e-mail: kkroszczyński@wat.edu.pl

Abstract

The paper presents the results of analyses of numerical experiments concerning GPS signal propagation delays in the atmosphere and the discrete mapping functions defined on their basis. The delays were determined using data from the mesoscale non-hydrostatic weather model operated in the Centre of Applied Geomatics, Military University of Technology. A special attention was paid to investigating angular characteristics of GPS slant delays for low angles of elevation. The investigation proved that the temporal and spatial variability of the slant delays depends to a large extent on current weather conditions.

Key words: slant delay, mapping function, mesoscale model, anisotropy.

1. INTRODUCTION

Atmospheric refraction plays an important role in precise determination of position by means of Global Positioning System (GPS) techniques. The influence of atmospheric pressure, temperature, and first of all humidity slows the GPS signal down, diffuses it and changes its propagation direction. In consequence, these factors cause the propagation time between the satellite and the receiver to be longer than in the vacuum. These delays, called slant tropospheric delays or corrections, are mainly related with the lower layers of the atmosphere where weather phenomena occur. Since the GPS waves are non-dispersive in the troposphere, the slant delays cannot be eliminated

and they appear in the observational equations as additional unknowns. In a specific measurement epoch there are as many delays as there are observed GPS satellites. From the computational point of view, this situation is unfavourable and it results in multiplication of the number of the unknowns and usually in a bad conditioning (peculiarity) of the observation matrix. In practice, in order to reduce the number of the unknowns, the slant delays are expressed using zenithal delays (common for all satellites) computed subsequently with other geodesic parameters in the estimation process of observation systems solutions. The relation between these delays is realized using mapping functions dependent on the direction of the satellites observations. Just a few years ago, the available Bernese (Dach *et al.* 2007) and GAMIT-GLOBK (King and Bock 2005) software packages used mainly the Niell mapping functions (Niell 1996). The coefficients of these functions were determined on the basis of the solutions of the GPS waves propagation equation, the so-called eikonal equation, using selected profiles of the standard atmosphere model and verified using upper air sounding data. This indirect way of determining the slant delays, due to neglecting the current weather conditions, is burdened with a large systematic error, especially for satellites at low elevation angles. For this reason, operational possibility to determine the mapping functions using meteorological data from global weather predicting models was brought to attention (Niell 2000). The research presented in Boehm *et al.* (2006) was a continuation of Niell's work, and the European Centre for Medium-Range Weather Forecasts (ECMWF) model was used for this. The discrete mapping functions were defined for the global network and the international IGS services (International GNSS Service), IVS (International VLBI Service), and IDS (International DORIS Service). Introducing these mapping functions significantly improved the results of the satellite geodesic techniques and made it possible to realize the precise positioning of a single point (PPP) with sub-centimetre accuracy (Urquhart *et al.* 2011). Data from the mesoscale non-hydrostatic Coupled Ocean/Atmosphere Mesoscale Prediction System – Naval Research Laboratory (COAMPS–NRL; Hodur 1997, Bosy *et al.* 2010) model are used for the research presented in this paper. This model, unlike the global ones, due to better spatial resolutions of the computational grids, enables to monitor the temporal and spatial variability of the meteorological elements fields with better effectiveness, especially the humidity which is the main cause of difficulties encountered in GPS signal delay modelling. The values of the delays, especially for low elevation angles, may have significant influence on the solutions of the observation systems. For this reason, the proper determination of the delays is extremely important for accuracy of spatial geodesic techniques and it has significant influence on the accuracy of geodesic param-

ters computations (e.g., Bogusz *et al.* 2011, Urquhart *et al.* 2011, Wielgosz *et al.* 2012).

2. SLANT DELAY OF GPS SIGNAL

The considered slant delay $\Delta\tau^s$ is a result of integrating the refraction coefficient n of the atmosphere along the GPS signal path s from the satellite to the receiver (Fig. 1) according to Eq. 1. In vacuum the signal propagates along a straight line while in the atmosphere (according to the Fermat principle) along a curve line. Beside the curvature called the geometric elongation effect $\Delta\tau^g$, the signal is first of all slowed down. The slowdown effect (additional elongation) is called the electromagnetic effect $\Delta\tau^e$:

$$\Delta\tau^s = \underbrace{\int_{\text{sat}}^{\text{rec}} n(s) ds - \int_{\text{sat}}^{\text{rec}} 1 ds}_{\Delta\tau^e} + \underbrace{\int_{\text{sat}}^{\text{rec}} 1 ds - \int_{\text{sat}}^{\text{rec}} 1 dl}_{\Delta\tau^g} , \tag{1}$$

where $\Delta\tau^s$ is the total delay of the signal in the atmosphere (expressed in units of length, e.g., meters), and ds, dl are the differential increments of the s and l (straight line) paths' lengths of the GPS signal from the satellite (sat) to the receiver (rec).

The $\Delta\tau^g$ element defines the curvature and represents the geometric delay resulting from the difference in length of the signal paths calculated along the curve and the straight line connecting the receiver and the satellite. The geometric delay τ^g has significant contribution (decimetres) for low elevation satellites; it is less than 1 cm for elevation angles greater than 15° and it equals zero in the direction of the zenith (Mendes 1999).

Meteorological parameters generated by the COAMPS mesoscale model, *i.e.*, potential temperature, total pressure, and specific humidity (Figurski *et al.* 2007, 2009), were used for determining tridimensional refraction fields n_{xyz} . The fields were interpolated from the computational space of the mesoscale model (Lambert's Conformal Projection with the Gal-Chen

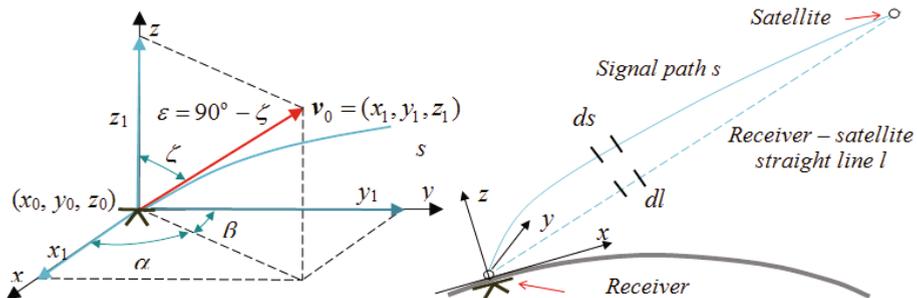


Fig. 1. The geometry of the GPS signal propagation.

and Somerville's nonlinear vertical coordinate) into the parallel planes of a local topocentric coordinate system defined by the GPS station position (Fig. 1). The Cartesian coordinate system used for realizing the computations simplifies preparation of the algorithm of the eikonal equation called also the ray Eq. 2 written in the form of ordinary differential equations system:

$$\frac{d}{ds} \left[n(\mathbf{r}) \frac{d\mathbf{r}}{ds} \right] = \nabla n(\mathbf{r}) \rightarrow \left\{ \frac{d\mathbf{r}}{ds} = \frac{1}{n(\mathbf{r})} \mathbf{v}, \frac{dv}{ds} = \nabla n(\mathbf{r}) \right\}, \quad (2)$$

where s is the path measured along the ray (Fig. 1) and defined by the vector $\mathbf{r} = [x(s), y(s), z(s)]$ with respect to the beginning of the topocentric reference system, \mathbf{v} is the vector tangential to the trajectory (ray) at point \mathbf{r} , and $n(\mathbf{r})$ is the refraction coefficient dependent on the radius vector \mathbf{r} .

The equation system 2 indicates that both the length of the path s and the value of the slant delay $\Delta\tau$ are dependent on the spatial heterogeneity of the refraction coefficient $n(\mathbf{r})$ defined by the gradient $\nabla n(\mathbf{r})$. The solution of the system 1 was formulated in the form of the following initial-value problem:

$$\begin{aligned} \mathbf{r} &= \mathbf{r}_0 \\ \mathbf{v}_0 &= (dx(s_0)/ds, dy(s_0)/ds, dz(s_0)/ds) = \cos \alpha, \cos \beta, \sin(90^\circ - \zeta). \end{aligned} \quad (3)$$

In Eq. 3 $\mathbf{r}_0 = [x(s_0), y(s_0), z(s_0)]$ is a vector of the GPS station antenna location, ζ is the zenithal angle, \mathbf{v}_0 is a versor tangential to the trajectory at point \mathbf{r}_0 which is determined by the directional cosines: $\cos \alpha, \cos \beta, \cos \zeta$ defining the direction to a selected point in the space, e.g., the satellite position. In order to determine signal delay of the particular satellite, shooting method (Hobiger *et al.* 2008) or collocation method (Zus *et al.* 2012) is used.

3. THE ALGORITHM OF DETERMINING THE SLANT DELAY

The path s of the ray and the slant delays $\Delta\tau$ related with it were determined using an iteration algorithm (Figurski *et al.* 2007, 2009). For every GPS station it uses atmospheric refraction fields defined in the topocentric systems (Fig. 1) on rectangular grids interpolated from irregular grids (of spatial resolution of 13 and 4.3 km in the Lambert's Conformal Projection) of the mesoscale model. The distribution of distances between the computational planes (levels) of these grids is not uniform. Distributions similar to that presented in Rocken *et al.* (2001) were used for the research. This solution saves the operational memory of the computer and shortens the time of computing while maintaining appropriate accuracy of the obtained values of the slant delay. It is also compatible with the distribution of atmospheric fields in real atmosphere, in the altitude reference system, and in mesoscale mod-

els. In the grids considered here, the cells are cuboids (Fig. 2) with different heights dh and constant lengths of the sides dx , dy equal to the distances between the nodes of the mesoscale model cells. The program for local three-line interpolation is introduced because of the applied procedures of solving the equations system 1 and thus the necessity of computing the values of refraction in any point of the computational cell (Fig. 2).

$$n_{xyz} = n_{000}(1-x)(1-y)(1-z) + n_{100}x(1-y)(1-z) + n_{010}(1-x)(1-y) + n_{101}x(1-y)z + n_{011}(1-x)yz + n_{110}xy(1-z) + n_{111}xyz \quad (4)$$

The variables (x, y, z) in Eq. 4 are normalized. The normalization is equivalent to an auxiliary transformation of the cuboids cells into unitary cubes (Fig. 2).

Polynomial form of the function n_{xyz} simplifies determination of gradient components n_{xyz} vector which are necessary for integrating realization procedure for the differential equations system 2 of the ray. The procedure determines the trajectory, *i.e.*, the GPS signal path. The classical constant step Runge–Kutta method of the fourth order (Ralson 1995) was used here. Identifications of the cells of the computations grid in which the determined point of the trajectory is placed (Fig. 2), is an important phase of the algorithm. For irregular grids, this identification is usually a complex process and computationally expensive. However, it is simple for the used rectangular grids. In this case, the cells are identified by the (i, j, k) indexes defined by:

$$(j, j, k) = \left(\text{floor}((x_0 + x^*)/dx), \text{floor}((y_0 + y^*)/dy), \text{floor}((z_0 + z^*)/dh) \right), \quad (5)$$

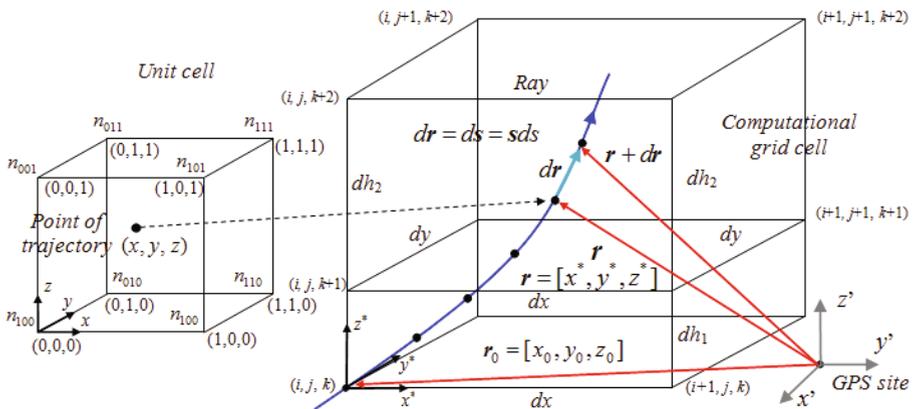


Fig. 2. Computational algorithm illustration – the three-line interpolation.

where “floor” is a function rounding the numbers down, and $\mathbf{r}_0 = [x_0, y_0, z_0]$ is the vector connecting the centre of the topocentric system with the local system (x^*, y^*, z^*) of the current cell of the grid (Fig. 2).

Procedure 5 may also be applied to regular spherical grids. The refraction value and the gradient 4 are computed for a new point of the trajectory after determining the point and identifying the cells of the grid (Fig. 2). The new position is then determined by means of the procedure of integration. The iteration process is continued until the modelled atmosphere is left. Recording the coordinates of the points of the trajectory and their related values of refraction n_{xyz} enables to compute the slant delay value $\Delta\tau^s$ (Eq. 1).

4. NUMERICAL EXPERIMENTS

4.1 Meteorological conditions

For the conducted basic research, every forecast from the model may be considered as the real state. However, having in mind future provision of mesoscale zenithal $\Delta\tau^z$ and slant $\Delta\tau^s$ delays for the users, as well as mapping functions $m = \Delta\tau^s/\Delta\tau^z$ determined on their basis, we have to be aware that the question of confronting the model results with the actual conditions, *i.e.*, appropriate monitoring and validation of the mesoscale model performance, will be inevitable. For this reason, investigating propagation of the GPS waves was related with periods in which the state of the atmosphere determined by the COAMPS mesoscale model correctly reflected the real atmospheric conditions.

To quantitatively verify the results, MSG 2 (Meteosat Second Generation) satellite images from the archives of the Dundee University, <http://www.sat.dundee.ac.uk/>, were used. Cloud systems related with atmospheric precipitation are observed in the IR7 (8.3-9.1 μm) images (Fig. 3). The satellite images are presented in the COAMPS model projection. It is observed that areas of precipitation and related cloud systems generated by the model are correlated with the cloud systems observed in the images. It is, among others, an effect of the observations data assimilation process which determines the initial conditions of the weather forecast. Partial quantitative verification of the atmosphere state, especially concerning the surface fields in the area of Poland, may be conducted using data provided by synoptic stations of the Institute of Meteorology and Water Management. If all sites were equipped with automatic systems measuring basic meteorological parameters, it would provide excellent auxiliary material from the Active Geodetic Network European Position Determination System (ASG EUPOS) network. Currently, EUREF Permanent Network (EPN) and the WAT1 sites are equipped with such stations.

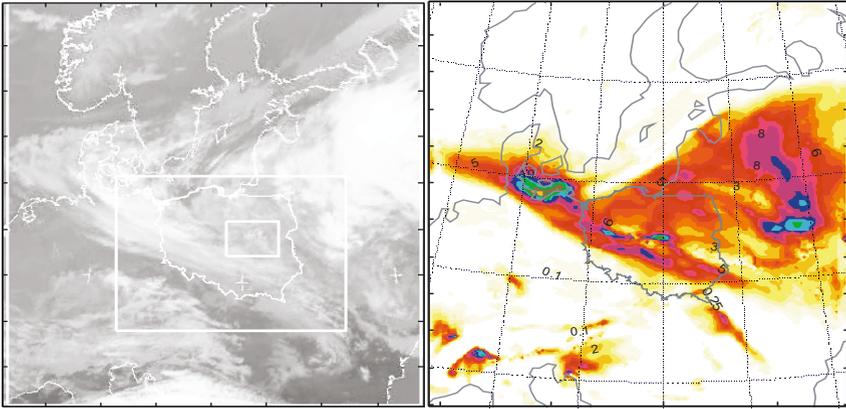


Fig. 3. Comparison of the cloud cover fields observed in the MSG 2 satellite images in the IR7 channel and the forecasted fields of precipitation for the 13 km computational grid of the COAMPS model at 6.00 on 17 April 2008.

4.2 The process of scanning the atmosphere

The process of scanning the atmosphere generated by the COAMPS mesoscale model was used in order to investigate the character of the slant delay spatial distribution (Figurski *et al.* 2009). The slant delays were determined according to the scheme presented in Fig. 4 for elevation angles ε within the range of $[3^\circ, 10^\circ]$ and $[15^\circ, 90^\circ]$ with increments of 1° and 5° , respectively, and for azimuths α within the range of $[0^\circ, 350^\circ]$ with increments of 10° . It takes into consideration $24 \times 36 = 864$ directions.

Such sampling of the space enables us to obtain appropriately accurate interpolated values of the delay for any direction. Due to the minimum value of the elevation angle, $\varepsilon = 3^\circ$, the computations could not be realized for all

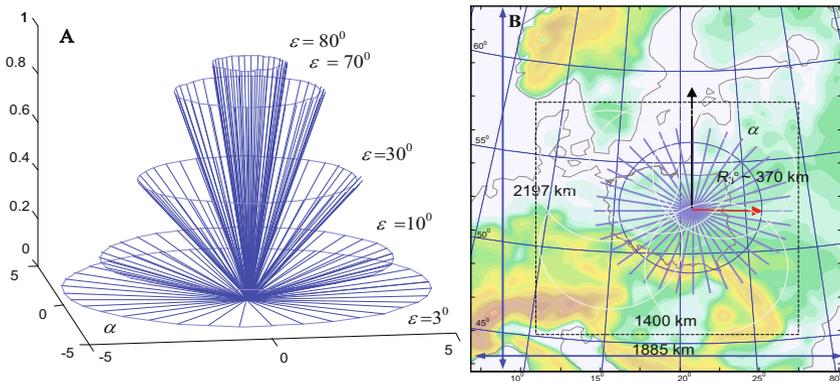


Fig. 4. The spatial process of scanning the atmosphere.

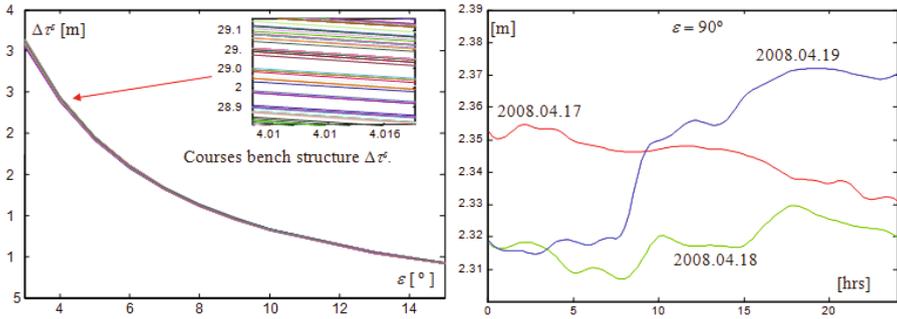


Fig. 5. Slant delay as a function of the elevation angles for fixed azimuth values and mesoscale model forecast time. Daily course of zenithal delay for the WAT1 site.

points of the mesoscale model space. For example, the area of 1885 by 2197 km (13 km grid), as presented in Fig. 4B, is appropriate for determining the full distributions of the slant delay $\Delta\tau^s$ for the rays inclined at 3° with respect to the plane of the horizon for all points belonging to the territory of Poland (except for those that are subject to visibility problems). The lengths of the rays projections R_z (Eq. 6) for the elevation angles ε (Fig. 4B) are defined by the (x, y) coordinates of the intersection of the sphere of $a + h$ radius and the straight line defined by the equation: $y = x \times \text{tg } \alpha$ and crossing the centre of the topocentric coordinate system:

$$R_z = \arctg(x/z) \leftarrow (y = x \times \text{tg } \varepsilon + a, x^2 + y^2 = (a+h)^2), \quad (6)$$

where $a = 6371.229$ km, and $h = 30$ km is the height of the COAMPS atmospheric model.

For $\varepsilon = 3^\circ$ Eq. 6 provides $R_{\varepsilon=3^\circ} \approx 370$ km. The R_ε parameter value usually influences the choice of the spatial resolution of the mesoscale model grid. Figure 4B shows that a square of the side length of about 1400 km is sufficient to determine $\Delta\tau^s$ for $R_{\varepsilon=3^\circ}$. The courses of slant delays $\Delta\tau^s$ presented in Fig. 5 as functions of the elevation angles for fixed azimuth values, mesoscale model forecast time and daily zenithal delays $\Delta\tau^z$ are the results of the computations.

4.3 Mesoscale mapping functions

Using the plots of Fig. 5, complete discrete mapping functions m_c may be derived which are dependent on GPS station position $x = (\lambda, \phi, h)$, time t , elevation ε , and azimuth α angles determined by the process of scanning the mesoscale model atmosphere:

$$m_c = (\varepsilon_j, \alpha_j, t_k, x) = \frac{\Delta\tau^s(\varepsilon_j, \alpha_j, t_j, x)}{\Delta\tau^s(\varepsilon = 90^\circ, t_k, x)} = \frac{\Delta\tau^s(\varepsilon_j, \alpha_j, t_j, x)}{\Delta\tau^z(t_k, x)}, \quad (7)$$

$$i = [3^\circ, 4^\circ, \dots, 9^\circ] \cup [10^\circ, 15^\circ, \dots, 90^\circ], \quad j = 0^\circ, 10^\circ, 20^\circ, \dots, 350^\circ, \quad k = 1, 2, \dots, 24 \text{ h},$$

where $\Delta\tau^z = \Delta\tau^s(\varepsilon = 90^\circ)$ is the zenithal delay.

In case of investigating the influence of water vapour content on GPS signals propagation or a reciprocal task, *e.g.*, GPS tomography of the atmosphere using the complete mapping function m_c , it is possible to extract its wet m_w and hydrostatic m_h parts. Determination of m_w and m_h (Rocken *et al.* 2001, Urquhart *et al.* 2011) is realized using the following relations:

$$\begin{aligned} n &= n_w + n_h \rightarrow (\Delta\tau_w^s, \Delta\tau_w^z) \\ (\Delta\tau_w^s, \Delta\tau_w^z) &\rightarrow (m_w = \Delta\tau_w^s / \Delta\tau_w^z, m_h = \Delta\tau_h^s / \Delta\tau_h^z), \end{aligned} \quad (8)$$

where $\Delta\tau_h^s, \Delta\tau_w^s, \Delta\tau_h^z, \Delta\tau_w^z, n_w, n_h$ are dry and wet parts of the slant $\Delta\tau^s$ and zenithal $\Delta\tau^z$ delays and of the refraction coefficient n .

For the $t = 0$ epoch, the m_c functions are computed using the analysis of meteorological fields, while for $t \neq 0$ using forecasts obtained from the mesoscale model. The analysis is based on real data which have to be collected from the meteorological data exchange network for the main synoptic times (00, 06, 12, 18 UTC). The delay related with it causes that operational application of m_c is feasible for ultra-short, *e.g.*, 6-hourly weather forecasts. Equation 8 shows that for the measurement epoch t , the m_c function differs from the $\Delta\tau^s$ by the multiplication factor which is a reciprocal of the $\Delta\tau^z$ delay. Due to relatively small values of changes of $\Delta\tau^z$ (Fig. 5), investigating m_c or $\Delta\tau^s$ in the range of low elevation angles ε is practically equivalent. For this reason, amplitudes m_c and $\Delta\tau^s$ are proportional, and their angular characteristics are identical. The functions described by Eq. 8 vary from the simplified ones routinely used in GPS analyses by anisotropy of the spatial distribution. The simplifications are first of all related to the assumed independence of azimuth α . The research presented below indicates that such an assumption is justified only for appropriately large elevation angles assumed in practice as the cut-off angles ($\varepsilon > 10^\circ$).

5. ANALYSIS OF SLANT DELAY DISTRIBUTIONS

Investigating the mapping functions 7 is not simple because of their multidimensionality. Therefore, the analysis method of projections on sub-spaces of various combinations of the $\varepsilon, \alpha, t, x$ parameters was used for this purpose. For example, the projection on the ε, α sub-space is equivalent to fixing the position x of the GPS site and fixing the time t of observation. The

first plot (Fig. 5) illustrates the relation 9 between the slant delay $\Delta\tau^s$ and the elevation angle ε for constant value of azimuth α and time t for the WAT1 reference site. The other plot (Fig. 5) presents one-dimensional temporal courses of the zenithal delay described by Eq. 10. They reflect, among others, the changes of meteorological conditions (*e.g.*, air humidity) above the observation point.

$$\Delta\tau^s = \Delta\tau^s(\varepsilon, \alpha = \text{const}, t = \text{const}, x), \quad (9)$$

$$\Delta\tau^s = \Delta\tau^s(\varepsilon = 90^\circ, \alpha = \text{const}, t, x = x_{\text{WAT1}}). \quad (10)$$

From a practical point of view, the sets of values of function 7 for elevation and azimuth angles determined by the varying in time position of GPS satellites are important for the GPS solutions. Generally, the slant delays $\Delta\tau^s$ for a selected GPS station are determined by means of fast methods of ray tracking according to the scanning scheme presented in Fig. 4. Forecasted refraction fields of the mesoscale model determined in one hour intervals are used in the computations. Tables 1 and 2 contain examples of angular distributions of slant delays for 17 and 19 April 2008, averaged with respect to azimuth and time (day).

The sets include the minimum $\Delta\tau_{\min}^s(\varepsilon)$, maximum $\Delta\tau_{\max}^s(\varepsilon)$, and average $\overline{\Delta\tau^s(\varepsilon)}$ values and the standard deviations of the delays for the elevation angles ε considered in the computations. They are determined for 900-element samples of data. The averaging $\overline{\Delta\tau^s(\varepsilon)}$ was done with respect to a set of azimuths α from the range of $(0^\circ, 360^\circ)$ with 10° increments. Analysis of the results obtained for two selected days indicates that for the elevation angle $\varepsilon = 3^\circ$, the azimuth differences of the delays are up to 1.1 m (on 17 April) and 1.6 m (on 19 April). They decrease with increasing ε , reaching 14 and 28 cm for the elevation angle of $\varepsilon = 15^\circ$. Additional comparison of estimated values of standard deviations $\sigma_{\Delta\tau^s(\varepsilon)}$ shows that the distribution for 19 April has greater spatial variability than the distribution for 17 April (Fig. 6). Therefore, it may be concluded that the GPS weather defined by the spatial variability and irregularity was worse for the meteorological conditions observed at 00 UTC on the second of the considered days (Table 2). It is interesting that the conclusion might be opposite if the average values $\overline{\Delta\tau^s(\varepsilon, \alpha)}$ were considered alone. In this case, the maximum difference is only 5 cm. The presented example illustrates the dissimilarity of the anisotropic mesoscale mapping functions developed in the research and the isotropic NMF or VMF1 type functions (Niell 1996, Boehm *et al.* 2006).

Table 1

Angular distribution of slant delays for the WAT1 reference station,
17 April 2008, 00 UTC

ε	$\Delta\tau_{\min}^s(\varepsilon)$ [m]	$\overline{\Delta\tau^s(\varepsilon)}$ [m]	$\Delta\tau_{\max}^s(\varepsilon)$ [m]	$\sigma_{\Delta\tau(\varepsilon)}$ [m]
3°	35.35	36.00	36.45	0.21
4°	28.79	29.16	29.46	0.14
5°	24.11	24.39	24.62	0.11
6°	20.69	20.91	21.09	0.09
7°	18.10	18.29	18.44	0.07
8°	16.08	16.24	16.37	0.06
9°	14.47	14.61	14.73	0.06
10°	13.16	13.29	13.39	0.05
13°	10.38	10.48	10.56	0.04
15°	9.14	9.22	9.28	0.03

Table 2

Angular distribution of slant delays for the WAT1 reference station,
19 April 2008, 00 UTC

ε	$\Delta\tau_{\min}^s(\varepsilon)$ [m]	$\overline{\Delta\tau^s(\varepsilon)}$ [m]	$\Delta\tau_{\max}^s(\varepsilon)$ [m]	$\sigma_{\Delta\tau(\varepsilon)}$ [m]
3°	35.00	35.96	36.60	0.38
4°	28.53	29.14	29.61	0.29
5°	23.93	24.39	24.75	0.24
6°	20.53	20.91	21.22	0.21
7°	17.96	18.29	18.55	0.18
8°	15.96	16.25	16.48	0.16
9°	14.36	14.62	14.82	0.14
10°	13.06	13.29	13.47	0.13
13°	10.31	10.49	10.63	0.10
15°	9.07	9.23	9.35	0.09

The plots of the delay differences and their average values for each hour of the forecast:

$$\Delta\tau_r^s = \Delta\tau^s - \overline{\Delta\tau^s}, \quad \overline{\Delta\tau^s} = \overline{\Delta\tau^s(\varepsilon, \alpha, t, x_{\text{WAT1}})}, \quad t = 0, 1, \dots, 24 \text{ h} \quad (11)$$

are presented in Fig. 6, where x_{WAT1} is the WAT1 reference site position vector.

Application of the differences presents in a clearer way the dependence of delays on the elevation ε and azimuth α angles. Observing them in the abso-

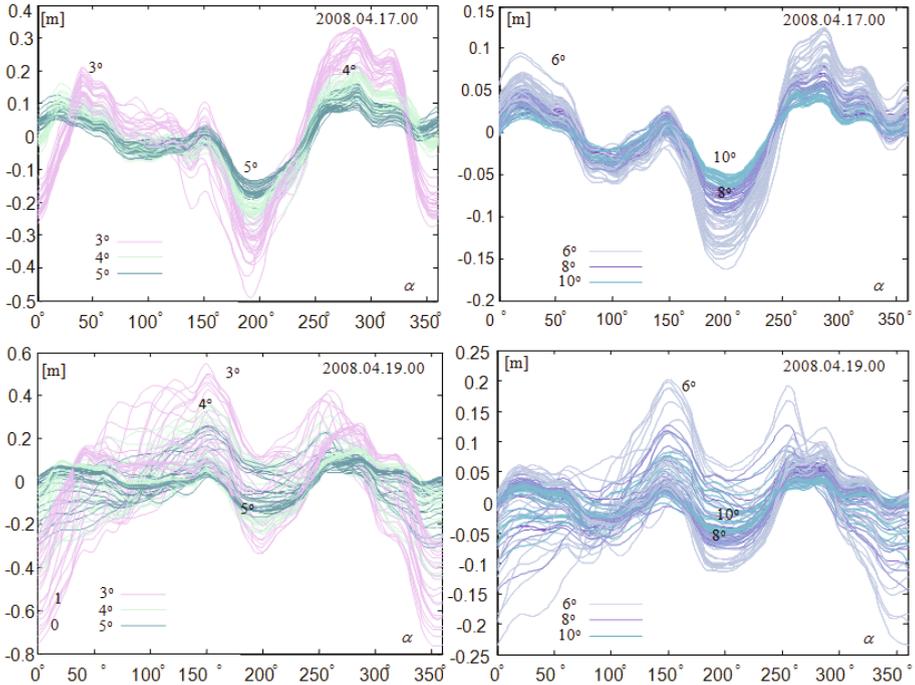


Fig. 6. Differences of slant delays as functions of azimuth and elevation angles for the WAT1 site.

lute courses is more difficult because the disturbances (differences) amplitude for the elevation angle of 5° is about 50 times smaller than the average value. The plots of Fig. 6 may also be presented in the polar form which is more comfortable for analysis.

Figures 7-12 present azimuth distributions of differences of slant delays $\Delta\tau_r^s$ for the WAT1 site for elevation angles $\varepsilon = 3^\circ, 4^\circ, 5^\circ$. They were constructed using the following equations:

$$\begin{aligned}
 x &= |\Delta\tau_r^s(\cdot, \alpha)| \cos \alpha, \quad y = |\Delta\tau_r^s(\cdot, \alpha)| \sin \alpha, \\
 \Delta\tau_r^s(\cdot, \alpha) &= \overline{\Delta\tau^s(\cdot, \alpha)} - \Delta\tau^s(\cdot, \alpha), \\
 \Delta\tau^s(\cdot, \alpha) &= \overline{\Delta\tau^s(\varepsilon \in \{3^\circ, 4^\circ, 5^\circ\}, \alpha, t \in \{1, 24\}, x_{\text{WAT1}})},
 \end{aligned}
 \tag{12}$$

where $\Delta\tau^s(\cdot, \alpha)$, $\overline{\Delta\tau^s(\cdot, \alpha)}$ is the discrete course and its average with respect to azimuth (Fig. 6) for each hour ($t = 0, 1, \dots, 24$) of the daily forecast; x, y are coordinates of radius vectors of the $|\Delta\tau_r^s(\cdot, \alpha)|$ lengths determined for the set of azimuths: $\alpha \in [0^\circ, 360^\circ)$.

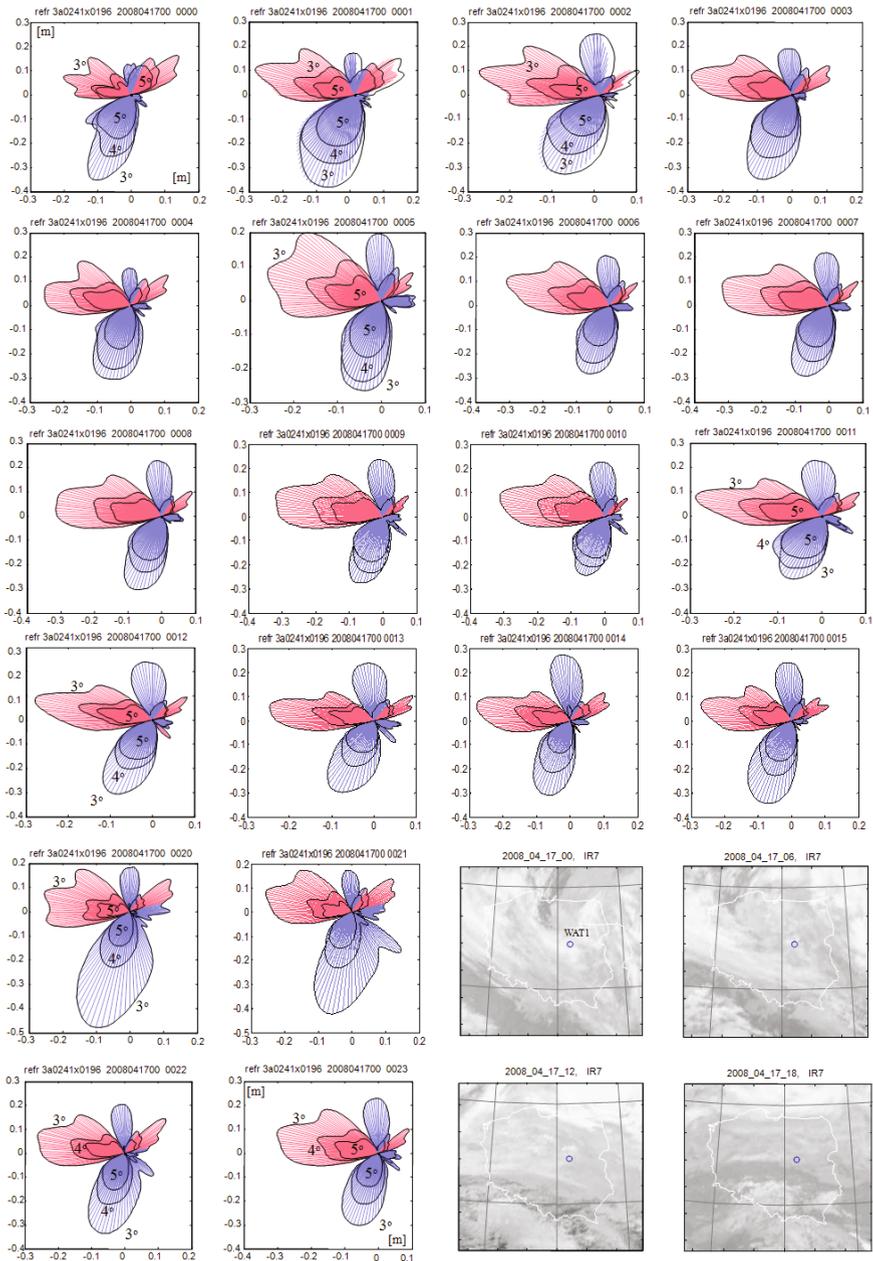


Fig. 7. Anisotropic distributions of the differences of the slant delay $\Delta\tau_r^s = \Delta\tau^s - \overline{\Delta\tau^s}$ ($\overline{\Delta\tau^s}$ is the average with respect to the azimuth) for a daily forecast on 17 April 2008. The synoptic situation is presented in MSG 2 IR7 satellite images. Units: [m].

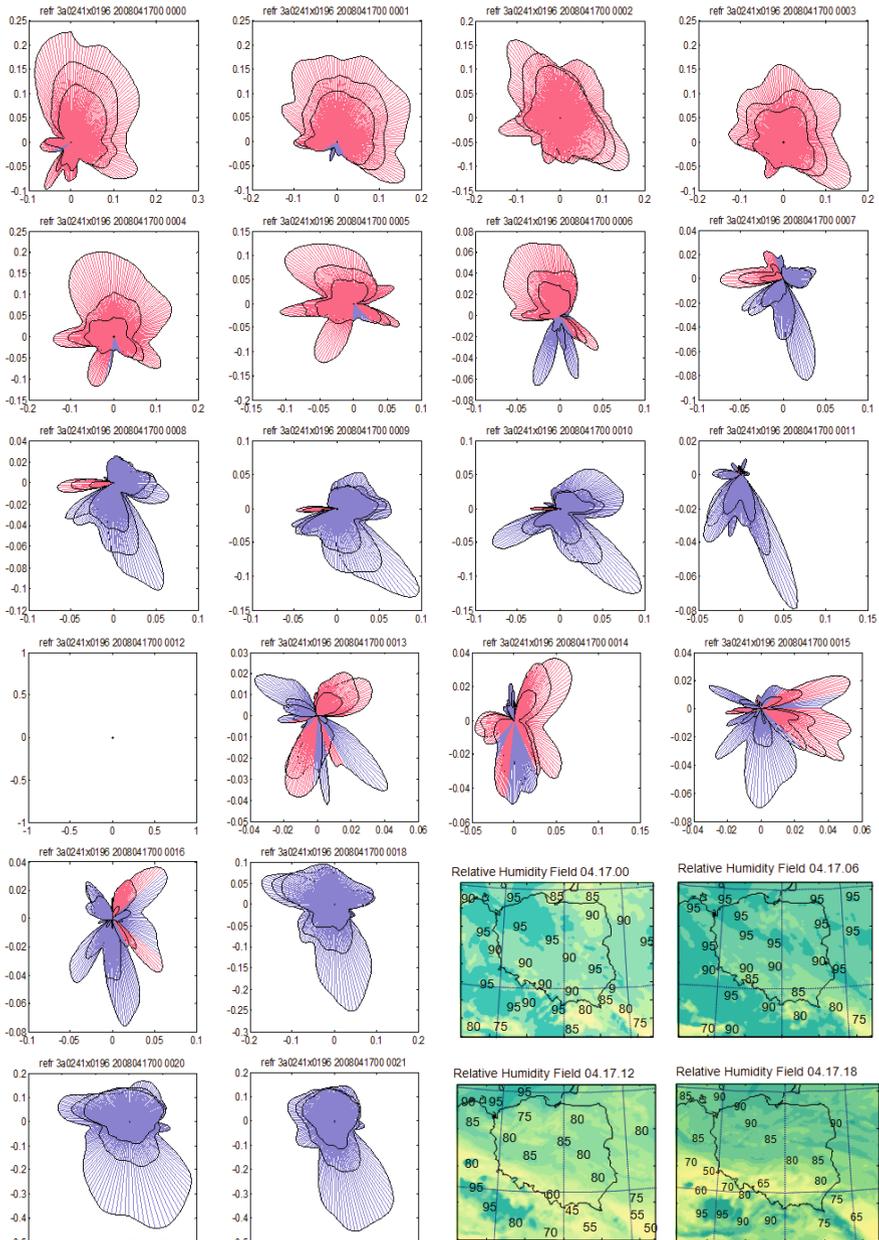


Fig. 8. Azimuth distributions of the differences of the slant delay $\Delta\tau_r^s = \Delta\tau^s - \Delta\tau^s(\cdot, t_{12})$ ($\Delta\tau^s(\cdot, t_{12})$ is the course for the 12 hour forecast) on 17 April 2008. The synoptic situation is presented in relative humidity fields. Units: [m].

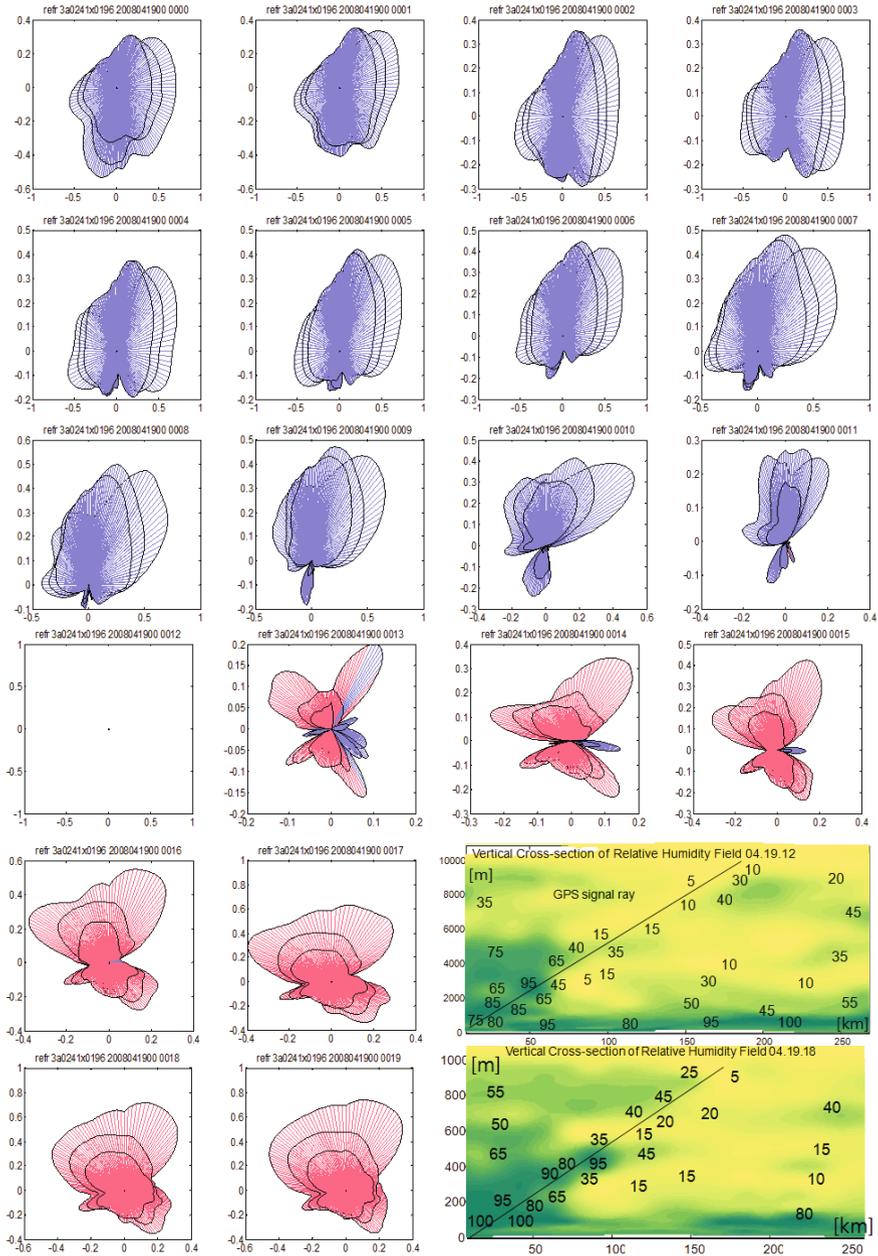


Fig. 10. Anisotropic distributions of the differences of the slant delay $\Delta\tau_r^s = \Delta\tau^s - \Delta\tau^s(\cdot, t_{12})$ at subsequent hours of the daily forecast on 19 April 2008. The GPS signal rays for the elevation angle of 3° are superimposed over vertical cross-sections of the humidity field. Units: [m].

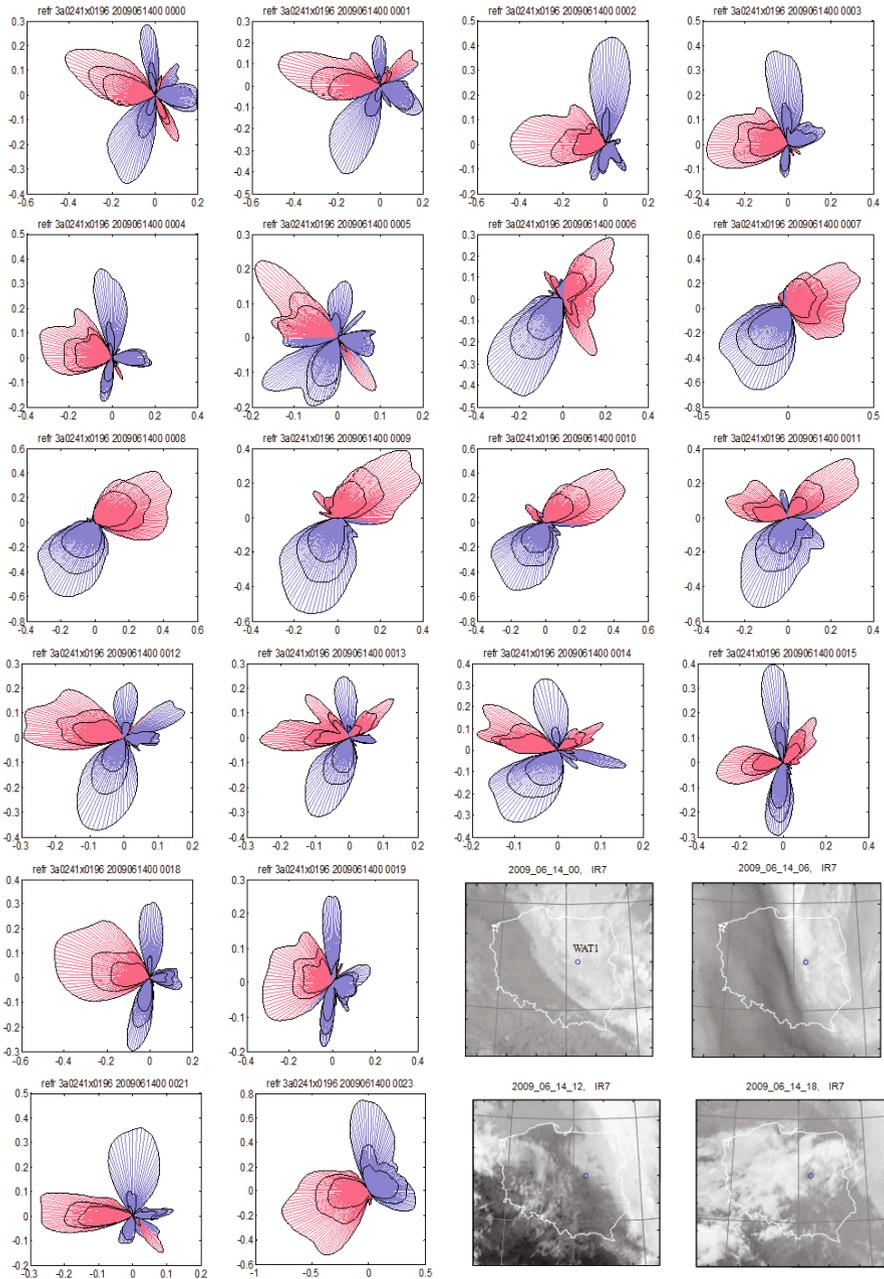


Fig. 11. Forecasted distributions of differences of slant delay $\Delta\tau_r^s = \Delta\tau^s - \overline{\Delta\tau^s}$ for 14 June 2009. The weather condition related with an atmospheric front is presented in MSG 2 satellite images. Units: [m].

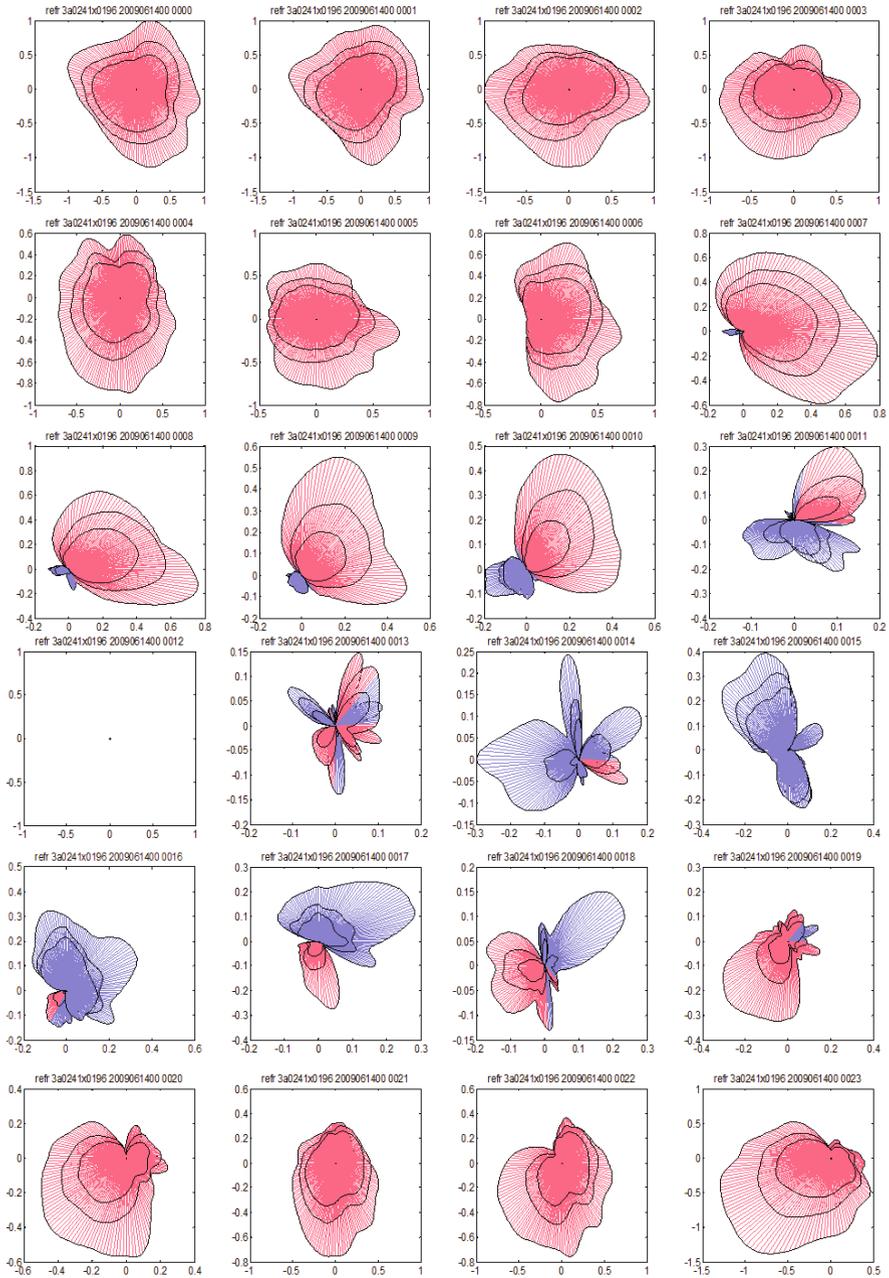


Fig. 12. Relative distributions of slant delay: $\Delta\tau_r^s = \Delta\tau^s - \Delta\tau^s(\cdot, t_{12})$ for 14 June 2009. Units: [m].

Distributions presented in Figs. 8, 10, and 12 were also obtained using Eq. 12 but for differences $\Delta\tau^s(\cdot, \alpha)$ defined by

$$\Delta\tau_r^s(\cdot, \alpha) = \Delta\tau^s(\cdot, \alpha) - \Delta\tau^s(\cdot, \alpha, t_{12}). \quad (13)$$

In this equation, $\Delta\tau^s(\cdot, \alpha, t_{12})$ corresponds to the distribution of slant delay at 12 UTC. The plots of the type as in Fig. 8 present differences between the azimuth course at a specific hour of forecast and the course at 12 UTC for each of the elevation angles ($\varepsilon = 3^\circ, 4^\circ, 5^\circ$). Red/blue colour indicates greater/smaller values than the averaged discrete course with respect to azimuth of the slant delay $\Delta\tau^s$. Additionally, MSG 2 geostationary satellite images from selected spectral channels, meteorological elements fields for the computational surfaces of the mesoscale model (*e.g.*, relative humidity), and vertical cross-sections indicating the conditions of the GPS signals propagation are included. Equations 12 and 13 do not use up all the possibilities of investigating the spatial distributions of slant delays. For example, $\overline{\Delta\tau^s}$ for given elevation angles may be azimuth daily series obtained as a result of averaging 24-hour forecasted courses. In this case, stronger variability of the distributions should be expected (Tables 1 and 2, and Fig. 13) because the daily average usually deviates significantly from the hourly averages used in Eq. 12. Other information related with the possibility of interpolating the delays for a receiver located in the neighbourhood of reference sites, *e.g.*, ASG-EUPOS network, may be obtained by means of determination of spatial differences of delays between the network GPS sites for the same measurement epoch.

Analysis of the azimuth distributions of slant delays presented in Fig. 7 shows that they are similar. Such a situation, occurring for each hour of the forecast, indicates the azimuth stationarity of the weather conditions and at the same time the stationarity of the signal propagation conditions in the atmosphere. At this time, the WAT1 reference site was under cloud cover of an atmospheric front moving over Poland and in the beginning under influence of characteristic north-westerly advection of an air mass. Positive (red) and negative (blue) values of differences of delays slightly fluctuate, showing changes of the order of a decimetre. It is also observed that the directions of their extremes do not change significantly. Figure 8 shows a greater spatial and temporal variability of the delays distributions. In the first hours of the forecast, the differences are positive, reaching the values of the order of two decimetres for the elevation angle of 3° . It means an increase of the length of the signal propagation path in the atmosphere in comparison with the one for 12 UTC which determines the reference epoch. In the following hours, the signs and the directions of the differences change, while their am-

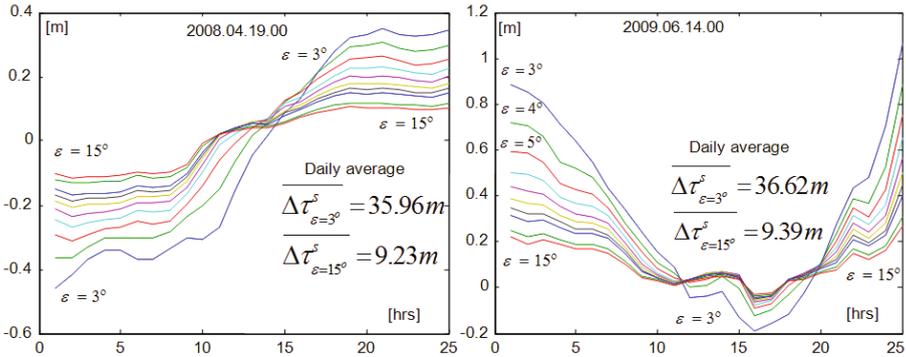


Fig. 13. Averaged azimuth differences of hourly and daily slant delays.

plitudes decrease below one decimetre. They are negative and much bigger (about 0.5 m) after 16 hours of forecast which means improvement in the propagation properties of the atmosphere. On 19 April 2008 (Fig. 9), the values and the changes of directions of the slant delay differences are bigger than on 17 April 2008 all day long. The positive component in the westerly direction decreases starting at 00 UTC. At the same time, the south-easterly component appears and develops reaching about 0.5 m. Then, in the afternoon hours of the forecast, the scenario is opposite. The plots in Fig. 10 indicate better transmission properties of the atmosphere in the first part of the day, and worsening of the properties in the following part. This is partly justified by cloud systems observed in the included satellite images in the vicinity of the WAT1 site. A similar situation, presented in Fig. 11, is observed for distributions investigated for comparison for 14 June 2009. In general, it may be concluded that positive differences (Figs. 7, 9, and 11) are usually correlated with directions indicating existence of a developed system of layer clouds, *e.g.*, frontal system, which is related with areas of larger quantities of water vapour, while negative values correspond to GPS waves propagation through cloudless areas of the atmosphere. Distributions presented in Fig. 11 show (in comparison with the above-discussed ones) greater spatial and temporal variability of positive and negative directions of slant delay differences. They reflect in a good way the motion of cloud structures over the WAT1 site, as observed in satellite images. Investigation of angular characteristics (Fig. 12) shows that the differences values of the slant delays with respect to the reference delay of 12 UTC are almost 1 m in the first and last hours of the forecast. The distributions presented in Figs. 7 through 12 reflect relative spatial and temporal changes of slant delays resulting from the atmospheric conditions evolution in the vicinity of the WAT1 site. Similar characteristics may be obtained for all reference sites of the ASG-EUPOS

system. Generally, they may be obtained for each point of the computational area of the mesoscale model which fulfils the criterion of feasibility of scanning (Eq. 12). It is said that they determine the conditions of microwaves propagation in the atmosphere, *i.e.*, the GPS weather, for the considered area and time of the model data archiving. Function 12 used for investigating the heterogeneousness of the slant delays is based on the differences of the absolute courses and average courses corresponding to them. Application of such a method enables to detect the azimuth differences. However, it does not contain all the information. It is the case because the averages with respect to the azimuth determined for each hour of the forecast change in time. Figure 13 presents the temporal courses for various elevation angles referred to daily azimuth averages, *i.e.*, computed using 24 values of hourly averages.

On 19 April 2008 the differences increased while on 14 June 2009 they decreased for some of the forecast time, and then they increased exceeding the initial values. In the analysed cases, the changes of the average delays for small elevation angles exceed 1 m. The time courses (Fig. 13) also reflect well the character of temporal changes of differences of the distributions presented in Figs. 10 and 12. All the considered courses and angular distributions indicate existence of a usually neglected azimuth relation. The discrepancy from symmetry may be illustrated in a cylindrical system by means of transferring two-dimensional distributions (Figs. 7 through 12) to tridimensional space (Fig. 14). Horizontal cross-sections (contour lines) of the constructed surfaces correspond to the azimuth distributions for fixed elevation angles ε , while vertical cross-sections correspond to profiles of differences of slant delays for a selected value of azimuth α .

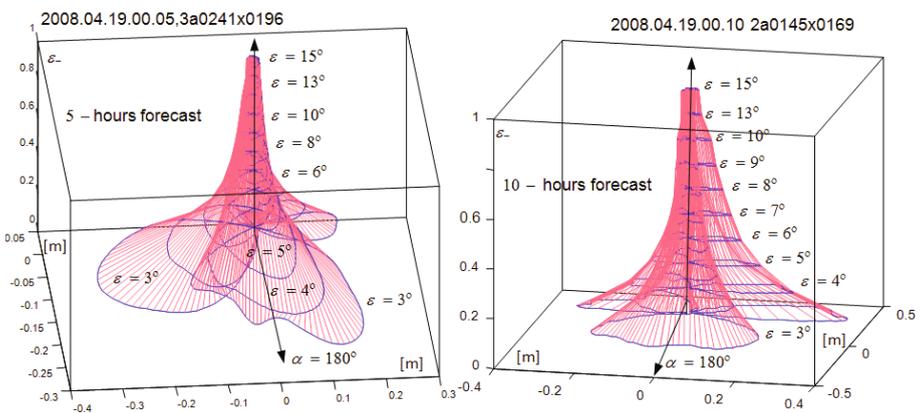


Fig. 14. Tridimensional anisotropic distributions of differences of slant delays and their hourly azimuth average values for the WAT1 reference site; α – azimuth, ε_- – arbitrary value of the elevation angle ($\varepsilon_- = 0 \rightarrow \varepsilon = 0^\circ$, $\varepsilon_- = 1 \rightarrow \varepsilon = 90^\circ$).

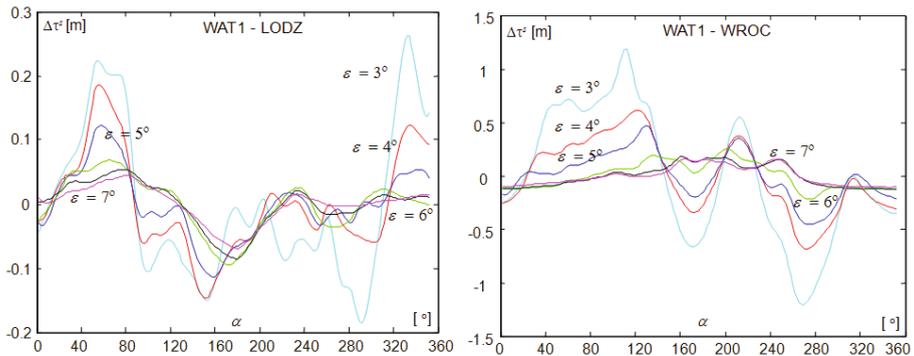


Fig. 15. Differences of slant delays between reference sites.

For appropriately large elevation angles ($\varepsilon > 20^\circ$), the function contour lines (Fig. 14) become circles (for $\varepsilon = 90^\circ$) they convert to a point. It means that for measurements at such values of the cut-off angles, the azimuth asymmetry is negligible. Tridimensional discrete mapping functions may be obtained by adding to each of the contour lines of the selected distributions the hourly averages corresponding to them, and then dividing the values by zenithal delays. They are determined by the positions of GPS reference sites for measurement epochs related with subsequent hours of the forecast. These sets of functions (delays) enable us to analyse their spatial differences (Fig. 15).

The differences of slant delays (Fig. 15) between the WAT1 and the WROC reference sites are a few times larger than those for the WAT1 and the LODZ sites which means that the propagation conditions in the atmosphere are dramatically different for those sites.

6. SUMMARY

The conducted research proved that non-hydrostatic mesoscale model forecast data enable to model spatial and temporal heterogeneousness of atmospheric refraction fields which may be used for investigating and simulating GPS waves propagation in the atmosphere. The results of computational experiments showed that the slant delay is a function of the azimuth of observation for small values of elevation angles. This relation reflects spatial heterogeneousness of the atmospheric state including the distribution of humidity along the GPS signal path. It was observed that the azimuth asymmetry (anisotropy) of the slant delays reaches the value of 1 m and it may be even larger in unfavourable weather conditions. Defining differential angular characteristics enabled us to present and investigate the anisotropic distributions of the slant delay in a better way. Application of the mesoscale model

data enabled us to determine them for various atmospheric conditions. A more precise explanation of heterogeneousness of the obtained distributions and their spatial and temporal evolution requires further research related with, *e.g.*, increasing the number of computational levels of the weather models, especially in the lower, humid part of the troposphere.

Acknowledgments. This research was financed by the Faculty of Civil Engineering and Geodesy of the MUT statutory research funds.

References

- Boehm, J., B. Werl, and H. Schuh (2006), Troposphere mapping functions for GPS and very long baseline interferometry from European Centre for Medium-Range Weather Forecasts operational analysis data, *J. Geophys. Res.* **111**, B2, B02406, DOI: 10.1029/2005JB003629.
- Bogusz, J., M. Figurski, K. Kroszczyński, and K. Szafranek (2011), Investigation of environmental influences to the precise GNSS solutions, *Acta Geodyn. Geomat.* **8**, 1, 5-15.
- Bosy, J., W. Rohm, A. Borkowski, K. Kroszczyński, and M. Figurski (2010), Integration and verification of meteorological observations and NWP model data for the local GNSS tomography, *Atmos Res.* **96**, 4, 522-530, DOI: 10.1016/j.atmosres.2009.12.012.
- Dach, R., U. Hugentobler, P. Fridez and M. Meindl (2007), Bernese GPS software version 5.0, Astronomical Institute, University of Bern, Bern, Switzerland.
- Figurski, M., M. Gałuszkiewicz, P. Kamiński, and K. Kroszczyński (2007), Mesoscale functions of GPS slant delay, *Acta Geodyn. Geomater.* **4**, 4, 141-144.
- Figurski, M., M. Gałuszkiewicz, P. Kamiński, and K. Kroszczyński (2009), Mesoscale anisotropy of GPS slant delay, *Bull. Geod. Geomatics* **2**, 99-110.
- Hobiger, T., R. Ichikawa, Y. Koyama, and T. Kondo (2008), Fast and accurate ray-tracing algorithms for real-time space geodetic applications using numerical weather models, *J. Geophys. Res.* **113**, D20, D20302, DOI: 10.1029/2008JD010503.
- Hodur, R.M. (1997), The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS), *Mon. Weather Rev.* **125**, 7, 1414-1430, DOI: 10.1175/1520-0493(1997)125<1414:TNRLSC>2.0.CO;2.
- King, R.W., and Y. Bock (2005), Documentation for the GAMIT GPS processing software, Release 10.2, Massachusetts Institute of Technology, Cambridge, USA.

- Mendes, V.B. (1999), Modeling the neutral atmosphere propagation delay in radiometric space techniques, Ph.D. Thesis, Tech. Rep. 199, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, Canada.
- Niell, A.E. (1996), Global mapping functions for the atmosphere delay at radio wavelengths, *J. Geophys. Res.* **101**, B2, 3227-3246, DOI: 10.1029/95JB03048.
- Niell, A.E. (2000), Improved atmospheric mapping functions for VLBI and GPS, *Earth Planets Space* **52**, 10, 699-702, DOI: 10.1186/BF03352267.
- Ralston, A. (1995), *Introduction to Numerical Analysis*, PWN, Warszawa (in Polish).
- Rocken, C., S. Sokolovskiy, J.M. Johnson, and D. Hunt (2001), Improved mapping of tropospheric delays, *J. Atmos. Ocean. Technol.* **18**, 7, 1205-1213, DOI: 10.1175/1520-0426(2001)018<1205:IMOTD>2.0.CO;2.
- Urquhart, L., F.G. Nievinski, M.C. Santos (2011), Ray-traced slant factors for mitigating the tropospheric delay at the observation level, *J. Geodesy* **86**, 2, 149-160, DOI: 10.1007/s00190-011-0503-x.
- Wielgosz, P., J. Paziewski, A. Krankowski, K. Kroszczyński, and M. Figurski (2012), Results of the application of tropospheric corrections from different troposphere models for precise GPS rapid static positioning, *Acta Geophys.* **60**, 4, 1236-1257, DOI: 10.2478/s11600-011-0078-1.
- Zus, F., M. Bender, Z. Deng, G. Dick, S. Heise, M. Shang-Guan, and J. Wickert (2012), A methodology to compute GPS slant total delays in a numerical weather model, *Radio Sci.* **47**, RS2018, DOI: 10.1029/2011RS004853.

Received 14 January 2015

Received in revised form 3 April 2015

Accepted 10 April 2015



Ionospheric and Solar Wind Variation during Magnetic Storm Onset and Main Phase at Low- and Mid-latitudes

Bolarinwa J. ADEKOYA¹ and Babatunde O. ADEBESIN²

¹Department of Physics, Olabisi Onabanjo University, Ago Iwoye, Nigeria
e-mail: adekoya.bolarinwa@oouagoiwoye.edu.ng (corresponding author)

²Department of Physical Sciences, Landmark University, Omu-Aran, Nigeria
e-mail: f_adebesin@yahoo.co.uk

Abstract

The relationship between the F2-layer critical frequency and solar wind parameters during magnetic storm sudden commencement (SSC) and main phase periods for intense (IS) and very intense (VIS) class of storms is investigated. The analysis covers low- and mid-latitude stations. The effects of ionospheric storm during SSC period is insignificant compared to the main phase, but can trigger the latter. The main phase is characterized by severe negative storm effect at both latitudes during VIS periods while it is latitudinal symmetric for IS observations. The IS reveal positive/negative storm phase in the low-/mid-latitudes, respectively. Ionization density effect is more prominent during VIS events, and is attributed to large energetic particle and solar activity input into the earth magnetosphere. However, ionospheric effect is more significant at the low-latitude than at the mid-latitude. Lastly, ionospheric storm effect during a geomagnetic storm may be related to the combinational effect of interplanetary and geomagnetic parameters and internal ionospheric effect, not necessarily the solar wind alone.

Key words: F2 layer critical frequency, SSC, very intense storm, intense storm, driver gas, ionization density.

1. INTRODUCTION

The effects of magnetic storms on the ionosphere are complex and deviate greatly from average ionospheric behavior. The global distribution of ionospheric storm effects is also complicated and differs considerably from one storm to another. These disturbances sometimes take the form of increases/decreases of the F-layer critical frequency (f_oF_2), and are referred to as the positive/negative ionospheric storms, respectively. The propagation of negative and positive ionospheric storms is strongly determined by the thermospheric disturbance gravity waves (Liu *et al.* 2010). According to Kane (2005), positive storm effects are a result of downwelling of neutral atomic oxygen and uplifting of the F-layer due to winds. Both of these rely on large scale changes in the thermospheric circulation caused by heating in the auroral zone. The storm negative phase in f_oF_2 and total electron content (TEC) occurs in a composition disturbance zone which reaches lower latitudes in summer than in winter, and has a preference for the night and morning sectors due to the local time variation of neutral winds.

The reaction of the ionosphere as seen at different ionospheric stations may be quite different during the same storm period depending on the station coordinates, local time effects, and some other parameters (Danilov 2001, Akala *et al.* 2010, Vijaya *et al.* 2011, Adebessin *et al.* 2013). At low latitude and equatorial zone, $E \times B$ drifts are affected by prompt penetration of magnetospheric convection electric fields, as well as by long-lived dynamo electric fields from the disturbance neutral winds and storm-related changes in ionospheric conductivity (Fejer 1997). In addition, changes in the neutral composition alter the balance between production and loss in the plasma and this affects the peak density of the ionosphere. An increase in the percentage of molecular neutrals, as would be the case if the thermosphere were heated, would lead to depletion in the ionospheric density (Davis *et al.* 1997). Such effects have been reproduced by coupled models of the ionosphere-thermosphere system.

Adekoya *et al.* (2012a) had demonstrated that investigations into the origin and nature of SSC on the ionospheric F-region have continued to engage the attention of scientists who constitute the space weather community. Some of these include Prölss (1995), Mikhailov and Perrone (2009), Burešová and Laštovička (2007, 2008), and Danilov (2013). While some researchers believed that the concept of SSC is delusion, others assumed that it is a reality. Danilov (2001) listed the SSC enhancements as one of the open problems of F-region physics and suggested that perhaps soft particle precipitation in the dayside cusp or magnetospheric electric field penetration might play a role in this phenomenon. Recently, Danilov (2013) affirmed and reported the appearance of ionospheric SSC enhancement in approximately

25-30% of ordinary storms, and in almost all prominent storms. Chukwuma (2007) asserted that the difficulty with the explanation of these phenomena is because in the studies of ionospheric storms it is assumed that the beginning of the disturbance is defined by storm sudden commencement (SSC) or main phase onset (MPO), which as a scheme restricts the geoeffectiveness of the solar wind to post-onset time. These foreclose the explanation of any aspect of the morphology of ionospheric storms whose origin precedes the onset reference time. Mikhailov and Perrone (2009) assumed a criterion for selecting SSC phenomenon which is that an SSC f_oF2 enhancement should precede the magnetic storm onset and take place within a reasonable time interval before the SSC and develop under quiet geomagnetic conditions. If an observed f_oF2 increase does not satisfy this requirement, there is no reason to consider it an SSC enhancement. This work therefore attempts to find the relationship between ionospheric and solar wind phenomena during SSC and main phase of geomagnetic storms at some selected low- and mid-latitude stations.

2. DATA AND METHODOLOGY

The geomagnetic index and solar wind data used consist of hourly UT values of the low latitude magnetic index Dst [nT], the solar wind flow speed V [km/s], the southward interplanetary magnetic field component (IMF B_z component [nT]), the plasma flow pressure P [nPa], the proton number density [N/cm^3], the plasma temperature K , plasma beta, and interplanetary electric field [mV/m^2]. These data were obtained from the National Space Science Centre's NSSDC OMNIWeb Service (<http://nssdc.gsfc.nasa.gov/omniweb>).

In like manner, the ionospheric data used are hourly UT values of f_oF2 obtained from Space Physics Interactive Data Resource (SPIDR's) network (<http://spidr.ngdc.noaa.gov>) of ionosonde stations located in the low- and mid-latitude regions. These stations are located in the Australian (Darwin, Learmonth), Euro-African (Grahamstown, Juliusruh/Rugen), and American sectors (Boulder). Their coordinates are listed in Table 1.

In order to contribute to the solution of the controversial problem of ionospheric F2 and geomagnetic storm phenomena during SSC event that leads to main phase disturbances, six intense geomagnetic storms were investigated: four very intense ($Dst \leq -250$ nT) and two intense (-100 nT $\leq Dst < -250$ nT) (e.g., Gonzalez *et al.* 2002, Adebessin 2008) storms at low and low-mid latitude (Table 1). It is important to note that paucity of data at some stations during the days under investigation restricted the choice of ionosonde stations. However, the normalized deviation of the critical frequency f_oF2 from the reference, which is used to denote the F2 region response to a geomagnetic activity, is given by

Table 1

List ionosonde stations used with their geomagnetic and geographic coordinates

Name of the ionospheric station	Station code	Geographic latitude and longitude	Geomagnetic latitude and longitude	Difference between LST and UT [hrs]
Darwin	DW41K	-12.5, 131.0	-22.90, 202.66	+9
Learmonth	LM42B	-21.9, 114.0	-34.15, 185.02	+8
Grahamstown	GR132	-33.3, 26.5	-33.92, 89.37	+2
Boulder	BC840	40.0, -105.3	48.90, 316.50	-7
Juliusruh/Rugen	JR055	54.6, 13.4	54.25, 99.73	+1

$$D(f_oF2) = \frac{f_oF2 - (f_oF2)_{ave}}{(f_oF2)_{ave}} \times 100\% \quad , \quad (1)$$

$D(f_oF2)$ variation is described in terms of percentage change in amplitude of the f_oF2 from the reference. According to Liu *et al.* (2008), positive and negative storms occur when the absolute maximum value of $D(f_oF2)$ exceeds 20%. Further, this limit is sufficiently large to prevent inclusion of random perturbation and disturbances of gravity waves, thereby making the indicated positive and negative storms represent real changes in electron density and not just redistribution of the existing plasma.

The data analyzed consist of $D(f_oF2)$ of respective hourly values of f_oF2 for 6-8 April 2000, 15-17 July 2000, 16-18 September 2000, 30 March – 1 April 2001, 10-12 April 2001, and 23-25 August 2005. The reference for each hour for each storm event is the average value of f_oF2 for that hour calculated from the four quiet days preceding the storm, based on the magnetic activity index ($A_p < 26$ corresponding to disturbance storm time $Dst \geq -25$) (Adeniyi 1986). Also these days are devoid of not only any significant geomagnetic activity, but also an absence of any considerable solar activity. This followed from the fact that high solar flare activity results in ionospheric disturbances due to their effects on thermospheric neutral density (Sutton *et al.* 2006).

3. RESULTS

3.1 Signature of interplanetary and geomagnetic parameters with ionospheric variations

3.1.1 Storm of 6 April 2000

Figure 1 presents the interplanetary, geomagnetic, and ionospheric observations for the 5-9 April 2000 storm event covering the initial, main, and re-

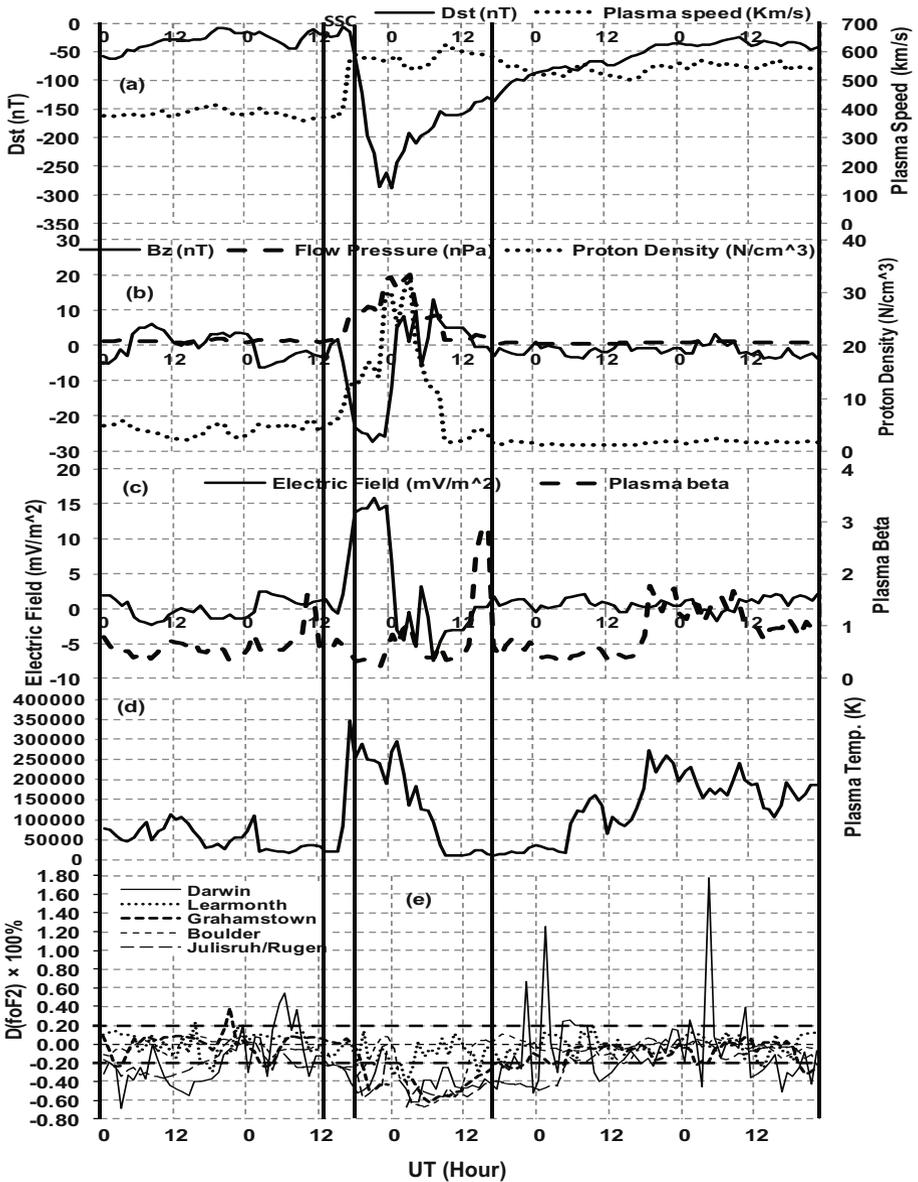


Fig. 1. Geomagnetic and interplanetary variations with $D(f_0F_2)$ response during storm-time period for low- and mid-latitude stations. The storm spans 5-9 April 2000. The thick horizontal dashed lines depict the disturbed time reference level on the $D(f_0F_2)$ plot.

covery phases, respectively. Figure 1a presents the geomagnetic index Dst (left side) and solar wind speed (right side). The storm is summarized using

available interplanetary data. However, Vieira *et al.* (2001) classified geomagnetic storm as weak (when $Dst > -50$ nT), moderate (when -100 nT < peak $Dst \leq -50$ nT), and intense (when $Dst < -100$ nT). From the early hour of 5 April the storm emerged with a moderate to weak amplitude activity up to around 12:00 UT on 6 April. Following this is the period expected to be dominated by sudden storm commencement phenomena (Balasis *et al.* 2006, Adekoya *et al.* 2012a), but the decrease in the westward ring current encircling the earth did not significantly enhanced the H component (little energy is entering into the earth magnetosphere). However, the Dst value was observed to be -6 nT around the onset period, which indicates that the storm is not preceded by SSC. The Dst thereafter decreases to a minimum peak value of -288 nT at 00:00 UT of the main phase. This coincided with the maximum plasma speed increase of 625 km/s at 09:00 UT. Also, this period of minimum Dst was observed to have an increasing plasma flow pressure with maximum peak value of 20.34 nPa at 03:00 UT (Fig. 1b). B_z was initially southwardly oriented with a magnitude of -27.3 nT at 21:00 UT before a northward orientation, which coincides with the Dst minimum peak. The proton density was initially decreased from the early hour of 5 April with decreasing average value of 3.50 N/cm³ which lasted for about a day and a half (Fig. 1b). The high electric field experienced during the main phase coincides with the B_z southward turning, as well as plasma temperature. The increase in plasma pressure and electron density consequently increased the plasma temperature, as well as a lifting in the electric field drift. This temperature rise (Fig. 1d) during the main phase corresponds to increase in plasma beta of about 2.92 magnitudes (Fig. 1c). It transpires from the high plasma beta and temperature values that the onset period was followed by ejecta, which may not be magnetic cloud type. According to Dal-Lago *et al.* (2004), the interaction of the high stream and ejecta results in an increase in speed, density, and temperature.

The ionospheric F2 effect observed over the low- and mid-latitude stations are presented in Fig. 1e using a normalized deviation of the critical frequency, $D(f_oF2)$, during the storm progression. Highlighted in the plot is the region for sudden storm commencement/storm onset (mark SSC), main phase (the region between the second and the third vertical dashed lines across the plot), and then the recovery phase. The horizontal dashed lines show the disturbed reference level. Discernible changes in the $D(f_oF2)$ were observed at the onset, main and recovery phases of the storm, respectively, which may be the consequence of geomagnetic storm effects. Here, the maximum ionization density of the F2 layer serves as a convenient parameter to specify the ionospheric behaviour. However, ionization density may either increase or decrease during the disturbed conditions, and these changes are designated as positive and negative ionospheric storms (P - and N -storm), re-

spectively. The ionosphere at Darwin emerges with a negative ionospheric disturbance (N -storm) from 00:00 UT on 6 April with peak electron density variation value of 34% at 03:00 UT. Thereafter, a sharply enhancement (positive -54%) was observed around 06:00 UT before a prolonged negative storm which extended to the SSC period. The intense negative phase observation at SSC was trailed by significant negative phase storm during the main and the recovery phase was largely controlled by the intense positive storm. Similarly, Learmonth with insignificant negative ionospheric storm variation during the storm-onset period was followed by noticeable negative storm effect at post-midnight and post-noon periods with percentage magnitudes of 43 and 31% at 02:00 and 14:00 UT during the main phase. The ionospheric storm response over Grahamstown was insignificant during the marked SSC period, while the main phase was largely predominated by an intense negative phase storm as a result of the depletion in $D(f_oF2)$. The ionosphere over Boulder and Juliusruh/Rugen shows similar $D(f_oF2)$ morphology with that observed over Learmonth. The main phase at these stations was completely controlled by negative storm effect which extended into the recovery phase. It has been reported that changes in neutral composition alter the balance between production and loss in the plasma, and subsequently affects the peak density of the ionosphere. If the thermosphere were heated, the molecular neutrals percentage would increase, and would result in the depletion of the ionospheric density (Davis *et al.* 1997).

3.1.2 Storm of 15 July 2000

Depicted in Fig. 2 is the interplanetary and geomagnetic response with the corresponding ionospheric observation during a very intense storm (VIS) of 16 July 2000 (see discussion section for the storm classification). Figure 2a-d presents the interplanetary and geomagnetic observation while Fig. 2e presents the ionospheric F2 effect during the geomagnetic storm and spans 14-18 July 2000, representing the initial (storm-onset periods is embedded in the initial phase period of the storm), main and the recovery phases, respectively. The characteristic signature of this magnetic storm is the depression in magnitude of Dst value within 17:00 UT on 15 July to 11:00 UT of 16 July known as storm main phase (*i.e.*, the H component of magnetic field) (Fig. 2a). This depression was a result of the ring current encircling the Earth in the westward direction (*e.g.*, Kamide *et al.* 1998). The Dst value decreases greatly to a minimum value of -301 nT at 01:00 UT on 16 July. This corresponds to the increase in solar wind flow speed to a peak value of 1107 km/s. On the Dst plot is the picture of the sudden positive increase known as the sudden storm commencement (SSC), which is generally recently known as the SSC (see Balasis *et al.* 2006, Mikhailov and Perrone 2009, Adekoya *et al.* 2012a) and spans 12:00-17:00 UT. This period is observed to be a turning

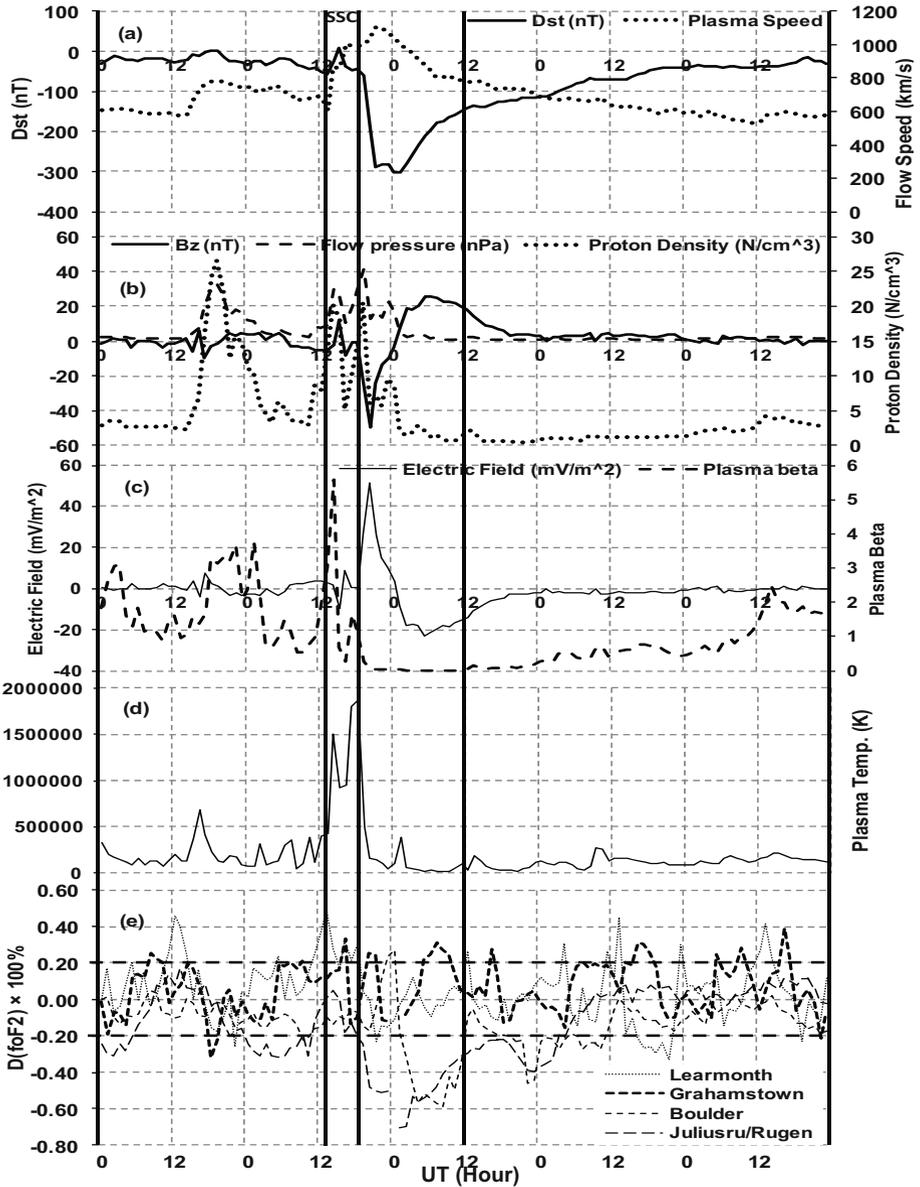


Fig. 2. Same as in Fig. 1 but for the storm of 14-18 July 2000.

point for the flow speed increase, as well as the southward orientation of IMF B_z (Fig. 2b). It is generally observed during this period that all the observed parameters responded with sudden changes. This is a result of the ef-

fect of compression of the front side of the magnetosphere of enhanced solar wind pressure, and subsequently increased the plasma temperature to its peak value around 17:00 UT (Fig. 2d). Consequently, the temperature decreased together with plasma beta to low values of 19 192 K and 0.03, respectively. The electric field increase to its peak value coincides with the peak southward orientation of B_z (-49.4 nT) and proton density (20.6 N/cm³) at about 20:00 UT. This storm was followed by a slow recovery that lasted for over 10 hours. The storm is characterized as a magnetic cloud type, with a characteristic low plasma beta, relatively high density ratio of He⁺⁺ to proton (Wang *et al.* 2003), low plasma temperature, and slowly varying strong magnetic field B_z (Gonzalez *et al.* 1999, 2002, Adekoya *et al.* 2012a).

The consequence of this storm on the ionospheric F2 layer over the low- and mid-latitude stations is presented in Fig. 2e. The horizontal dashed lines show the disturbed reference level of both positive and negative storm phases. Between the first two vertical lines is the period marked SSC. We observed that the ionospheric F2 electron density during the SSC period at Learmonth and Grahamstown was enhanced with an average f_oF_2 value of about 32%. Boulder and Juliusruh are in the Northern Hemisphere. Their $D(f_oF_2)$ variation was depleted significantly compared to the Southern Hemispheric stations (Learmonth and Grahamstown) with weak positive storm appearance during the main phase. The observed ionospheric storm event during the main phase may have originated from the mass input of energetic particle that change the daytime eastward electric field at this stations. The main phase depletion could be attributed to composition changes, which directly influence the electron concentration in the F2 region. In other words, the ionospheric storm intensity is more pronounced during the main phase, and may have arise from more penetration of the connective electric field and neutral wind (*e.g.*, Foster and Rich 1998).

3.1.3 Storm of 17 September 2000

The solar wind parameters and magnetic index observation for the intense geomagnetic storm that spans 15-19 September 2000 are illustrated in Fig. 3a-d. This storm had a minimum Dst excursion of -201 nT. The storm-onset occurred within 12:00-21:00 UT on 17 September. The Dst amplitude started decreasing from around 19:00 UT and reached its minimum amplitude of -201 nT at 23:00 UT. It was accompanied with a steep drop and polarity reversal in IMF B_z (Fig. 3b). It turns southward and attained maximum negative value of -23.9 nT at 21:00 UT. Afterwards, there was a steep rise noticed in the northward direction. During the main phase, the solar wind speed reached a peak of 839 km/s, with a corresponding increase in flow speed during the SSC period. The compression in the earth magnetosphere subsequently increased the flow pressure and the temperature to a peak

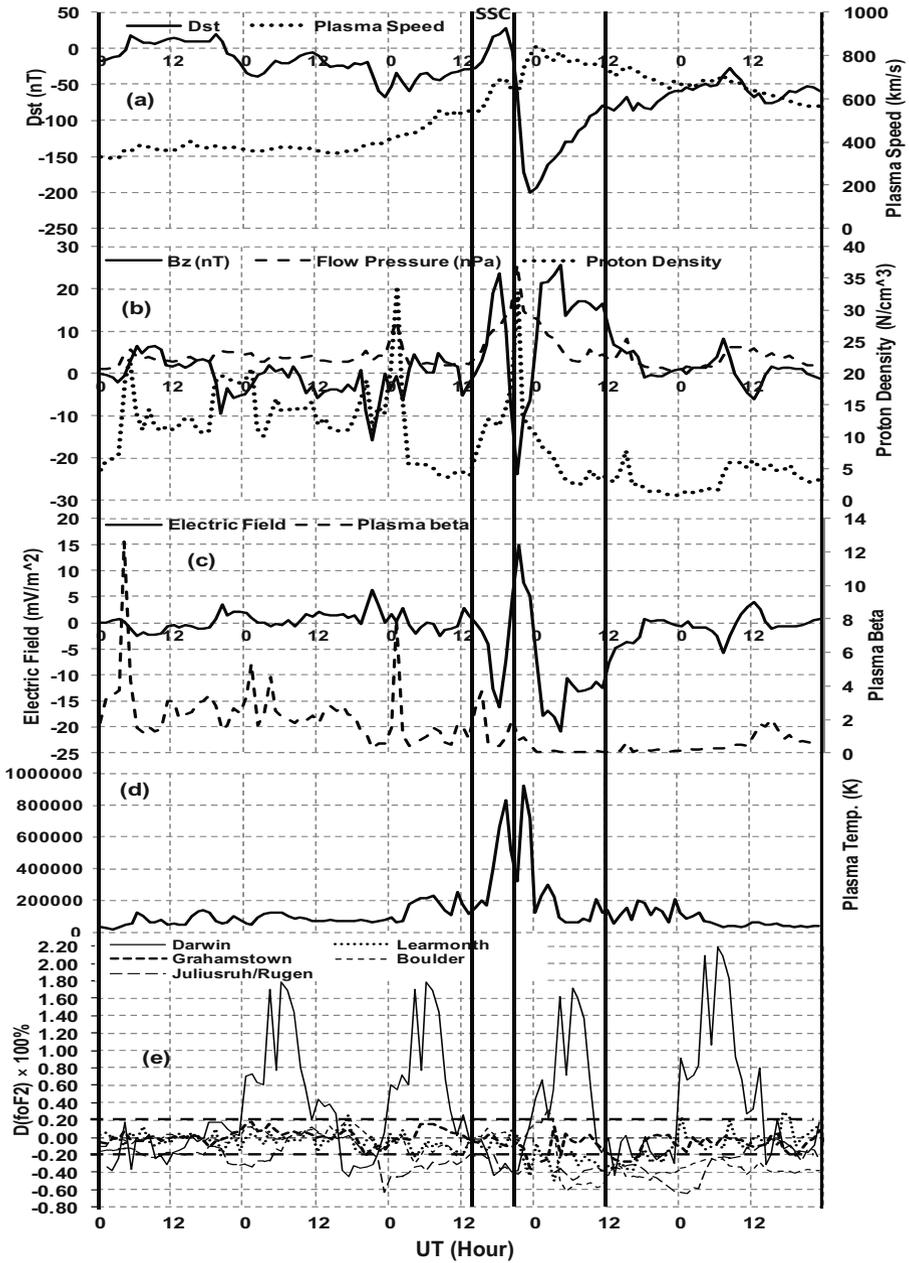


Fig. 3. Same as in Fig. 1 but for the storm of 15-19 September 2000.

value; which in effect increased the proton number density during the main phase. The electric field activity during the SSC period is observed to turn

Table 2

The distribution of storms according to their southward IMF B_z source

Storm date	Peak Dst [nT]	V [km/s]	B_z [nT]	Plasma temperature [K]	Plasma beta	Proton density [cm^{-3}]	Source of storm event
6 Apr 2000	-288	625	-27.3	L	1.06	33.4	sheath
16 Jul 2000	-301	1107	-49.4	L	0.03	9.7	cloud
17 Sep 2000	-201	839	-23.9	H	0.93	32.8	sheath
31 Mar 2001	-387	821	-44.7	H	1.61	25.2	sheath + ejecta
11 Apr 2001	-271	687	-20.5	H	1.22	24.2	sheath + cloud
24 Aug 2005	-216	721	-38.3	H	6.29	29.6	sheath + ejecta

Explanations: H – high plasma temperature (*i.e.*, $T \geq 400\,000$ K), L – low plasma temperature (*i.e.*, $T < 400\,000$ K).

southward with a reduction in magnitude up till a peak value of -16.28 mV/m² during the SSC period (Fig. 3c). It thereafter rotates northward and reached a peak value of 14.19 mV/m² during the main phase boundary. It is overwhelming that the solar wind dawn-to-dusk electric field directly drives magnetospheric convection. These electric fields are caused by a combination of increase solar wind velocity and southward IMF. This storm event was observed to be a sheath B_z source storm event (see Table 2 and Section 4 for more details).

The ionospheric observations to this storm are presented in Fig. 3e. The onset period occurred during the daytime of 17 September. The significant positive disturbance is majorly noticed at Darwin throughout the studied storm period, the SSC responded with depletion in electron density which was trailed by an enhanced $D(f_oF2)$ value of 168% around 05:00 UT at the main phase. The electron density variation for Learmonth, Grahamstown, Boulder, and Juliusruh/Rugen were disturbed with negative storm effect during the main phase compared to Darwin. Only Learmonth during the SSC was enhanced with noticeable storm disturbances. However, the ionospheric response over the low-latitude stations was more significant than it appears over the low latitude than in the mid-latitude of the Southern Hemisphere. Prölss (1993) had suggested that at low latitudes, the energy dissipation of the two traveling ionospheric disturbances (TIDs) launched in both hemispheres causes an increase in the upper atmosphere temperature and in the gas densities. These TIDs may be connected to the excessive and impulsive energy input in the high latitude during the storm that dissipates to the equator with high speed. The noticeable feature is that the intense main phase ionospheric storm emerged from SSC and more pronounced in the Northern Hemisphere over the mid-latitude stations.

3.1.4 Storm of 31 March 2001

Figure 4 presents the superposed plot of *Dst*, plasma flow speed, the IMF *Bz*, and other solar wind parameters together with the corresponding *D*(foF2)

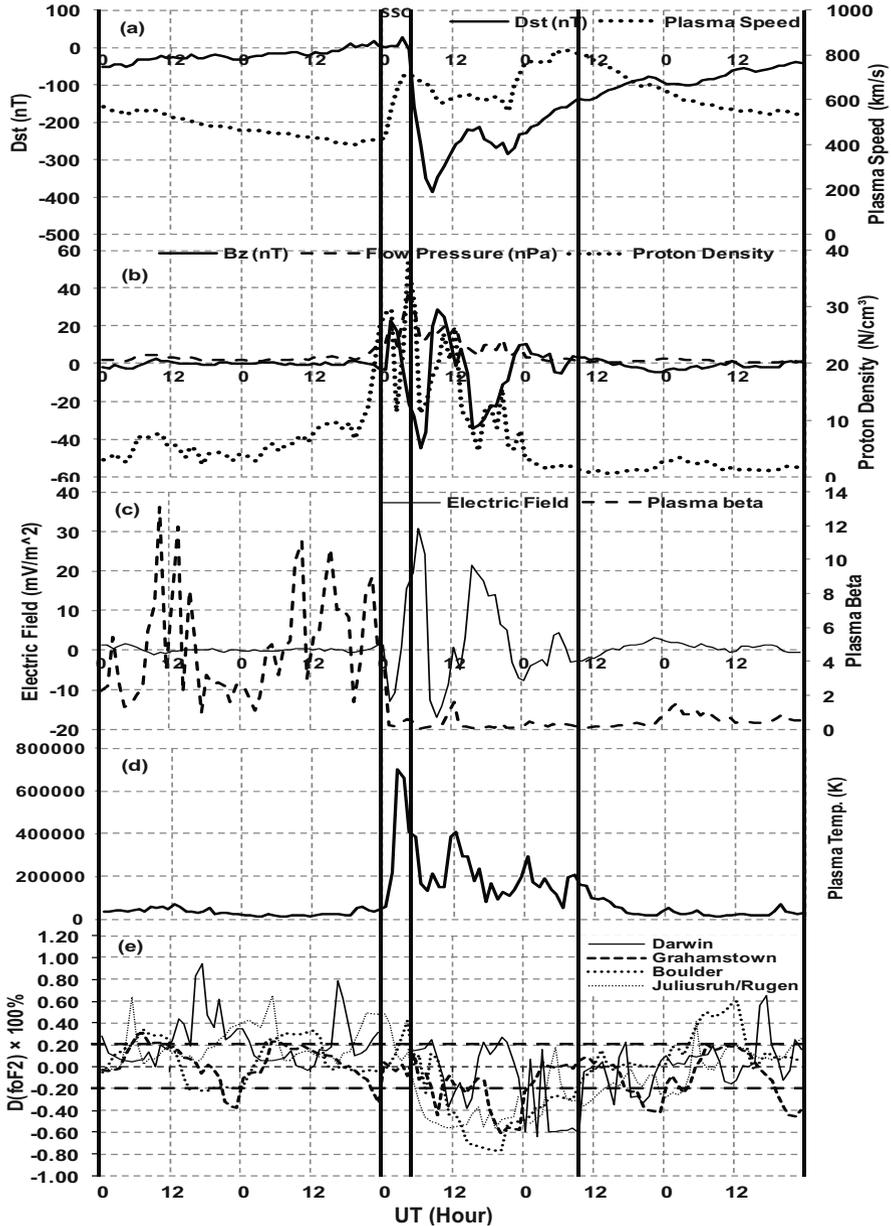


Fig. 4. Same as in Fig. 1 but for the storm of 29 March – 2 April 2001.

observations for the storm that occurred within 00:00 UT 29 March – 23:00 UT 2 April 2001. From Fig. 1a, the Dst plot shows a reduction in solar activity for a couple of hours at the initial phase. This progressive low storm signature continued till around 00:00 UT 31 March when a sudden positive storm enhancement was followed by a large depression in the Dst index to a value of -387 nT at 08:00 UT. On the contrary, the flow speed plot shows a high speed stream during the main phase of the storm with a peak value of 821 km/s. The maximum Bz southward orientation is observed with a peak value of -44.7 nT (Fig. 1b); this coincides with the Dst minimum peak value at 08:00 UT. The intense IMF is believed to be associated with essentially two activities: firstly the high-speed stream, intrinsic fields, and plasma associated with the coronal ejecta; and secondly, the shocked and compressed fields and plasma due to the collision of the high-speed stream with slower solar wind preceding it (Gonzalez *et al.* 1994). Observing Fig. 1a, b, and d, one can see that the simultaneous increase in the flow speed stream and plasma pressure may be responsible for an increase in plasma temperature during the onset period. However, the electric field was marked with an increase in northward field flow of 30.62 mV/m² during the main phase period. Moreover, the increase in plasma beta and high electric field orientation during the main phase may be an indicative that the storm driver gas was the ejecta associated with a southward IMF Bz sheath (*e.g.*, Gonzalez *et al.* 2002, Dal-Lago *et al.* 2004).

Figure 4e depicts the $D(foF2)$ variation for this storm. Starting from 30 March, the plot shows a concurrent increase in $D(foF2)$ for all the stations. This positive storm excursion was extended to the SSC period. During the SSC, the ionospheric F2 effect over Darwin and Grahamstown (in the Southern Hemisphere) was insignificant, while the effect is noticeable over Boulder and Juliusruh with a positive ionospheric storm effect. Subsequently, all the stations simultaneously depleted which resulted in an intense negative ionospheric storm at the main phase.

3.1.5 Storm observation during 11 April 2001

The plot of the response of 11 April 2001, was shown in Fig. 5. The plot spans 9–13 April 2001. Figure 5a presents the Dst and plasma flow speed observations. The Dst observation revealed moderate storm appearance at the beginning of 9 April. This dominates the first observational day. Thereafter, a progressive increase in Dst value > -25 nT was observed throughout the second day. Thereafter, the Dst decreased to -105 nT at 18:00 UT, reaching the minimum peak value of -271 nT at 23:00 UT, before a sharp recovery. The increasing plasma flow speed observed during the SSC was preceded by a low solar wind flow stream. During the period mark SSC the Dst was observed to be -2 nT, which is an indicative that the storm **does not precede**

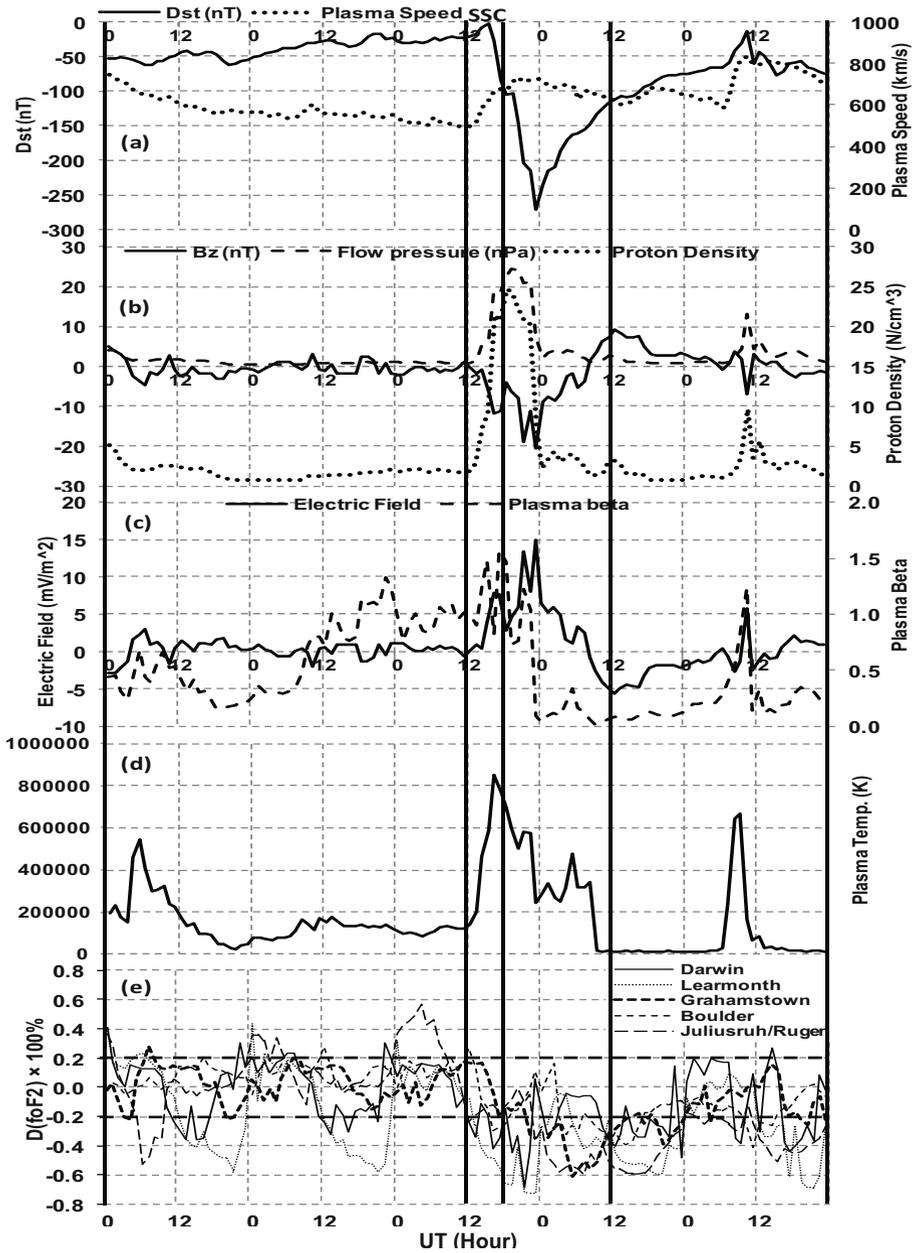


Fig. 5. Same as in Fig. 1, but for the storm of 9-13 April 2001.

by SSC. The B_z southward rotation emerges from the SSC period, reaching a minimum value of -20.5 nT during the main phase (Fig. 5b). As observed

in Fig. 5d, the sudden compression of the magnetosphere by solar wind increased the plasma temperature, which thereafter decreased after the impact. This as well increases the proton density and flow pressure to peak values of 24.7 N/cm^3 and 24.47 nPa at the main phase. The electric field orientation during the SSC period was northward and records a maximum field stream during the main phase. At the right hand side of Fig. 5c is the plasma beta plot. It has a relatively high value (*i.e.*, > 1) at the main phase.

Figure 5e depicts the corresponding ionospheric response. The electron density concurrently decreased during the SSC and the main phase of the storm. The ionospheric effect is more pronounced over the low-latitudes than in the mid-latitudes at both storm-onset and main phase periods. This shows an evident of latitudinal dependence. Also, the critical observation of the hemispheric dependence shows that the ionospheric storm effect is more pronounced in the Southern Hemisphere than in the Northern Hemisphere. This intense negative storm record extends to the recovery phase and decreases with time. This negative ionospheric storm effect during the main phase may be connected with a prompt eastward penetration of electric field (PPEF) and equatorward neutral winds (Balan *et al.* 2010) and change in thermospheric composition generated during geomagnetic storms at auroral latitude which are then transported to lower latitudes by the disturbed thermospheric wind circulation produced by joule heating and particle precipitation in the auroral region (Prölss 1995).

3.1.6 Storm of 22-26 August 2005

The plot of the response of this storm is as shown in Fig. 6. The plot spans 22-26 August 2005. Throughout the initial phase the storm was quiet with the Dst peak value not exceeding $\pm 16 \text{ nT}$. During the onset period the Dst increases as a result of sudden positive enhancement in the H component to about 30 nT before it later decreases to the minimum peak value of 216 nT around 11:00 UT during the main phase. The decrease is a result of depression in the ring current encircling the Earth in the westward direction. The higher plasma density and velocity combine to form a much larger solar wind ram pressure. This pressure compresses the Earth's magnetosphere and increases the field magnitude near the equator. This Dst decrease is coincident with a high speed stream of solar wind with a peak of 721 km/s at 13:00 UT, high proton temperature, increased proton density, and flow pressure at 13:00 UT, as well as high southward turning of B_z of magnitude -38.3 nT was evident. The B_z later rotates northward, reaching a peak value of 19.6 nT at 13:00 UT (in Fig. 6b). However, it transpired from the concurrent enhancement in plasma beta and proton temperature that the shock produced was followed by ejecta which are not of magnetic cloud type.

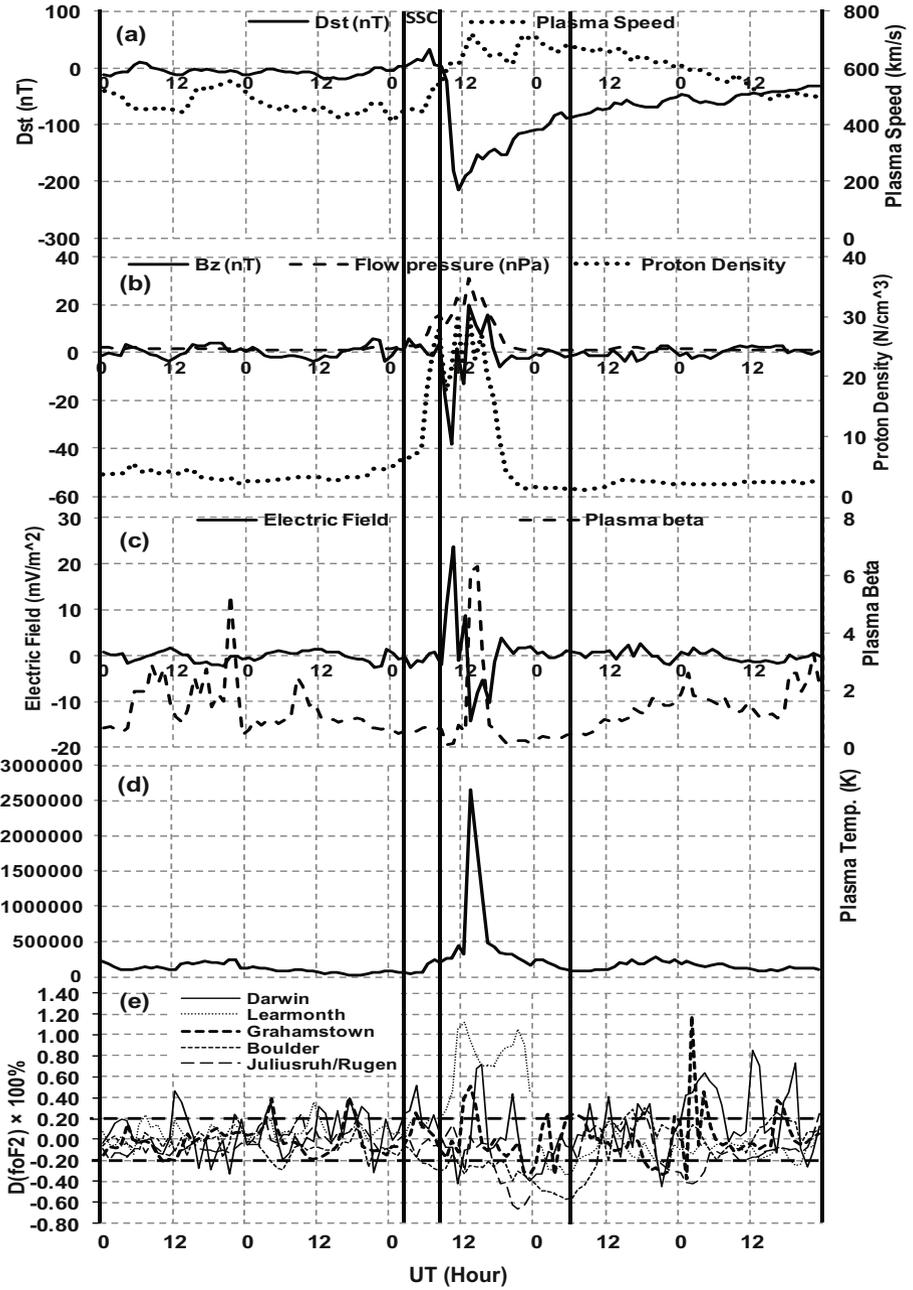


Fig. 6. Same as in Fig. 1, but for the storm of 22-26 August 2005.

Consequently, the ionospheric effect of this storm is highlighted in Fig. 6e. The plot spans 22-26 August 2005 indicating the initial, main, and recovery phases, respectively. As discussed in Section 4, the storm is an intense storm driven by complex ejecta gas. The ionospheric effect is latitudinal dependence. Severe long lasting increase/decrease of ionization at low-/mid-latitude constituted the typical ionospheric response to the intense geomagnetic storm. The $D(f_oF2)$ wavelike variation over Darwin elongated to the SSC period with electron density increase (positive storm effect). Similarly, the ionosphere over Grahamstown had initial records of positive storm phase with a magnitude of 39% and 40% at 04:00 and 17:00 UT before the SSC period. Consequent to the storm onset, the ionosphere responded with a weak positive storm of magnitude 25%. The effect over Learmonth ionosphere is more significant during the main phase, and insignificant during both initial and recovery phases. The ionosphere over Boulder and Juliusruh/Rugen shows similar $D(f_oF2)$ pattern during the main phase. Their respective SSC period response appears to be negative with weak amplitude at Boulder, and Juliusruh experiencing quiet condition. The main phase over both stations was abruptly depleted with a significant negative storm impact.

4. DISCUSSION AND COMPARISON WITH PREVIOUS RESULTS

According to Gonzalez *et al.* (2002), the dominant interplanetary phenomena that are frequently associated with intense magnetic storms are the interplanetary manifestations of fast coronal mass ejections (CMEs). Two such interplanetary structures, involving an intense and long duration B_s component of the IMF are: the sheath region behind a fast forward interplanetary shock, and the CME ejecta itself. These structures, when combined, lead sometimes to the development of very intense storms, especially when an additional interplanetary shock is found in the sheath plasma of the primary structure accompanying another stream (Gonzalez *et al.* 2002). For the intense magnetic storms (IS) ($-250 \text{ nT} \leq Dst < -100 \text{ nT}$) and very intense magnetic storm (VIS) ($Dst \leq -250 \text{ nT}$), the solar wind speed and the IMF intensity must be substantially higher than their "average" values of $V \approx 400 \text{ km/s}$, $B_z = -10 \text{ nT}$ and over a period exceeding 3 h (Gonzalez and Tsurutani 1987, Gonzalez *et al.* 2002, Vieira *et al.* 2001, Adebessin 2008, Adebessin and Chukwuma 2008). With the aforementioned geomagnetic storm classification and characteristics, it is obvious that the storms of 7-9 April 2000, 16-18 July 2000, 29 March – 2 April 2001, and 9-13 April 2001 are very intense storms (VIS), while the storms of 17-19 September 2000 and 22-26 August 2005 are intense geomagnetic storm (IS). Moreover, some geomagnetic storms, especially the largest one, begin with a sudden impulse which signals the arrival of an interplanetary shock structure (Gonzalez *et al.*

2002). This generally coincides with the onset of a period of increased ram pressure (initial phase).

We highlighted in Table 2 the distribution of storms according to their IMF B_z sources. According to Gonzalez *et al.* (2007), there are several drivers that are responsible for the excitation of disturbed conditions. These are: (i) the corotating interactive region (CIR), which in many cases may be associated with high speed stream; (ii) the interplanetary coronal mass ejections (ICME) which are the main causes of magnetic cloud (MC) driver; (iii) the “Sh + MC” for a sheath B_s (B_z southward) field which is followed by a magnetic cloud; (iv) the “S compressed MC” for a magnetic cloud compressed by a shock; and (v) “Complex” for a case in which none of the other cases were identified. However, Echer *et al.* (2005) classified the magnetic cloud as having high magnetic field strength, smooth rotation in the B_z or B_y component, low proton temperature, and plasma beta. The sources of the storms were related to the intensified shock/sheath field, interplanetary magnetic cloud’s field, or the combination of sheath-cloud or sheath-ejecta field (Dal Lago *et al.* 2004). Only the storm of 16 July 2000 is observed to be driven by MC, while all others are either driven by sheath and cloud, sheath followed by ejecta, or ordinary sheath (see Table 2). Echer *et al.* (2005 and reference therein) have reported the occurrence of low proton temperature in solar wind at main phase and postulated the existence of magnetic cloud. Furthermore, they characterized the driver gas region by low solar wind proton density and temperature associated with intense and smooth magnetic fields during a visualization of the evolution of the ICMEs in interplanetary space. Sometime within the gas, strong north-south IMF orientations occur. This occurs mainly in a low plasma beta within an interval of 0.03-0.8, with 0.1 typical (*e.g.*, Choe *et al.* 1992). We found here for the MC driver gas, a low plasma beta of 0.03, low proton density of 9.7, high north-south IMF component, and low plasma temperature, which are common ICME driver characteristics.

During a geomagnetic storm, the disturbed solar wind-magnetosphere interactions could affect the low- and mid-latitude ionospheric F region due to intense transient magnetospheric (prompt or direct penetration) convective electric fields and neutral air wind (ionospheric disturbance dynamo) (Adekoya *et al.* 2013 and references therein), which result in changes to rates of production and loss of ionization. These electric fields redistribute the plasma, affecting production and loss rates (Buosanto 1999). Storm-related electric fields may also destabilize the plasma, producing irregularities. It is to be noted that ionization density may either increase or decrease during disturbed conditions, and these changes are traditionally designated as positive and negative ionospheric storms, respectively. A decrease in the

mean molecular mass (increasing the O density relative to N₂ and O₂) from downwelling through constant pressure surfaces would lead to increases in electron density (positive storm) while an increase in the mean molecular mass (*i.e.*, decreases in the O/N₂ and O/O₂ neutral density ratios) due to upwelling leads to decreases in electron density (negative storm) (Buosanto 1999). Mikhailov *et al.* (1995) reported that an increase in the O density is more important than an increase in the O/N₂ ratio in causing positive storm effects. Parameters such as latitude, hemispheric, and phase of the storm are crucial for determining the occurrence and magnitude of the positive and negative storm effects. Many observations and modeling studies have shown the mechanisms that are responsible for these storm time effects.

Though many attempts are going on, there remain gaps in the present understanding of how the ionosphere responds to geomagnetic storms. Danilov (2001) had explained the principal features of the positive and negative phase distribution and variations on the basis of the principal concept: during a geomagnetic disturbance there is an input of energy into the polar ionosphere, which changes thermospheric parameters, such as composition, temperature, and circulation. Composition changes directly influence the electron concentration in the ionospheric F2 region. The circulation spreads the heated gas to lower latitudes. The conflict between the storm-induced circulation and the regular one determines the spatial distribution of the negative and positive phases in various seasons. According to Davis *et al.* (1997), large ionospheric currents cause joule heating of the atmosphere and if these currents are sufficiently long lived, they can cause an upwelling of the neutral thermosphere, bringing more molecular species such as N₂ and O₂ to the height of f_oF2 . At F region altitudes, these molecules are converted into molecular ions by charge exchange or ion-atom interchange reactions with oxygen ions and then relatively rapidly undergo dissociative recombination to cause depletion in the F region plasma density. The origin of negative ionospheric storms is well understood. It is attributed to changes in the composition of neutral density in response to a storm, which will take place only after sufficient particle precipitation from the auroral oval.

Danilov (2013) has recently proposed in his comprehensive review paper that SSC enhancements could occur in the absolutely geomagnetically quiet background, so it could not be related directly to geomagnetic activity but manifests some sort of interaction between the ionosphere and neutral atmosphere (the coupling from below) to geomagnetic activity. However, around this period, little energy is entering into the earth magnetosphere, which significantly decreases the ring current encircling the earth magnetosphere in westward direction (*e.g.*, Gonzalez *et al.* 1994, Kamide *et al.* 1998). The storms of 6 April 2000 and 11 April 2001, are not preceded by SSC, so their storm onset may not be associated with SSC. The observed

$D(f_oF_2)$ variation for the five ionospheric stations in the low and low-mid latitude did not show a concomitant variation during sudden storm commencement, main phase and the recovery phase, respectively. At low latitude the SSC period is recorded with a weak storm appearance and mid latitude stations remains in quiet mode. This may result from the fact Adekoya *et al.* (2012a) raised. They reported that ionospheric F2 response during the SSC period is thought puzzling; they may either increase or decrease or remain in quiet mode. Also, they reported that low to moderate variation in the ionospheric F2 during SSC may signal the upcoming of large ionospheric disturbances at the main phase.

Liu *et al.* (2008) studied the enhancement of the electron concentration in the ionosphere during three geomagnetic storms (21 April 2001, 29 May 2003, and 22 September 2001) using ionosonde observations and total electron content measurement along the 120°E meridian in the Asia/Australia sector. All three events show quite similar features. The strong SSC enhancements during these events are simultaneously presented in f_oF_2 and TEC and enhancements have latitudinal dependence, tending to occur at low latitudes with maxima near the northern and southern equatorial ionization anomaly (EIA) crests and depletions in the equatorial region. This is quite different from what was reported by Burešová and Laštovička (2007) for middle latitudes. They found no systemic latitudinal dependence in SSC enhancements over European region. However, the difference in the latitudinal and longitudinal intervals considered in both papers could be a natural explanation of that difference (Danilov 2013). An SSC is a period during which little energy is entering into the earth magnetosphere regardless of the speed and number of density of particle in the solar wind. This is referred to as the “pre-storm” period in other related literatures (*e.g.*, Kane 1975, Burešová and Laštovička 2007, 2008, Mikhailov and Perrone 2009, Adekoya *et al.* 2012a, Danilov 2013).

Prior to the storms, some of the stations recorded some degree of positive/negative ionospheric effect. This was consistent with earlier studies of Kane (1975). He pointed out that sometimes there appears positive phase of an ionospheric storm before the SSC of a magnetic storm. Blagoveshchensky and Kalishin (2009) carried out a detailed study of positive phases of ionospheric storms preceding SSC. They came to a conclusion that ordinary mechanisms operating during the main phases of the magnetic storm cannot provide the necessary enhancement in f_oF_2 prior to the storm. The sudden decrease in the intensity of electron density during the SSC period was immediately followed by a severe response during the main phase.

Furthermore, it is very important to note the discernible changes between the latitudinal ionospheric effect during IS and VIS events. During the IS, the atmosphere at the low-latitude region responded with an intense positive

ionospheric storm effect and the effect was *vice versa* over the mid-latitude station. This observation was consistent with the recent work of Adekoya and Adebessin (2014) and Adebiyi *et al.* (2014). However, the ionospheric effect is higher in the low latitude than it appears in the mid-latitude. The $D(f_oF2)$ variation is concurrent during VIS, the ionosphere is dominated by intense negative storm. This simultaneous depletion in the $D(f_oF2)$ at the main phase during the VIS event may be connected with a substantial increase in the average values of the solar wind parameters and the interplanetary magnetic field. Observing the latitudinal dependence at the low and mid-latitudes during the main and the recovery phases of the storm, the negative ionospheric storm effect at the main phase was followed by a positive storm during the recovery phase at low-latitudes and overturned at the mid-latitudes. The mid-latitude was generally depleted with a negative storm effect prior to variation during the SSC period. This is consistent with the results obtained by Adebiyi *et al.* (2014) and Adekoya *et al.* (2012b). Adekoya *et al.* (2012b) had investigated the effect of geomagnetic storm on middle latitude ionospheric F2 during the storm of 2-6 April 2004, and reported that positive-negative (PN) storm phase in $D(f_oF2)$ variation during SSC will signal the upcoming of an intense negative storm during the main phase. For the observed low latitudes, the low, negative storm (N) variation during the SSC is trailed by a severe negative storm effect of the main phase. If ionospheric storms show an initial positive excursion followed by negative, it is classified as PN-storms (*e.g.*, Vijaya *et al.* 2011). Furthermore, regardless of the peak variation of f_oF2 at the initial phase, the pre-storm period is always characterized with low ionospheric storm which is immediately followed by an intense ionospheric variation at the main phase.

Tables 3 and 4 presents the peak deviation of the ionospheric critical frequency $D(f_oF2)$ during the SSC and main phase period for each of the geomagnetic storm events at the low and mid-latitude stations with their corresponding peak values of Dst index, solar wind parameters (*i.e.*, flow speed (V), southward interplanetary magnetic field (IMF B_z), and plasma pressure (P)) as well as the combine solar wind speed, and the B_z (VB_z field) at the main phase; they all follow into the sun-Earth's magnetosphere-ionosphere coupling. Figures 7a and b show the regression plots for the ionospheric F2 variation with the aforementioned geomagnetic storm parameters for each of the five ionospheric stations during the main phase period; this is clarified in Table 5. Both tables, however, present the correlation coefficient between $D(f_oF2)$ and the corresponding peak values of Dst , solar wind parameters and VB_z field, respectively. However, it is imperative to note that not all the observed storms are preceded by SSC. However, those that are preceded by SSC are associated with high solar wind speed (*i.e.*, $V \geq 700$ km/s). SSC or storm onset periods resulted from the sudden decrease in

Table 3
The peak values of geomagnetic, solar wind phenomena, and ionospheric stations $D(\text{foF2})$ peak variation at main phase

Storm date	Peak value of solar wind phenomena at main phase					Ionospheric stations $D(\text{foF2})$ peak variations at main phase				
	Peak D_{st} [nT]	V [km/s]	B_z [nT]	Pressure [nPa]	V/B_z [nT]	Darwin	Learmonth	Grahamstown	Boulder	Juliusruh/Rugen
6 Apr 2000	-288	625	-27.3	20.34	17062.5	-0.63	-0.43	-0.59	-0.67	-0.67
16 Jul 2000	-301	1107	-49.4	41.20	57798.0	*	-0.23	0.31	-0.59	-0.71
17 Sep 2000	-201	839	-23.9	25.54	20052.1	1.68	-0.49	-0.34	-0.63	-0.49
31 Mar 2001	-387	821	-44.7	20.78	36698.7	-0.64	-	-0.61	-0.77	-0.56
11 Apr 2001	-271	687	-20.5	24.47	14083.5	-0.68	-0.72	-0.61	-0.31	-0.60
24 Aug 2005	-216	721	-38.3	30.98	27614.3	0.72	1.12	0.50	-0.57	-0.67

*no data available

Table 4
The peak values of geomagnetic, solar wind phenomena, and ionospheric stations $D(\text{foF2})$ peak variation for storm-onset period

Storm date	Peak value of solar wind phenomena during SSC					Ionospheric stations $D(\text{foF2})$ peak variations during SSC				
	Peak D_{st} [nT]	V [km/s]	B_z [nT]	Pressure [nPa]	V/B_z [nT]	Darwin	Learmonth	Grahamstown	Boulder	Juliusruh/Rugen
6 Apr 2000	-6	564	1.8	8.13	1015.2	-0.52	-0.23	0.06	-0.33	-0.28
16 Jul 2000	28	610	11.6	30.15	7076.0	*	0.32	0.33	-0.06	0.05
17 Sep 2000	7	687	23.7	8.34	16281.9	-0.39	-0.10	-0.07	0.11	-0.41
31 Mar 2001	26	703	22.9	38.7	16098.7	0.16	-	-0.08	0.43	0.49
11 Apr 2001	-2	670	1.31	19.3	877.7	-0.42	-0.55	0.16	0.27	-0.17
24 Aug 2005	32	545	5.9	14.24	3215.5	0.52	0.23	0.25	-0.28	-0.16

*no data available

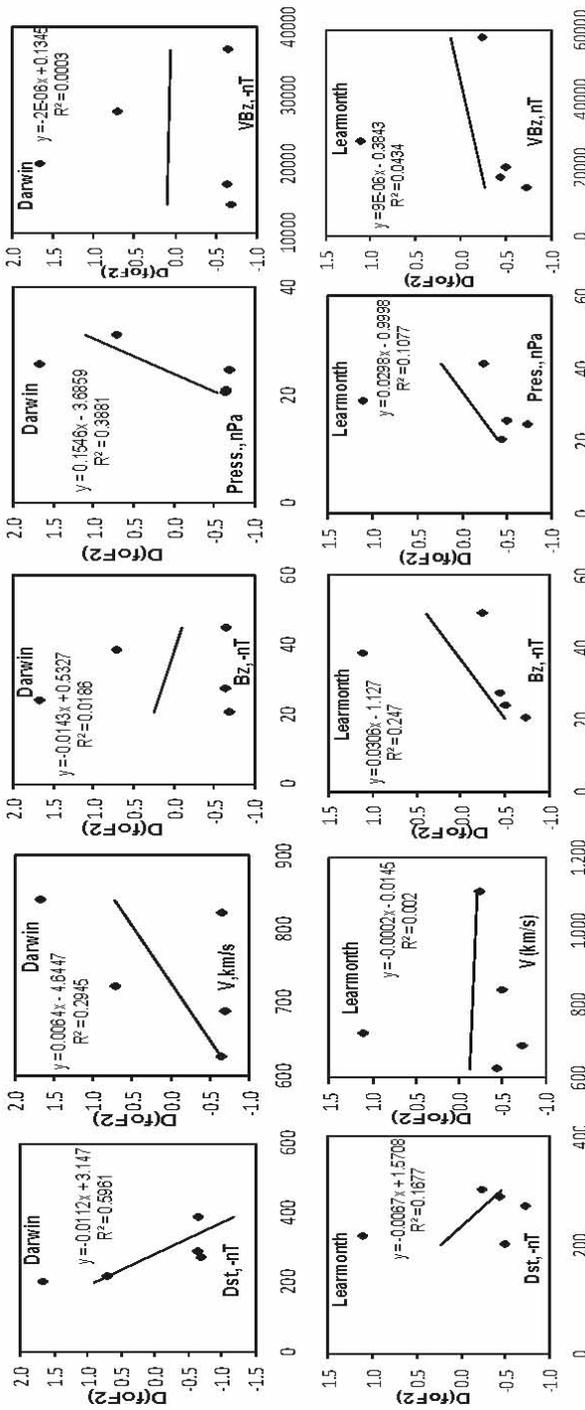


Fig. 7a. Showing the regression plots for ionospheric $D(f_0F_2)$ peak variations with Dst (first column), plasma flow speed V (second column), IMF Bz (third column), plasma pressure (fourth column), and VBz (last column) for low-latitude stations for the years under consideration at the main phase.

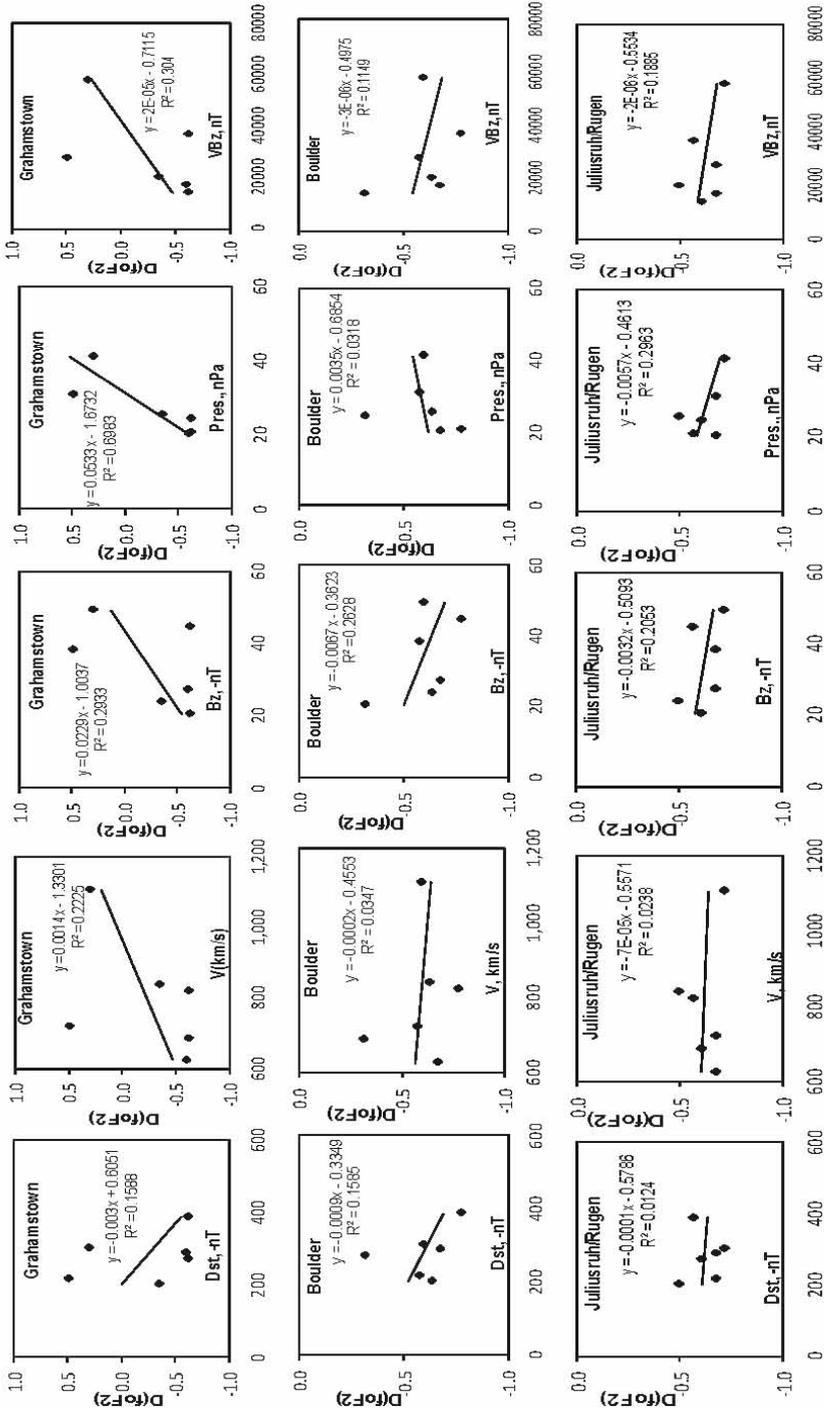


Fig. 7b. Same as in Fig. 7a, but for mid-latitude stations.

Table 5

Correlation coefficient of $D(f\text{o}F2)$ variations during each of the 6 storm events versus Dst , V , Bz , plasma pressure, and VBz at main phase respectively; for each of the 5 stations

Stations	Latitude position	$D(f\text{o}F2)$ vs peak Dst [nT]	$D(f\text{o}F2)$ vs V [km/s]	$D(f\text{o}F2)$ vs Bz [nT]	$D(f\text{o}F2)$ vs P [nPa]	VBz [nT]
Darwin	Low	0.772	0.543	0.136	0.623	0.017
Learmonth	Low	0.410	0.045	0.497	0.328	0.208
Low latitude average		0.591	0.294	0.317	0.476	0.113
Grahamstown	Mid	0.399	0.472	0.542	0.836	0.551
Boulder	Mid	0.398	0.186	0.513	0.178	0.339
Juliuruh/Rugen	Mid	0.111	0.154	0.453	0.544	0.434
Mid latitude average		0.303	0.271	0.502	0.519	0.442
Total averaged		0.418	0.280	0.428	0.502	0.310
Aprox. total averaged [%]		42	28	43	50	31

Note: The total averaged is the mean correlation coefficient for each variable plotted against $D(f\text{o}F2)$.

the ring current encircling the earth's magnetosphere and subsequently suddenly enhanced the H component (*i.e.*, Dst) of the field. This generally coincides with the onset of a period of increased ram pressure (initial phase) that is followed by sustained southward IMFs (main phase) and then by a return-to-normal condition (the recovery phase) (*e.g.*, Gonzalez *et al.* 1994). Consequently, the large class of storms is characterized with flow speed exceeding 700 km/s and begins with SSC.

Figures 7a and b present the regression scatter plots for the ionospheric F2 stations under investigation during the main phase of the aforementioned geomagnetic storms. The first column presents the regression plots of Dst [nT], the second for V [km/s], third for Bz [nT], fourth for P [nPa], and fifth for VBz [nT] in that order against their corresponding ionospheric F2 response for the low- and mid-latitude stations under consideration. The summary of the correlation values are highlighted in Table 5. It was observed that the correlation of $D(f\text{o}F2)$ against Dst is more appreciative for Darwin, with percentage coefficient of 77% when compared to other stations. The corresponding ionospheric F2 variation against flow speed V [km/s] is depicted in the second column in Fig. 7a. It can be seen from the scatter plot and Table 5 that only Darwin showed a better correlation between $D(f\text{o}F2)$

and V . The correlation coefficient is 54% compare to Grahamstown (47%), Boulder (18%), Juliusruh/Rugen (15%), and Learmonth (4%). In relation to B_z , it was observed that Grahamstown (54%) and Boulder (51%) have a good correlation, its relationship with Learmonth was approximately (50%) better compared to Juliusruh (40%) and Darwin (13%). However, only Learmonth and Boulder show an almost insignificant correlation between $D(foF2)$ and plasma pressure. The average sum of the low- and mid-latitude values are use in observing the latitudinal dependence. At low-latitudes, only Dst is the main feature for ionospheric storm, with average percentage correlation of 59%. However, at mid-latitude, B_z and plasma pressure are the deterministic feature for causing intense ionospheric storm. Adebessin and Kayode (2012) have suggested that at high- and mid-latitude flow speed is the most geoeffective parameter with the F2 ionosphere. However, the present studies confirmed that at low-latitude Dst is the most geoeffective, while plasma pressure (52%) and B_z (54%) might be the important parameters causes ionospheric F2 variation at mid-latitude.

Moreover, it should be noted that the ionospheric F2 effect over the low-latitude region is more significant than it appears over the mid-latitude stations during the VIS events. The effect overturned during the IS event. This may be associated with the internal ionospheric effect and not necessary solar wind criteria feature alone. This is because the correlation between $D(foF2)$ and solar wind parameters did not show much significance in their coefficients except for the $D(foF2)$ and Dst . Although, the persistency of solar wind critical feature in the magnetosphere is as a result of combinational solar wind and internal magnetospheric activity rather than solar wind variation alone (Balasis *et al.* 2006). As a consequence, we combined the solar wind plasma speed (V) and the southward interplanetary field (B_z) to look at their relationship with the $D(foF2)$, since the VB_z is the most geoeffective solar wind variables with Dst index (Balasis *et al.* 2006, Bakare and Chukwuma 2010). Their relationship shows on the average a poor correlation (see Table 5), the low-latitude average is 11%, mid-latitude average is 44%, and the overall average is 31%. The consequent of this is that the increase in the ionospheric effect during geomagnetic storm may be related to the combination of interplanetary and geomagnetic parameters, and internal ionospheric effect rather than solar wind variation alone.

5. SUMMARY AND CONCLUSION

We have presented the relationship between ionospheric F2 and solar wind phenomena during the SSC and main phase of geomagnetic storms for six (6) geomagnetic storms. Two are intense ($-100 \text{ nT} \leq Dst < -250 \text{ nT}$) and four are very intense ($Dst \leq -250 \text{ nT}$) in the low and low-mid latitudes. The SSC phenomena that lead to an intense ionospheric positive and negative

storm phase during the *Dst* minimum were investigated. These are *Dst* positive increase, southward turning of B_z (≤ -10 nT), increase in plasma speed, and high plasma pressure for both low- and low-mid-latitudes. It is pertinent to note that the largest geomagnetic storm with flow speed exceeding 700 km/s begins with SSC. Saranya *et al.* (2011) found that *Dst* index has no influence on the pre-storm enhancement phenomenon. Blagoveshchensky and Kalishin (2009) assumed that the mechanisms of the pre-storm enhancements are related to the impact of rapid particles in the foreshock region of the solar wind on the Earth's magnetosphere. Regardless of the peak variation of $D(f_oF2)$ the SSC period is always characterized with low ionospheric storm, which is immediately followed by an intense ionospheric variation during the main phase. The implication is that ionospheric storm during SSC period is insignificant compared to the main phase but can trigger the main phase storm phenomenon. Only one out of the entire storms is driven by magnetic cloud gas, which is the common ICME driver.

We observed that ionospheric storm effect during SSC is latitudinal symmetric. It is more pronounced at the low-latitude with negative storm effect than at mid-latitude with positive storm effect. However, the main phase is characterized with severe negative storm effect at both latitudes during VIS event periods. This is attributed to the large energetic particle and solar activity input into the earth magnetosphere. The negative ionospheric storm effect at the main phase was followed by a positive storm during the recovery phase at low-latitudes and overturned at mid-latitudes. In contrary during the IS, the atmosphere at the low-latitude region responded with an intense positive ionospheric storm effect and the effect was *vice versa* over the mid-latitude station. In a nut shell, the ionospheric effect is more significant in the low-latitude than at mid-latitude.

Additionally, SSC period is the quality of upcoming geomagnetic storm and ionospheric disturbance. Adekoya *et al.* (2012a, b) had suggested that SSC period might be related to some kind of pathway for penetration of energy into the terrestrial ionosphere and magnetosphere at the main phase. The recovery phase is simultaneously depleted with a significant ionospheric storm effect throughout the studied geomagnetic storm events across the latitude except for Darwin whose ionosphere responded with a severe positive storm during 6-8 April 2000 event.

The poor correlation between the solar wind parameters and $D(f_oF2)$ indicates that increase in ionospheric storm effect during geomagnetic storms may be related to the combination effect of interplanetary and geomagnetic parameters and internal ionospheric effect. At low-latitude *Dst* is the most geoeffective, while plasma pressure (52%) and B_z (54%) might be the important parameters causing ionospheric F2 effect at mid-latitudes. The ionospheric F2 effect over the low-latitude region is more significant than it

appears over the mid-latitude stations during the VIS while the effect is overturned during the IS period.

Acknowledgement. The authors are grateful to working team of National Space Science data Centres (NSSDC's) OMNI database (<http://nssdc.gsfc.nasa.gov/omniweb>), and also to the National Geophysical Data Center's SPIDR (Space Physics Interactive Data Resource) network's (<http://spidr.ngdc.noaa.gov>). The authors also appreciate the reviewers for the constructive comments on the structure of the paper and for their useful suggestion. This has tremendously improved the quality of the paper.

References

- Adebesin, B.O. (2008), Roles of interplanetary and geomagnetic parameters in “intense” and “very intense” magnetic storms generation and their geoeffectiveness, *Acta Geod. Geoph. Hung.* **43**, 4, 383-408, DOI: 10.1556/AGeod.43.2008.4.2.
- Adebesin, B.O., and V.U. Chukwuma (2008), On the variation between D_{st} and IMF B_z during “intense” and “very intense” geomagnetic storms, *Acta Geod. Geoph. Hung.* **43**, 1, 1-15, DOI: 10.1556/AGeod.43.2008.1.1.
- Adebesin, B.O., and J.S. Kayode (2012), On the coexistence of ionospheric positive and negative storm phases during January–December, 2000 geomagnetic activities at East Asian Sector, *Adv. Appl. Sci. Res.* **3**, 1, 349-370.
- Adebesin, B.O., S.O. Ikubanni, S.J. Adebisi, and B.W. Joshua (2013), Multi-station observation of ionospheric disturbance of March 9, 2012 and comparison with IRI-model, *Adv. Space Res.* **52**, 4, 604-613, DOI: 10.1016/j.asr.2013.05.002.
- Adebisi, S.J., I.A. Adimula, O.A. Oladipo, B.W. Joshua, B.O. Adebesin, and S.O. Ikubanni (2014), Ionospheric response to magnetic activity at low and mid-latitude stations, *Acta Geophys.* **62**, 4, 973-989, DOI: 10.2478/s11600-014-0205-x.
- Adekoya, B.J., and B.O. Adebesin (2014), Hemispheric, seasonal and latitudinal dependence of storm-time ionosphere during low solar activity period, *Adv. Space Res.* **54**, 11, 2184-2193, DOI: 10.1016/j.asr.2014.08.013.
- Adekoya, B.J., V.U. Chukwuma, N.O. Bakare, and T.W. David (2012a), On the effects of geomagnetic storms and pre storm phenomena on low and middle latitude ionospheric F2, *Astrophys. Space Sci.* **340**, 2, 217-235, DOI: 10.1007/s10509-012-1082-x.
- Adekoya, B.J., V.U. Chukwuma, N.O. Bakare, and T.W. David (2012b), Effects of geomagnetic storm on middle latitude ionospheric F2 during storm of 2-6 April 2004, *Indian J. Rad. Space Phys.* **41**, 6, 606-616.

- Adekoya, B.J., V.U. Chukwuma, and S.A. Salako (2013), On the coexistence of positive and negative ionospheric storm during geomagnetic storms and pre storm phenomena on low and low-mid latitude ionospheric F2. **In: Proc. 5th Nat. Ann. Conf. Nigerian Union of Radio Science (NURS)**, 15-28.
- Adeniyi, J.O. (1986), Magnetic storm effects on the morphology of the equatorial F2-layer, *J. Atmos. Sol.-Terr. Phys.* **48**, 8, 695-702, DOI: 10.1016/0021-9169(86)90019-X.
- Akala, A.O., E.O. Oyeyemi, E.O. Somoye, A.B. Adeloje, and A.O. Adewale (2010), Variability of foF2 in the African equatorial ionosphere, *Adv. Space Res.* **45**, 11, 1311-1314, DOI: 10.1016/j.asr.2010.01.003.
- Bakare, N.O., and V.U. Chukwuma (2010), Relationship between Dst and solar wind conditions during intense geomagnetic storms, *Indian J. Rad. Space Phys.* **39**, 3, 150-155.
- Balan, N., K. Shiokawa, Y. Otsuka, T. Kikuchi, D. Vijaya Lekshmi, S. Kawamura, M. Yamamoto, and G.J. Bailey (2010), A physical mechanism of positive ionospheric storms at low latitudes and mid latitudes, *J. Geophys. Res.* **115**, A2, A02304, DOI: 10.1029/2009JA014515.
- Balasis, G., I.A. Daglis, P. Kaperis, M. Manda, D. Vassiliadis, and K. Eftaxias (2006), From pre-storm activity to magnetic storms: a transition described in terms of fractal dynamics, *Ann. Geophys.* **24**, 3557-3567, DOI: 10.5194/angeo-24-3557-2006.
- Blagoveshchensky, D.V., and A.S. Kalishin (2009), Increase in the critical frequency of the ionospheric F region prior to the substorm expansion phase, *Geomagn. Aeron.* **49**, 2, 200-209, DOI: 10.1134/S0016793209020091.
- Buonsanto, M.J. (1999), Ionospheric storms – A review, *Space Sci. Rev.* **88**, 3-4, 563-601, DOI: 10.1023/A:1005107532631.
- Burešová, D., and J. Laštovička (2007), Pre-storm enhancements of foF2 above Europe, *Adv. Space Res.* **39**, 8, 1298-1303, DOI: 10.1016/j.asr.2007.03.003.
- Burešová, D., and J. Laštovička (2008), Pre-storm electron density enhancements at middle latitudes, *J. Atmos. Sol.-Terr. Phys.* **70**, 15, 1848-1855, DOI: 10.1016/j.jastp.2008.01.014.
- Choe, G.S., N. LaBelle-Hamer, B.T. Tsurutani, and L.C. Lee (1992), Identification of a driver gas boundary layer, *EOS Trans. AGU* **73**, 485.
- Chukwuma, V.U. (2007), On positive and negative ionospheric storms, *Acta Geod. Geoph. Hung.* **42**, 1, 1-21, DOI: 10.1556/AGeod.42.2007.1.1.
- Dal-Lago, A., L.E.A. Vieira, E. Echer, W.D. Gonzalez, A.L. Clúa de Gonzalez, F.L. Guarneri, L. Balmaceda, J. Santos, M.R. da Silva, A. de Lucas, and N.J. Schuch (2004), Great geomagnetic storms in the rise and maximum of solar cycle 23, *Braz. J. Phys.* **34**, 4B, 1542-1546, DOI: 10.1590/S0103-97332004000800008.
- Danilov, A.D. (2001), F2-region response to geomagnetic disturbances, *J. Atmos. Sol.-Terr. Phys.* **63**, 5, 441-449, DOI: 10.1016/s1364-6826(00)00175-9.

- Danilov, A.D. (2013), Ionospheric F-region response to geomagnetic disturbances, *Adv. Space Res.* **52**, 3, 343-366, DOI: 10.1016/j.asr.2013.04.019.
- Davis, C.J., M.N. Wild, M. Lockwood, and Y.K. Tulunay (1997), Ionospheric and geomagnetic responses to changes in IMF Bz: a superposed epoch study, *Ann. Geophys.* **15**, 2, 217-230, DOI: 10.1007/s00585-997-0217-9.
- Echer, E., M.V. Alves, and W.D. Gonzalez (2005), A statistical study of magnetic cloud parameters and geoeffectiveness, *J. Atmos. Sol.-Terr. Phys.* **67**, 10, 839-852, DOI: 10.1016/j.jastp.2005.02.010.
- Fejer, B.G. (1997), The electrodynamics of the low-latitude ionosphere: Recent results and future challenges, *J. Atmos. Sol.-Terr. Phys.* **59**, 13, 1465-1482, DOI: 10.1016/S1364-6826(96)00149-6.
- Foster, J.C., and F.J. Rich (1998), Prompt midlatitude electric field effects during severe geomagnetic storms, *J. Geophys. Res.* **103**, A11, 26367-26372, DOI: 10.1029/97JA03057.
- Gonzalez, W.D., and B.T. Tsurutani (1987), Criteria of interplanetary parameters causing intense magnetic storms ($D_{st} < -100$ nT), *Planet. Space Sci.* **35**, 9, 1101-1109, DOI: 10.1016/0032-0633(87)90015-8.
- Gonzalez, W.D., J.A. Joselyn, Y. Kamide, H.W. Kroehl, G. Rostoker, B.T. Tsurutani, and V.M. Vasyliunas (1994), What is a geomagnetic storm?, *J. Geophys. Res.* **99**, A4, 5771-5792, DOI: 10.1029/93JA02867.
- Gonzalez, W.D., B.T. Tsurutani, and A.L. Clúa de Gonzalez (1999), Interplanetary origin of geomagnetic storms, *Space Sci. Rev.* **88**, 3-4, 529-562, DOI: 10.1023/A:1005160129098.
- Gonzalez, W.D., B.T. Tsurutani, R.P. Lepping, and R. Schwenn (2002), Interplanetary phenomena associated with very intense geomagnetic storms, *J. Atmos. Sol.-Terr. Phys.* **64**, 2, 173-181, DOI: 10.1016/S1364-6826(01)00082-7.
- Gonzalez, W.D., E. Echer, A.L. Clua-Gonzalez, and B.T. Tsurutani (2007), Interplanetary origin of intense geomagnetic storms ($D_{st} < -100$ nT) during solar cycle 23, *Geophys. Res. Lett.* **34**, 6, L06101, DOI: 10.1029/2006GL028879.
- Kamide, Y., N. Yokoyama, W. Gonzalez, B.T. Tsurutani, I.A. Daglis, A. Brekke, and S. Masuda (1998), Two-step development of geomagnetic storms, *J. Geophys. Res.* **103**, A4, 6917-6921, DOI: 10.1029/97JA03337.
- Kane, R.P. (1975), Global evolution of the ionospheric electron content during some geomagnetic storms, *J. Atmos.-Terr. Phys.* **37**, 4, 601-611, DOI: 10.1016/0021-9169(75)90055-0.
- Kane, R.P. (2005), Ionospheric foF2 anomalies during some intense geomagnetic storms, *Ann. Geophys.* **23**, 2487-2499, DOI: 10.5194/angeo-23-2487-2005.
- Liu, J., B. Zhao, and L. Liu (2010), Time delay and duration of ionospheric total electron content responses to geomagnetic disturbances, *Ann. Geophys.* **28**, 3, 795-805, DOI: 10.5194/angeo-28-795-2010.

- Liu, L., W. Wan, M.-L. Zhang, and B. Zhao (2008), Case study on total electron content enhancements at low latitudes during low geomagnetic activities before the storms, *Ann. Geophys.* **26**, 4, 893-903, DOI: 10.5194/angeo-26-893-2008.
- Mikhailov, A.V., and L. Perrone (2009), Pre-storm NmF2 enhancements at middle latitudes: delusion or reality? *Ann. Geophys.* **27**, 3, 1321-1330, DOI: 10.5194/angeo-27-1321-2009.
- Mikhailov, A.V., M.G. Skoblin, and M. Förster (1995), Daytime F2-layer positive storm effect at middle and lower latitudes, *Ann. Geophys.* **13**, 5, 532-540, DOI: 10.1007/s00585-995-0532-y.
- Pröls, G.W. (1993), On explaining the local time variation of ionospheric storm effects, *Ann. Geophys.* **11**, 1, 1-9.
- Pröls, G.W. (1995), Ionospheric F-region storms. **In:** H. Volland (ed.), *Handbook of Atmospheric Electrodynamics*, Vol. 2, CRC Press, Boca Raton, 195-248.
- Saranya, P.L., K. Venkatesh, D.S.V.V.D. Prasad, P.V.S. Rama Rao, and K. Niranjana (2011), Pre-storm behaviour of NmF₂ and TEC (GPS) over equatorial and low latitude stations in the Indian sector, *Adv. Space Res.* **48**, 2, 207-217, DOI: 10.1016/j.asr.2011.03.028.
- Sutton, E.K., J.M. Forbes, R.S. Nerem, and T.N. Woods (2006), Neutral density response to the solar flares of October and November 2003, *Geophys. Res. Lett.* **33**, 22, L22101, DOI: 10.1029/2006GL027737.
- Vieira, L.E.A., W.D. Gonzalez, A.L. Clua de Gonzalez, and A. Dal Lago (2001), A study of magnetic storms development in two or more steps and its association with the polarity of magnetic clouds, *J. Atmos. Sol.-Terr. Phys.* **63**, 5, 457-461, DOI: 10.1016/S1364-6826(00)00165-6.
- 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.
- Wang, C.B., J. K. Chao, and C.H. Lin (2003), Influence of the solar wind dynamic pressure on the decay and injection of the ring current, *J. Geophys. Res.* **108**, A9, 1341, DOI: 10.1029/2003JA009851.

Received 17 April 2014

Received in revised form 4 September 2014

Accepted 26 November 2014



LAGEOS Sensitivity to Ocean Tides

Krzysztof SOŚNICA

Astronomical Institute, University of Bern, Canton Bern, Switzerland;
e-mail: krzysztof.sosnica@aiub.unibe.ch

Institute of Geodesy and Geoinformatics, Wrocław University of Environmental
and Life Sciences, Wrocław, Poland; e-mail: krzysztof.sosnica@igig.up.wroc.pl

Abstract

Satellite Laser Ranging (SLR) to LAGEOS has a remarkable contribution to high-precise geodesy and geodynamics through deriving and validating various global geophysical models. This paper validates ocean tide models based on the analysis of satellite altimetry data, coastal tide gauges, and hydrodynamic data, *i.e.*, CSR3.0, TOPEX4.0, CSR4.0A, FES2004, GOT00.2, and the CSRC Schwiderski model. LAGEOS orbits and SLR observation residuals from solutions based on different ocean tide models are compared and examined. It is found that LAGEOS orbits are sensitive to tidal waves larger than 5 mm. The analysis of the aliasing periods of LAGEOS orbits and tidal waves reveals that, in particular, the tidal constituent S_2 is not well established in the recent ocean tide models. Some of the models introduce spurious peaks to empirical orbit parameters, which can be associated with S_2 , S_a , and K_2 tidal constituents, and, as a consequence, can be propagated to fundamental parameters derived from LAGEOS observations.

Key words: satellite geodesy, ocean tides, SLR, LAGEOS, orbit determination.

1. INTRODUCTION

Satellite Laser Ranging (SLR) greatly contributes to deriving various geodetic and geophysical parameters (Smith and Turcotte 1993), *e.g.*, to deriving Earth orientation parameters (Schutz *et al.* 1989, Sośnica *et al.* 2014), static and time-variable Earth's gravity field (Cheng *et al.* 1997, Bianco *et al.* 1998, Maier *et al.* 2012), station coordinates and velocities (*e.g.*, Schillak and Wnuk 2003, Lejba and Schillak 2011), crustal deformations (Schillak *et al.* 2006), and elastic Earth parameters (Rutkowska and Jagoda 2010, 2012). The orbits of LAGEOS satellites equipped with laser retro-reflectors can be very well determined through the SLR observations with the accuracy at a level of a few millimeters, as well as through the minimized area-to-mass ratios of LAGEOS, and, as a consequence, a minimized impact of non-gravitational orbit perturbations. This implies that LAGEOS orbits are subject to orbit perturbations of gravitational origin to the greatest extent, and thus, are very well-suited for the validation of geodynamical models, *e.g.*, Earth gravity field and ocean tide models.

The tidal forces are caused almost uniquely by the gravitational attraction of the Moon and the Sun. These forces are responsible for the solid Earth, atmosphere, and ocean tides as a consequence of the mass redistribution and gravity changes in the system Earth. Up to now, ocean tide models (OTM) were typically validated by computing monthly tidal elevation differences between models (Wünsch *et al.* 2005), by comparing with global tide gauge datasets (Ponchaut *et al.* 2001, Zahran *et al.* 2006), by comparing with sea level Topex/Poseidon (T/P) time series analysis (Shum *et al.* 1997) or by comparing of the simulations of tidal elevation differences at footpoints of GRACE (Wünsch *et al.* 2005). In this paper, OTM are evaluated using LAGEOS and the impact of the OTM on the orbit determination of the LAGEOS satellites is investigated. An intrinsic comparison is carried out using seven different OTM based on analysis of satellite altimetry data, *e.g.*, TOPEX/Poseidon: CSR3.0 (Eanes and Bettadpur 1996), TOPEX4.0 (Egbert *et al.* 1994, Egbert and Erofeeva 2002), a model based on analysis of gravity field changes observed by different satellite missions: CSR4.0A (Eanes 2004), and hydrodynamic models with assimilation from observed tidal data, *i.e.*, altimeter: FES2004 (Lyard *et al.* 2006), GOT00.2 (Ray 1999), EOT08A (Savcenko and Bosch 2008), and coastal tide gauges: CSRC Schwiderski (Schwiderski 1980).

2. LAGEOS ORBIT MODELING

The orbits of LAGEOS satellites are estimated using a development version of the Bernese GNSS Software version 5.2 (Dach *et al.* 2007) with the SLR extensions. The definition of the satellite orbits consists of the list of models

applied along with estimated deterministic (Keplerian elements) and empirical orbital parameters. Estimating empirical parameters, which absorb insufficiently known perturbing forces, is necessary, because of modelling deficiencies in some forces acting upon a satellite. These parameters are related to modelling deficiencies of the solar radiation pressure (direct and re-radiated by the Earth’s surface), thermal thrust, mismodellings and variations in gravity field and ocean tide models.

In the 7-day LAGEOS solutions, a multistep collocation method of 2-minute intervals and a polynomial degree of 12 was adopted for the numerical integration. The full unconstrained set of six osculating orbital parameters is estimated together with the empirical force model, which comprises:

In along-track: $S_0 + S_S \sin(u) + S_C \cos(u)$,
 In out-of-plane: $W_S \sin(u) + W_C \cos(u)$,

where S_0 stands for a constant acceleration in the along-track direction, S_S and S_C are once-per-revolution sine and cosine along-track accelerations, respectively, W_S and W_C are once-per-revolution sine and cosine accelerations in the out-of-plane direction, and u is an argument of latitude of the satellite within the orbital plane. Additionally estimated parameters are: stations coordinates (all stations), range biases for selected stations – one set of parameters per every week, and Earth orientations parameters (*i.e.*, pole coordinates and length-of-day) – one set of parameters per day.

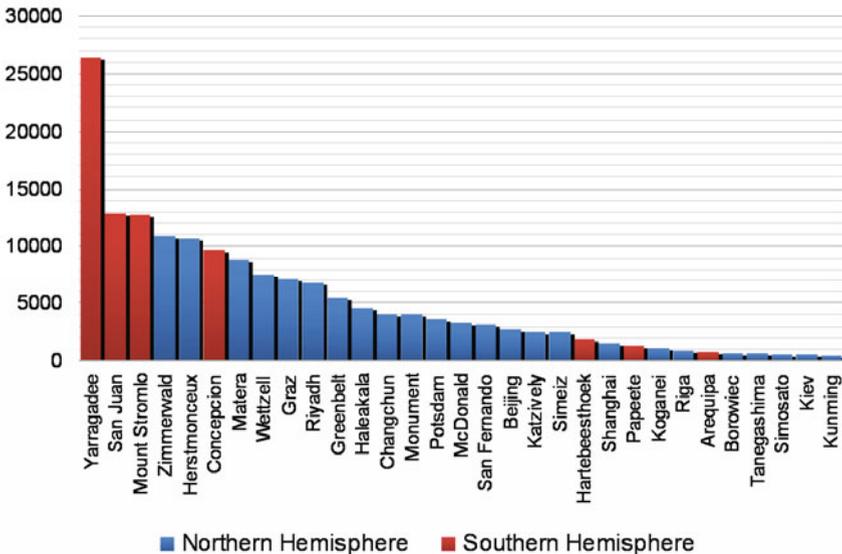


Fig. 1. Number of SLR normal points gathered by ILRS stations to LAGEOS-1 and LAGEOS-2 in 2008.

The reference frame SLRF2008 and Earth orientation IERS-08-C04 series (Bizouard and Gambis 2014) were used as *a priori* information. In general, the 7-day LAGEOS solutions and all models used are in very good agreement with the solutions derived by the International Laser Ranging Service (ILRS; Pearlman *et al.* 2002) Analysis Centers with some modifications, *e.g.*, using of atmospheric tidal loading model for crust displacements

Table 1
Ocean tide models validated using LAGEOS orbits (after Sośnica 2014)

Type of model	Description
Length of arc	7 days
Data editing	2.5 sigma editing, maximum overall sigma: 25 mm
Satellite center of mass	Station- and satellite-specific
Troposphere delay	Mendes–Pavlis delay model + mapping function (Mendes and Pavlis 2004)
Cut-off angle	3 degrees
Gravity field model	EGM2008 up to degree and order 30 (Pavlis <i>et al.</i> 2012)
Relativity	Light time propagation correction and Schwarzschild orbit perturbation according to IERS Conventions 2010 (Petit and Luzum 2010)
Third-body	Earth's Moon, Sun, Venus, Mars, Jupiter, ephemeris: JPL DE405 (Folkner <i>et al.</i> 1994)
Subdaily pole model	IERS2000 (Kolaczek <i>et al.</i> 2000)
Tidal forces	Solid Earth tide model, Earth pole tide model, and ocean pole tide model applied – IERS Conventions 2010, Atmospheric tidal loading (Ray and Ponte 2003)
Nutation model	IAU2000
Solar radiation pressure	Direct radiation: applied with fixed solar radiation coefficient $C_R = 1.13$
Numerical integration	Interval: 2 minutes, polynomial degree: 12, collocation method (Beutler 2005)
Earth orientation parameters (<i>a priori</i>)	<i>A priori</i> C04 series from IERS, consistent with ITRF2008 (Bizouard and Gambis 2014)
Reference frame (<i>a priori</i>)	SLRF2008/ITRF2008 (Altamimi <i>et al.</i> 2011)
Range biases	Estimated for: Simeiz, Zimmerwald (IR)

and the gravity field model EGM2008 (Pavlis *et al.* 2012) up to degree and order 30. *A priori* terrestrial reference frame SLR2008 is the ILRS version of ITRF2008 (Altamimi *et al.* 2011) with some additional stations that were not included in ITRF2008 solution (namely, for some older and some newer SLR sites or SLR sites affected by earthquakes). A list of the most important models used in the LAGEOS solutions can be found in Table 1.

Data of the year 2008 are adopted for the comparison of seven ocean tide models. Figure 1 illustrates the number of SLR normal points to both LAGEOS satellites gathered by the ILRS stations. In total 139 000 SLR normal points are available in this year. The number per week varies between 1932 and 3804. For two selected ocean tide models, namely for FES2004 and CSR4.0A, 10-year LAGEOS solutions (2002-2012) are additionally generated in order to investigate the influence of particular tidal constituents.

3. OCEAN TIDE MODELS (OTM)

Ocean tide models are typically expressed by the coefficients of amplitudes and waves of particular discrete frequencies (Petit and Luzum 2010):

$$\xi(\phi, \lambda, t) = \sum_f Z_f(\phi, \lambda) \cos(\theta_f(t) - \psi_f(\phi, \lambda)), \quad (1)$$

where Z_f is the amplitude of wave f , λ is the longitude, and Φ is the latitude of the point, ψ_f is the phase longitude (referred to the Greenwich meridian) and θ is the Doodson argument. The hydrodynamical effects of ocean tides can be expanded as periodic variations of the normalized Stockes' coefficients of degree n and order m . After the expansion into spherical harmonic functions the equation yields (Petit and Luzum 2010):

$$\xi(\phi, \lambda, t) = \sum_f \sum_{n=1}^N \sum_{m=0}^n P_{nm}(\sin\phi) \sum_{\pm} C_{f,n,m}^{\pm}(\theta_f + \chi_f \pm m\lambda) + S_{f,n,m}^{\pm}(\theta_f + \chi_f \pm m\lambda), \quad (2)$$

where $C_{f,n,m}^{\pm}$, $S_{f,n,m}^{\pm}$ are prograde and retrograde normalized spherical harmonic coefficients of the main wave f of degree n and order m , χ_f is the phase bias according to the Shureman conventions and P_{nm} represents the normalized associated Legendre function. The practical implementation of ocean tide models consists of the list of astronomical amplitudes for semi-diurnal waves M_2 , S_2 , N_2 , K_2 , the diurnal waves K_1 , O_1 , P_1 , Q_1 , the long-period waves M_f , M_m , S_a , S_{sa} , and also for other waves, *e.g.*, quarter-diurnal waves. For an extensive study concerning the impact of the particular tidal waves on the LAGEOS-1 and LAGEOS-2 orbits please refer to Iorio (2001), whereas the analysis of the spectrum of tidal perturbations on the orbital elements of LAGEOS-1 is described by Dow (1990).

4. SENSITIVITY OF LAGEOS ORBITS TO OCEAN TIDES

The truncation of OTM up to a particular degree/order (d/o) or to a minimum size of a tidal wave is important on one hand to minimize computational time, and on the other hand, to investigate the sensitivity of LAGEOS orbits to small tidal waves. In order to study the impact of the maximum d/o of OTM on LAGEOS orbits, six solutions were generated using the CSR4.0A model up to maximum d/o 2, 4, 8, 12, 20, and 30. The RMS of the observation residuals and the comparison between estimated and predicted orbits are shown in Table 2.

Table 2

The sensitivity of LAGEOS orbits to maximum degree/order of OTM and the size of tidal wave using CSR4.0A (mean values for 2008)

OTM up to d/o	Maximum size of tidal wave [mm]	RMS of residuals [mm]	Comparison of estimated and predicted orbits		
			RMS of radial prediction [mm]	RMS of along-track prediction [mm]	RMS of out-of-plane prediction [mm]
2	∞	568.94	–	–	–
4	∞	10.29	23.5	405.9	205.4
8	∞	7.42	29.9	399.1	200.6
12	∞	7.42	29.8	399.0	200.6
20	∞	7.41	29.8	398.9	200.6
30	∞	7.41	29.8	398.7	200.4
30	2000	9.63	30.3	404.7	203.8
30	500	7.91	29.9	397.0	199.9
30	50	7.45	29.9	398.3	200.4
30	5	7.42	29.8	398.9	200.4
30	0.5	7.41	29.8	398.7	200.4

Using maximum d/o 2 or 4 is definitely insufficient, because it gives a significant loss of accuracy of the solution (RMS: 569 and 10 mm, respectively). LAGEOS satellites are very sensitive up to d/o 8 of OTM. Using OTM up to d/o 30 may slightly improve RMS of observation residuals (difference of 0.01 mm), and the orbit prediction, especially in the along-track direction (by 0.8 mm). Degree 8 is an absolute minimum needed for reasonable LAGEOS solutions, but it is recommended that OTM should be used up to d/o 30 for LAGEOS orbit determination in order to avoid orbit degradation due to model truncation.

The sensitivity of LAGEOS solutions to amplitudes of ocean tides is analyzed by choosing only the tides exceeding the particular threshold. Table 2

shows six different test cases with the maximum considered size of the tides set to: 2000, 500, 50, 5, and 0.5 mm, and the approach, where regardless of the size, all waves are considered (∞). Big values of RMS residuals for maximum wave size of 2000 and 500 mm (9.63 and 7.91 mm, respectively) indicate that taking into consideration smaller waves is obligatory when processing LAGEOS data. Small differences are visible between the solutions based on OTM maximum size of 50 and 5 mm. Waves smaller than 0.5 mm have no impact on LAGEOS solutions. Therefore, taking into account at least all waves larger or equal 5 mm is highly recommended.

The high sensitivity of LAGEOS orbits is striking, in particular because the uncertainties of amplitudes (formal errors) of tidal waves in OTM may exceed 50 mm, *i.e.*, about one order of magnitude more than the sensitivity of LAGEOS orbits. This confirms that using the LAGEOS satellites is suitable for validating the low-degree part of OTM.

5. VALIDATION OF OCEAN TIDAL MODELS

Different OTM are validated by comparing the quality of LAGEOS-1 and LAGEOS-2 orbits (see Table 3). Three hydrodynamic models based on tide gauge observations (CSRC Schwiderski) or tide gauges and satellite altimetry (FES2004, GOT00.2) are compared with the hydrological models based on satellite altimetry data (CSR3.0, CSR4.0A, TOPEX4.0) and with one empirical model based on many satellite missions (EOT08A). Most of the OTM are based on the analysis of satellite altimetry data stemming from TOPEX/Poseidon (T/P) mission. Above and below latitude of 66°N and 66°S (given by the T/P or Jason-1/2 satellite inclination), and for shallow sea areas, the tidal waves are of inferior quality. Moreover, some of the models contain missing water areas in the model description (*e.g.*, the Baltic Sea, Black Sea, and Red Sea).

Table 3

Ocean tide models validated using LAGEOS orbits

Model	Type	Mainly based on	Reference
CSR3.0	hydrological	Topex/Poiseidon (T/P)	Eanes and Bettadpur (1996)
CSR4.0A	hydrological	T/P, GRACE	Eanes (2004)
TOPEX4.0	hydrological	T/P	Egbert <i>et al.</i> (1994)
EOT08A	empirical	T/P, ERS-1/-2, GFO, Jason-1, Envisat	Savcenko and Bosch (2008)
FES2004	hydrodynamic	tide gauges, T/P, ERS	Lyard <i>et al.</i> (2006)
GOT00.2	hydrodynamic	tide gauges, T/P, ERS	Ray (1999)
CSRC Schwiderski	hydrodynamic	tide gauges	Schwiderski (1980)

FES2004 is the model that is recommended by the IERS Conventions 2010 (Petit and Luzum 2010). Compared to the CSR3.0, which was recommended by IERS Conventions 2003 (McCarthy and Petit 2004), FES2004 has the benefit that the treatment of the secondary waves is specified.

5.1 RMS of observation residuals and orbit predictions

Seven series of LAGEOS solutions are generated, each based on a different OTM. Table 4 shows the mean values of the RMS of the observations residuals of LAGEOS solutions. Although all solutions show RMS errors at a comparable level, an association of the solutions with four groups is possible. The first group with the smallest RMS of residuals contains those models which are based, to a great extent, on the T/P satellite mission. These are: CSR3.0, CSR4.0A, and TOPEX4.0 with the RMS values of 7.40, 7.41, and 7.44 mm, respectively. It seems that the models based on LEO altimetry satellite missions may give a benefit to higher satellite like LAGEOS. EOT08A model has an average RMS of observation residuals about 0.6 mm larger than the aforementioned models and 0.4 mm smaller than FES2004 and GOT00.2. EOT08A is based on six different satellite missions and a long time series of observations. The third group contains two hydrodynamic models with T/P observations, namely FES2004 and GOT00.2. The RMS of residuals for these models yield 8.41 mm, *i.e.*, it is about 1 mm (12%) worse than for exclusively T/P based models (CSR3.0 and TOPEX4.0). The oldest

Table 4

RMS of observation residuals and comparison between predicted and estimated 7-day LAGEOS orbits using the Helmert transformation (mean values for 2008)

Ocean tide model	RMS of residuals [mm]	Comparison of estimated and predicted orbits including the Helmert transformation			
		RMS of radial prediction [mm]	RMS of along-track prediction [mm]	RMS of out-of-plane prediction [mm]	Scale [ppb]
CSR3.0	7.40	30.1	387.7	200.5	-0.10
CSR4.0A	7.41	29.8	398.7	200.4	-0.05
TOPEX4.0	7.44	30.1	392.9	202.8	-0.10
EOT08A	8.02	30.8	497.2	257.9	0.04
FES2004	8.41	31.1	536.6	299.6	0.17
GOT00.2	8.41	31.2	542.7	303.3	0.15
CSRC Schwiderski	8.72	31.1	553.2	280.9	0.00

model, namely the CSRC (Schwiderski 1980) hydrodynamic model, is characterized by the largest RMS of the residuals of 8.72 mm.

The differences between solutions using different OTM (max. 1.32 mm of RMS) are slightly larger than the differences between solutions using different Earth gravity field models (max. 1.16 mm of RMS, see Sošnica *et al.* 2012). This implies that the modelling of ocean tides is one of the key factors influencing the quality of LAGEOS orbits.

Table 4 contains the result of the orbit comparisons between estimated and predicted orbits when the Helmert transformation parameters are estimated. A classification of all models into three groups is possible on the basis of the along-track (S) and the out-of-plane (W) components from Table 4. The first group, with RMS in S about 400 mm and in W about 200 mm, contains CSR3.0, CSR4.0A, and TOPEX4.0, *i.e.*, this group corresponds to the first group when classifying the models according to RMS. Then, the second group with EOT08A that has RMS about 500 mm in S and 260 mm in W . Finally, FES2004, GOT00.2, and the CSRC Schwiderski which have the RMS of S and W predictions at the level of 540 and 300 mm, respectively. The radial (R) component does not indicate any significant differences between compared models.

A comparison between predicted and estimated orbits without estimating Helmert parameters is shown in Fig. 2. The direct comparison of orbits (without estimating rotations, translations, and a scale) shows even bigger RMS in the S component, implying some difficulties in establishing the rota-

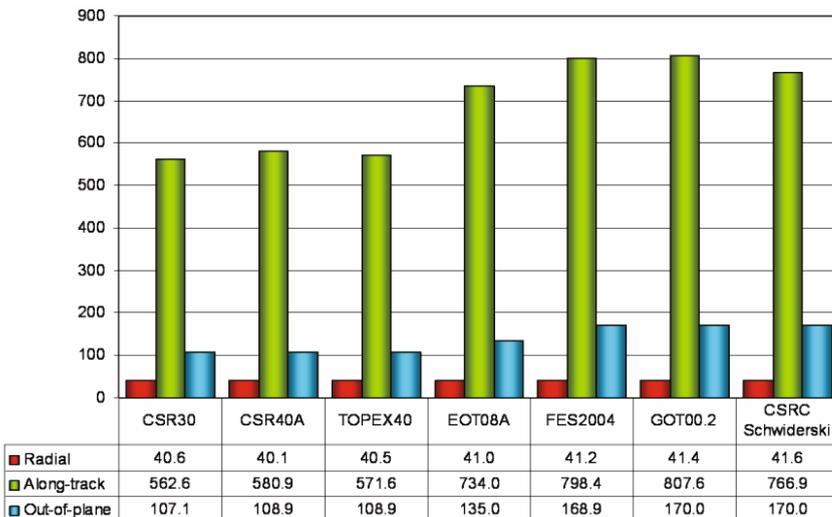


Fig. 2. Comparison between predicted and estimated orbits using different OTM, mean values for 2008; units: mm.

tion parameters in weekly SLR solutions. On the other hand, discrepancy of the W component is smaller than the discrepancy of the same component from Table 4.

The largest differences in OTM are between d/o 10 and 30. LAGEOS orbits are mostly sensitive to OTM up to d/o 8. Therefore, the older class hydrological OTM with a poorer spatial resolution as compared to the hydrodynamic models, may possibly lead to better LAGEOS orbits, provided that they contain good estimates of the low-degree spherical harmonic coefficients. The larger RMS of residuals in LAGEOS solutions using hydrodynamic models may be due to some deficiencies in the “datum definition” of these models. For lower orbiting satellites, *e.g.*, for GRACE, the best results can be achieved using EOT08A and FES2004 (Meyer *et al.* 2012), due to the satellite orbit sensitivity to all the tidal constituents, and not only to low degree coefficients of OTM.

5.2 Empirical orbit parameters

Empirical parameters absorb, to a certain extent, deficiencies in modelling the gravitational and non-gravitational forces acting on the LAGEOS satellites. Thus, the comparison of these parameters estimated using different OTM allows us to study the magnitude of errors still present in particular models.

5.2.1 Empirical orbit parameters in out-of-plane

Figure 3 top illustrates W_S values in 2008 for LAGEOS-1 and LAGEOS-2. The average value of W_S for LAGEOS-1 is about $5 \times 10^{-10} \text{ ms}^{-2}$, whereas it is negative for LAGEOS-2 and yields about $7 \times 10^{-10} \text{ ms}^{-2}$. The variations W_S series mainly correspond to variations of C_{20} in 2008 (see Sośnica *et al.* 2012).

Figure 3 does not illustrate any offsets for one or more OTM w.r.t. other models, but different variations can easily be noticed between days 105 and 222. In this period, FES2004 and GOT00.2 exhibit bigger amplitudes of variations in the consecutive weeks than the other models. In case of the LAGEOS-1 satellite, these variations could be associated with an eclipsing period. In 2008 there is one eclipsing period for LAGEOS-1 lasting from day 110 till day 203, which agrees with the periods of large variations in the FES2004 and GOT00.2 solutions.

The cosine once-per-revolution out-of-plane term (W_C , see Fig. 3 bottom) is very sensitive to OTM, as opposed to the different Earth gravity field models for which W_C does not exhibit any differences (*cf.* Sośnica *et al.* 2012). The differences of OTM are more apparent in case of W_C than W_S , because W_C is free from the impact of the variations of C_{20} . The CSR4.0A

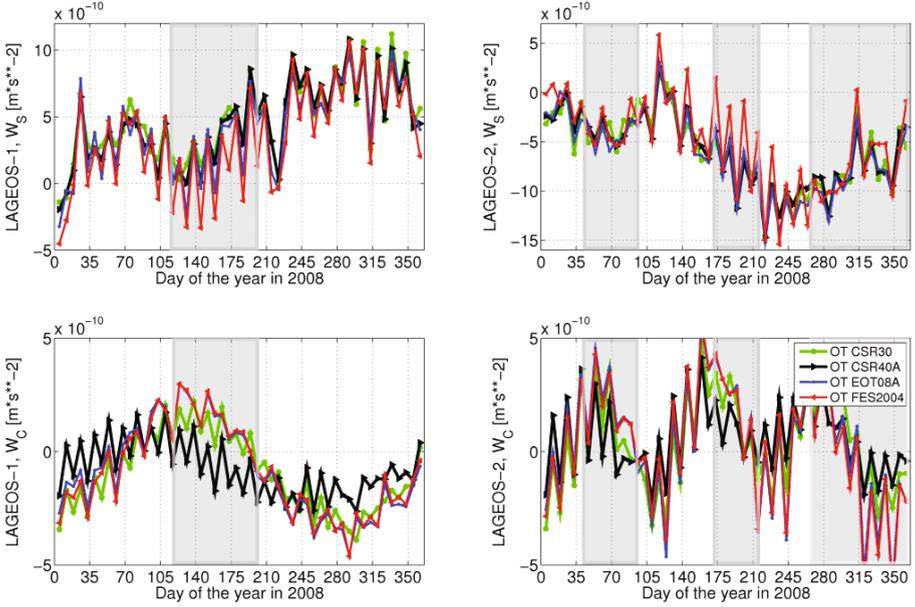


Fig. 3. Empirical orbit parameters in the out-of-plane direction for selected OTM. Eclipsing periods are shaded.

model shows the smallest amplitude and a different phase in the W_C series (see Fig. 3 bottom), whereas FES2004 and EOT08A indicate large variations in W_C for consecutive weeks.

The perturbing accelerations in the W direction influence the three orbital elements, defining the orientation of the satellite orbit in the inertial frame: right ascension of ascending node Ω , argument of perigee ω , and the inclination angle i . The relationships between the Euler angles and the accelerations in W read, according to Beutler (2005), as:

$$\frac{d\Omega}{dt} = \frac{r \sin u}{na^2 \sqrt{1-e^2} \sin i} W', \quad (3)$$

$$\frac{d\omega}{dt} = \frac{1}{e} \sqrt{\frac{p}{GM}} \left[-\cos v R' + \left(1 + \frac{r}{p} \right) \sin v S' \right] - \frac{r \sin u}{na^2 \sqrt{1-e^2} \tan i} W', \quad (4)$$

$$\frac{di}{dt} = \frac{r \cos u}{na^2 \sqrt{1-e^2}} W', \quad (5)$$

where W' , R' , and S' denote the accelerations in W , R , and S , respectively. The symbol a denotes a satellite semi-major axis, e is the orbital eccentricity, n is the mean motion, r is the length of satellite state vector, p is the semi-

latus rectum, ν denotes the true anomaly, and GM is a product of gravitational constant and Earth's mass. It can be seen in the three equations (3-5) that the right ascension of the ascending node and the inclination angle are solely perturbed by forces in \mathcal{W} , whereas the argument of perigee is also sensitive to perturbations in \mathcal{R} and \mathcal{S} .

Table 5 relates the impact of the major ocean tide constituents to the perturbations of orbital elements: Ω of LAGEOS-1, Ω of LAGEOS-2, and ω of LAGEOS-2. The ω of LAGEOS-1 is neglected here, because the orbit of LAGEOS-1 is near-circular, and thus, ω cannot be well established.

The spectral analysis of 10-year LAGEOS solutions using FES2004 and CSR4.0A reveals periods clearly referring to the tidal constituents perturbing orbital elements. Figure 4 exhibits the amplitudes of Fourier analysis of the \mathcal{W}_C for LAGEOS-1 (top) and LAGEOS-2 (bottom). The solution using FES2004 shows peaks around 14.1, 24.8, 28.2, 28.6 days, and the largest peak for 280 days for LAGEOS-1. For LAGEOS-2 the peaks are for 15.1, 24.8, 26.3, 28.8, 33.5, 86, 111, 137, and 285 days. Most of the peaks refer to

Table 5

Aliasing periods and perturbations of LAGEOS orbits due to the ocean tides for degree 2, after Iorio (2001)

Tide	Ω of LAGEOS-1		Ω of LAGEOS-2		ω of LAGEOS-2	
	Period [days]	Amplitude [mas]	Period [days]	Amplitude [mas]	Period [days]	Amplitude [mas]
K_1	1043.67	156.55	569.21	35.69	569.21	177.76
O_1	1043.67	151.02	569.21	34.43	569.21	171.48
P_1	221.35	11.49	138.26	3.00	138.26	14.95
Q_1	788.90	24.67	690.88	9.03	690.88	44.98
K_2	521.83	6.24	284.60	6.24	284.60	5.95
M_2	14.02	75.59	13.03	75.65	13.03	72.12
	521.83		284.60		284.60	
S_2	280.93	9.45	111.20	6.87	111.20	6.55
N_2	449.30	12.93	312.00	16.49	312.00	15.72
T_2	158.80	0.28	85.27	0.27	85.27	0.26
M_m	27.55	0.54	27.55	1.00	27.55	0.69
S_a	365.27	20.55	365.27	37.71	365.27	26.17
M_f	13.66	0.62	13.66	1.13	13.66	0.78
S_{Sa}	182.62	5.98	182.62	10.98	182.62	7.62

Note: The periods related to the tidal constituents found in the spectral analysis of empirical orbit parameters are indicated in bold.

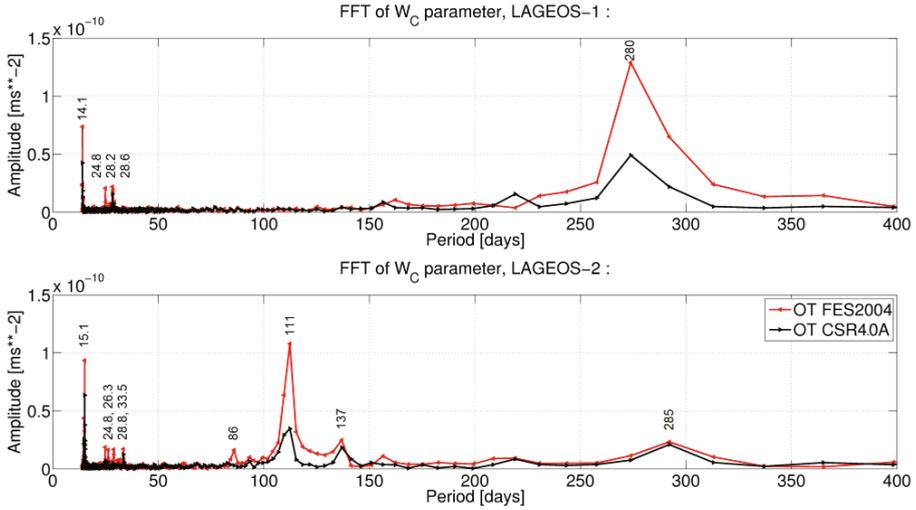


Fig. 4. Amplitude spectra (Fourier transform) of W_C empirical parameter from 10-year LAGEOS solutions using FES2004 and CSR4.0A.

the particular tidal constituents perturbing LAGEOS orbits. The periods of tidal constituents found in the spectral analysis of W_C or other empirical orbit parameters are indicated in bold in Table 5.

The largest peaks in Fig. 4, which are equal to 280 days and 111 days for LAGEOS-1 and LAGEOS-2, respectively, can be explained by the resonance between the diurnal and semi-diurnal tides and LAGEOS orbits. Due to the drift of ascending node $\dot{\Omega}_{L1}$, which is mainly due to C_{20} , the time interval between two consecutive passes (in the same direction) of the Sun through the orbital plane (the so-called draconitic year) can be expressed for LAGEOS-1 (with a prograde drift of $\dot{\Omega}_{L1} = 1050$ days) as:

$$\frac{360^\circ}{365.25 - \frac{360^\circ}{\dot{\Omega}_{L1}}} \text{days} = \frac{365.25 \cdot \dot{\Omega}_{L1}}{365.25 \cdot \dot{\Omega}_{L1} - 360^\circ} \text{days} = 560 \text{ days} . \quad (6)$$

In analogy for LAGEOS-2 ($\dot{\Omega}_{L2} = 570$ days, retrograde), the draconitic year is equal to 222 days. The diurnal tidal constituent S_1 imposes orbit perturbations having a period of the draconitic year, whereas the semi-diurnal constituent S_2 imposes orbit perturbations with a period of the half of the draconitic year (the semi-draconitic year). The perturbations due to S_2 have periods of 280 days and 111 days for LAGEOS-1 and LAGEOS-2, respectively. This explains the peaks found in Fig. 4 and the relations to the eclips-

ing seasons found in Fig. 3. Figure 4 suggests as well that deficiencies in the S_2 tide are the main quality limiting factor for the LAGEOS solutions. This can be, both, due to the deficiencies in the S_2 atmospheric tide, or due to the deficiencies in the S_2 ocean tide. The atmospheric tides are, however, much smaller as compared to ocean tides (Sośnica *et al.* 2013), and thus, cannot solely explain the large orbit perturbations. The differences between amplitudes of peaks in Fig. 4 suggest that, in particular, FES2004 contains the S_2 tide of inferior quality.

Most of the ocean tide models have largest residuals in the Polar Regions due to the absence of TOPEX/Poseidon and Jason data. The residuals, *e.g.*, of EOT08A for tidal constituent S_2 in the Arctic Sea and near Antarctica exceed 5 cm (Savcenko and Bosch 2008), whereas from the study in Section 4 we know that the LAGEOS orbits are sensitive to the ocean tides larger than 5 mm.

For high-orbiting satellites the semi-draconitic year corresponds exactly to the eclipsing periods of the satellite orbits (with an exception of satellites in sun-synchronous orbits). This explains the relation between eclipsing periods and differences in empirical parameters found in Fig. 3. In fact, the differences are not directly due to the eclipsing seasons, but due to the alias with S_2 tide, which imposes the orbit perturbations of the same period as the eclipsing seasons.

Figure 4 top shows, in addition, a small peak around 285 days for the LAGEOS-2 orbit. The residuals of K_2 in the Polar Regions exceed 3 cm, whereas the residuals of M_2 are of the order of 5 cm, but only for the shelf areas, *e.g.*, for EOT08A (Savcenko and Bosch 2008). Therefore, the peak of 285 days should be associated with deficiencies in K_2 , rather than M_2 , despite that both tides impose the perturbations on the LAGEOS satellites of the same period (see Table 5). The uncertainties for these tidal waves significantly exceed the sub-centimeter sensitivity of LAGEOS orbits to tidal waves found in Section 4. The perturbations imposed on LAGEOS-1 orbits due to the K_2 and M_2 tides have a period of 521.8 days; thus, an analysis of a longer time series is needed in order to detect these periods with a sufficient accuracy.

The large peaks from Fig. 4, of 14.1 and 15.1 days for LAGEOS-1 and LAGEOS-2, respectively, can be explained by the imposition of the annual tidal signal S_a and the groundtrack repeatability of satellites. The groundtrack repeatability of LAGEOS-1 is $gr_{L1} = 7\text{d } 23\text{h } 45\text{ min}$ and of LAGEOS-2 it is $gr_{L2} = 8\text{d } 22\text{h } 58\text{min}$. The peaks close to 14 days can be explained by overlapping of the groundtrack repeatability, S_a , and generated 7-day arcs. The overlapping period for LAGEOS-1 reads as:

$$2 \frac{2\pi}{\frac{2\pi}{gr_{L1}} + 7 \frac{2\pi}{365.25}} = 14.0 \text{ days} \quad (7)$$

and for LAGEOS-2:

$$2 \frac{2\pi}{\frac{2\pi}{gr_{L2}} + 7 \frac{2\pi}{365.25}} = 15.1 \text{ days} . \quad (8)$$

Lemoine *et al.* (2004) found that the amplitudes of some constituents in FES2004 are underestimated as compared to those obtained from the LAGEOS multi-year solutions. The differences in the amplitudes of 18.6-year tide and 9.3-year tide even reach 6000% and the phases are shifted even by 140°. Lemoine *et al.* (2004) found also large differences in S_a and S_{Sa} constituents, but they could be explained in terms of the mass displacement in the system Earth. Ray *et al.* (2014) tried to explain the spurious peaks around 14 days in GNSS series by subdaily Earth orientation parameter tide model errors. They also found the differences in tides between 10 and 20%, which correspond to differences of 2 cm at GPS altitudes. All in all, some of the ocean tide constituents require a further improvement, because of the large differences of their amplitudes in different OTMs.

In conclusion, the analysis of the W_C empirical parameter has revealed deficiencies in ocean tide constituents and some substantial differences between OTMs. The largest perturbations correspond to the S_2 tide and the resonance between the S_a tide and the groundtrack repeatability of LAGEOS orbits. Smaller perturbations due to the K_2 tide (alternatively due to M_2) have also been detected.

5.2.2 Empirical orbit parameters in along-track

Now, the empirical forces in the along-track direction are discussed. The S_0 reveals almost no differences for most of OTM (thus not shown here). The differences in S_0 are at the $8 \times 10^{-13} \text{ ms}^{-2}$ level; therefore, it can be stated that S_0 is insignificantly affected by different OTM.

Tapley *et al.* (1993) claim that errors in the odd-degree diurnal and semi-diurnal ocean tide coefficients determine variability in both the real and imaginary parts of eccentricity excitation, while variability in the odd zonal harmonics causes variations with the same spectrum in the real part of the excitation of eccentricity vector. Therefore, bigger variations in S_C are expected, which is related to the real part of the excitation of the eccentricity vector. Figure 5 illustrates the series of S_S and S_C for LAGEOS-1. As expected, S_C demonstrates bigger differences between ocean tide models. Nev-

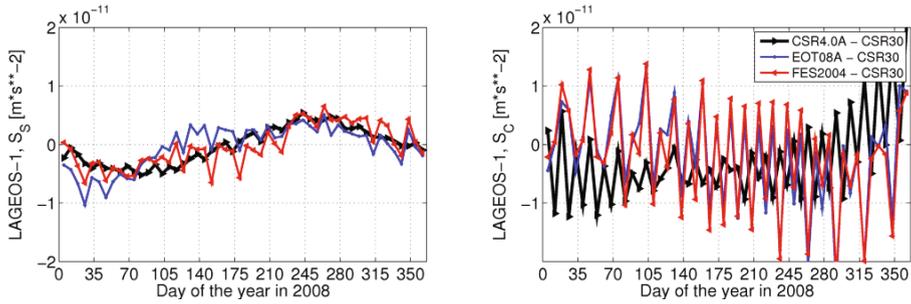


Fig. 5. Empirical LAGEOS-1 orbit once-per-revolution parameters in along-track for selected OTM. Differences w.r.t. CSR3.0.

ertheless, these differences are more than one order of magnitude smaller than for W_S and W_C , and smaller than those reported by Tapley *et al.* (1993).

For the S_S variations are smaller than for S_C for all models. However, the OTM affects the along-track empirical parameters only to a small extent. The major impact of OTM is reflected in the out-of-plane once-per-revolution parameters.

5.3 Orbit comparison

The orbits based on one OTM are compared with orbits based on all other models. The results of the direct comparison without estimating the Helmert parameters are presented in Table 6.

Table 6

Comparison between estimated orbits based on different ocean tide models:
RMS of direct orbit differences (mean values for 2008; in mm)

Model \ Model	CSR40 A	TOPEX40	EOT08A	FES2004	GOT00.2	SCR Schwiderski
CSR3.0	4.5	2.0	10.6	14.6	14.5	13.6
CSR40 A	–	4.5	8.4	12.2	12.0	12.7
TOPEX40	–	–	10.3	14.1	14.1	13.6
EOT08A	–	–	–	7.3	7.5	13.1
FES2004	–	–	–	–	1.6	16.0
GOT00.2	–	–	–	–	–	15.8

The OTM based on analysis of satellite altimetry data (CSR3.0, CSR4.0A, and TOPEX4.0) agree very well. Especially, the RMS for the differences between TOPEX4.0 and CSR3.0 is very small, *i.e.*, 2.0 mm. There is also an excellent consistency (1.6 mm) between FES2004 and GOT00.2

(both dynamical models with assimilation from observed tidal altimeter data). Orbits based on EOT08A are quite similar to those based on CSR4.0A, FES2004, and GOT00.2 (RMS of 8.4, 7.3, and 7.5 mm, respectively). Good agreement between orbits based on dynamic OTM and satellite altimetry OTM is only achieved for EOT08A, which acts like “a link” between these two types of OTM.

The agreement between satellite altimetry OTMs (CSR3.0, CSR4.0A, and TOPEX4.0) and dynamical OTMs (FES2004 and GOT00.2) is rather poor (the RMS is at the level 12-15 mm). The RMS of differences is largest when comparing orbits based on the CSRC Schwiderski model with other models, *i.e.*, the RMS of 13 mm to OTMs based on satellite altimetry data, and to 16 mm for other dynamical OTMs.

Chapter 6 of IERS Conventions 2010 (Petit and Luzum 2010) indicates a 7 mm 3D-RMS difference of LAGEOS-1 orbits, when using FES2004 and CSR3.0. In this study it was found that the orbit difference is even 14.6 mm for 7-day LAGEOS-1 and -2 solutions. However, such a difference strongly depends on the applied orbit parameterization and the procedure used for orbit comparison, *e.g.*, a direct comparison, or a comparison with the estimation of the Helmert transformation parameters, *etc.*

5.4 Earth Orientation Parameters

Now, the impact of different OTM on the estimation of Earth Orientation Parameters (EOP, *i.e.*, pole coordinates and length-of-day) from the LAGEOS solutions is investigated. Ocean tides are one of the main contributors to the EOP variations (Beutler 2005), and, thus, EOP estimations can be used as indicators of the OTM quality. The mean differences with respect to IERS-

Table 7
Comparison between estimated Earth orientation parameters and the IERS-08-C04 series from the solutions based on different OTM (mean values for 2008)

Model	Mean bias			RMS of differences		
	X pole [μ as]	Y pole [μ as]	LoD [μ s]	X pole [μ as]	Y pole [μ as]	LoD [μ s]
CSR3.0	49	1	-4	186	179	48
CSR4.0A	61	0	-3	186	182	48
TOPEX4.0	60	1	-4	188	179	48
EOT08A	82	8	-6	206	197	51
FES2004	94	8	-4	220	208	52
GOT00.2	91	8	-4	219	207	52
CSRC Schwiderski	84	25	-6	236	223	55

08-C04 series, as well as weighted RMS of pole coordinates and length-of-day values are shown in Table 7. The X pole coordinate shows a bias when comparing to C04 series, amounting from 49 μs for CSR3.0 to 94 μs for FES2004. The mean bias for the X pole coordinate is smallest for CSR3.0, CSR4.0A, and TOPEX4.0 models and largest for FES2004 and GOT00.2. The Y pole coordinate, as well as the length-of-day, do not show any significant biases with respect to IERS-08-C04 series. The RMS of estimated pole coordinates ranges from 186 and 179 μs for CSR3.0 to 236 and 223 μs for the SCRS Schwiderski model, for the X and Y components, respectively. This implies that the estimated EOP values can be degraded by about 25% when using older-class or improper OTM.

6. DISCUSSION AND CONCLUSIONS

In all tests the most appropriate OTM for LAGEOS orbits are those based on altimetry observations from TOPEX/Poseidon (*i.e.*, CSR3.0, TOPEX4.0), or models based on T/P containing additional observations derived from GRACE and other gravity satellite missions (*i.e.*, CSR4.0A, EOT08A). Hydrological models supported by T/P (*i.e.*, FES2004 and GOT00.2) show big discrepancies, and the hydrological model based almost uniquely on coastal tide gauges (*i.e.*, CSRC Schwiderski) shows the largest deviations in most cases.

Even though the differences in OTM are at cm or sub-cm level, they can be detected by LAGEOS, because LAGEOS satellites are sensitive to tidal waves bigger or equal to 5 mm, whereas the errors of some tidal constituents exceed 50 mm in the current models. The minimum requested degree/order of ocean tide model expansion is 8 when processing SLR observations to LAGEOS. Using the OTM up to d/o 30 is, however, recommended, in order to avoid the degradation of the orbit determined through the model truncation. There is a good agreement between OTM based on the same type of observations and a significant disagreement between models based on different assumptions. Only orbits based on the EOT08A model indicate a quite high convergence of both the hydrological and hydrodynamic models (see Table 6).

The empirical orbit parameters indicate the smallest variations for CSR3.0, CSR4.0A, TOPEX4.0, and EOT08A in W_S parameter and for CSR4.0A in W_C parameter. The spectral analysis of W_C and W_S series reveals the deficiencies in the S_2 constituent, especially in the Polar Regions, due to the lack of altimetry data above 66°N and below 66°S, and some minor deficiencies in the annual S_a and S_{sa} tides, as well.

When estimating gravity field parameters from LAGEOS, the W_C and W_S parameters are not estimated due to a direct correlation with C_{20} (Sośnica *et*

al. 2012). In such a case, all spurious peaks from W_C and W_S may leak to LAGEOS-derived gravity field coefficients, in particular to C_{20} , which is one of the SLR core products.

The problems of orbit alias with the S_2 tide are well-known for GRACE gravity field solutions (*e.g.*, Chen *et al.* 2009, Meyer *et al.* 2012). Remarkably, the S_2 constituent limits the GRACE capability of recovering C_{20} variations. In this study it was found that not only GRACE solutions are affected by deficiencies in the S_2 tide in the latest OTM, but also SLR solutions and LAGEOS-derived parameters suffer from S_2 mismodellings.

The study reveals that the current OTM have a larger impact on the LAGEOS orbits than the *a priori* Earth gravity field models. The mean differences of RMS of residuals between solutions using different OTM (max. 1.32 mm of RMS) are larger than the mean differences between solutions using different Earth gravity field models with maximum values of RMS amounting to 1.16 mm (Sośnica *et al.* 2012).

The background models have a crucial impact not only on LAGEOS orbits, but also on all other SLR-derived parameters. In particular, polar motion or length-of-day are very sensitive to *a priori* ocean tide models. The difference in RMS of SLR-derived pole coordinates with respect to IERS-08-C04 series varies even up to 25% when using different ocean tide models.

Acknowledgments. A sincere gratitude to the International Laser Ranging Service is expressed for publishing the SLR data and to a global network of SLR stations for providing state-of-the-art SLR observations.

References

- Altamimi, Z., X. Collilieux, and L. Métivier (2011), ITRF2008: an improved solution of the international terrestrial reference frame, *J. Geod.* **85**, 8, 457-473, DOI: 10.1007/s00190-011-0444-4.
- Beutler, G. (2005), *Methods of Celestial Mechanics*, Springer, Berlin Heidelberg.
- Bianco, G., R. Devoti, M. Fermi, V. Luceri, P. Rutigliano, and C. Sciarretta (1998), Estimation of low degree geopotential coefficients using SLR data, *Planet. Space Sci.* **46**, 11-12, 1633-1638, DOI: 10.1016/S0032-0633(97)00215-8.

- Bizouard, C., and D. Gambis (2014), The combined solution C04 for Earth Orientation Parameters consistent with International Terrestrial Reference Frame 2008, IERS Earth Orientation Product Centre, Paris, France.
- Chen, J.L., C.R. Wilson, and K.W. Seo (2009), S_2 tide aliasing in GRACE time-variable gravity solutions, *J. Geod.* **83**, 7, 679-687, DOI: 10.1007/s00190-008-0282-1.
- Cheng, M.K., C.K. Shum, and B.D. Tapley (1997), Determination of long-term changes in the Earth's gravity field from satellite laser ranging observations, *J. Geophys. Res.* **102**, B10, 22377-22390, DOI: 10.1029/97JB01740.
- Dach, R., U. Hugentobler, P. Fridez, and M. Meindl (2007), Bernese GPS software version 5.0, Astronomical Institute, University of Bern, Bern, Switzerland.
- Dow, J.M. (1990), Ocean tides and tectonic plate motions in high precision orbit determination, *Adv. Space Res.* **10**, 3-4, 229-238, DOI: 10.1016/0273-1177(90)90353-2.
- Eanes, R.J. (2004), CSR4.0A global ocean tide model, Center for Space Research, University of Texas, Austin, USA, <ftp://ftp.csr.utexas.edu/pub/tide>.
- Eanes, R.J., and S. Bettadpur (1996), The CSR 3.0 global ocean tide model: Diurnal and semi-diurnal ocean tides from TOPEX/POSEIDON altimetry, Technical Report CRS-TM-96-05, Center for Space Research, University of Texas, Austin, USA.
- Egbert, G.D., and S.Y. Erofeeva (2002), Efficient inverse modeling of barotropic ocean tides, *J. Atmos. Oceanic Technol.* **19**, 2, 183-204, DOI: 10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2.
- Egbert, G.D., A.F. Bennett, and M.G.G. Foreman (1994), TOPEX/POSEIDON tides estimated using a global inverse model, *J. Geophys. Res.* **99**, C12, 24821-24852, DOI: 10.1029/94JC01894.
- Folkner, W.M., P. Charlot, M.H. Finger, J.G. Williams, O.J. Sovers, X.X. Newhall, and E.M. Standish Jr. (1994), Determination of the extragalactic-planetary frame tie from joint analysis of radio interferometric and lunar laser ranging measurements, *Astron. Astrophys.* **287**, 1, 279-289.
- Iorio, L. (2001), Earth tides and Lense–Thirring effect, *Celest. Mech. Dyn. Astr.* **79**, 3, 201-230, DOI: 10.1023/A:1017963306722.
- Kolaczek, B., H. Schuh, and D. Gambis (eds.) (2000), High frequency to subseasonal variations in Earth Rotation, IERS Technical Note No. 28, Paris: Central Bureau of IERS – Observatoire de Paris, 91 pp.
- Lejba, P., and S. Schillak (2011), Determination of station positions and velocities from laser ranging observations to Ajisai, Starlette and Stella satellites, *Adv. Space Res.* **47**, 4, 654-662, DOI: 10.1016/j.asr.2010.10.013.
- Lemoine, J.M., R. Biancale, and G. Bourda (2004), Processing 18.6 years of Lageos data. In: *Proc. 14th Int. Laser Ranging Workshop, 7-11 June 2004, San Fernando, Spain*.

- Lyard, F., F. Lefevre, T. Letellier, and O. Francis (2006), Modelling the global ocean tides: modern insights from FES2004, *Ocean Dynam.* **56**, 5-6, 394-415, DOI: 10.1007/s10236-006-0086-x.
- Maier, A., S. Krauss, W. Hausleitner, and O. Baur (2012), Contribution of satellite laser ranging to combined gravity field models, *Adv. Space Res.* **49**, 3, 556-565, DOI: 10.1016/j.asr.2011.10.026.
- McCarthy, D.D., and G. Petit (2004), IERS Conventions 2003, IERS Technical Note No. 32, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, Germany, 127 pp.
- Mendes, V.B., and E.C. Pavlis (2004), High-accuracy zenith delay prediction at optical wavelengths, *Geophys. Res. Lett.* **31**, 14, L14602, DOI: 10.1029/2004GL020308.
- Meyer, U., A. Jäggi, and G. Beutler (2012), Monthly gravity field solutions based on GRACE observations generated with the Celestial Mechanics Approach, *Earth Planet. Sci. Lett.* **345-348**, 72-80, DOI: 10.1016/j.epsl.2012.06.026.
- Pavlis, N.K., S.A. Holmes, S.C. Kenyon, and J.K. Factor (2012), The development and evaluation of the Earth Gravitational Model 2008 (EGM2008), *J. Geophys. Res.* **117**, B4, B04406, DOI: 10.1029/2011JB008916.
- Pearlman, M.R., J.J. Degnan, and J.M. Bosworth (2002), The International Laser Ranging Service, *Adv Space Res.* **30**, 2, 125-143, DOI: 10.1016/S0273-1177(02)00277-6.
- Petit, G., and B. Luzum (2010), IERS Conventions 2010, IERS Technical Note No. 36, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, Germany, 179 pp.
- Ponchaut, F., F. Lyard, and C. Le Provost (2001), An analysis of the tidal signal in the WOCE Sea level dataset, *J. Atmos. Oceanic Technol.* **18**, 1, 77-91, DOI: 10.1175/1520-0426(2001)018<0077:AAOTTS>2.0.CO;2.
- Ray, J., J. Griffiths, X. Collilieux, and P. Rebischung (2014), Subseasonal GNSS errors in IGS products. **In:** *Proc. EGU General Assembly 2014, 27 April – 2 May 2014, Vienna, Austria*, EGU2014-8504.
- Ray, R.D. (1999), A global ocean tide model from TOPEX/POSEIDON altimetry: GOT99.2, NASA Tech. Memo. 209478, Goddard Space Flight Centre, Greenbelt, MD, USA.
- Ray, R.D., and R.M. Ponte (2003), Barometric tides from ECMWF operational analyses, *Ann. Geophys.* **21**, 8, 1897-1910, DOI: 10.5194/angeo-21-1897-2003.
- Rutkowska, M., and M. Jagoda (2010), Estimation of the elastic Earth parameters using the SLR LAGEOS 1 and LAGEOS 2 data, *Acta Geophys.* **58**, 4, 705-716, DOI: 10.2478/s11600-009-0062-1.
- Rutkowska, M., and M. Jagoda (2012), Estimation of the elastic Earth parameters using SLR data for the low satellites STARLETTE and STELLA, *Acta Geophys.* **60**, 4, 1213-1223, DOI: 10.2478/s11600-012-0045-5.

- Savcenko, R., and W. Bosch (2008), EOT08a – empirical ocean tide model from multi-mission satellite altimetry, DGFI Report No. 81, Deutsches Geodätisches Forschungsinstitut (DGFI), München, Germany.
- Schillak, S., and E. Wnuk (2003), The SLR stations coordinates determined from monthly arcs of LAGEOS-1 and LAGEOS-2 laser ranging in 1999-2001, *Adv. Space. Res.* **31**, 8, 1935-1940, DOI: 10.1016/S0273-1177(03)00169-8.
- Schillak, S., E. Wnuk, H. Kunimori, and T. Yoshino (2006), Short note: Crustal deformation in the Key Stone network detected by satellite laser ranging, *J. Geod.* **79**, 12, 682-688, DOI: 10.1007/s00190-005-0020-x.
- Schutz, B.E., M.K. Cheng, C.K. Shum, R.J. Eanes, and B.D. Tapley (1989), Analysis of earth rotation solution from Starlette, *J. Geophys. Res.* **94**, B8, 10167-10174, DOI: 10.1029/JB094iB08p10167.
- Schwiderski, E.W. (1980), On charting global ocean tides, *Rev. Geophys.* **18**, 1, 243-268, DOI: 10.1029/RG018i001p00243.
- Shum, C.K., P.L. Woodworth, O.B. Andersen, G.D. Egbert, O. Francis, C. King, S.M. Klosko, C. Le Provost, X. Li, J.M. Molines, M.E. Parke, R.D. Ray, M.G. Schlax, D. Stammer, C.C. Tierney, P. Vincent, and C.I. Wunsch (1997), Accuracy assessment of recent ocean tide models, *J. Geophys. Res.* **102**, C11, 25173-25194, DOI: 10.1029/97JC00445.
- Smith, D.E., and D.L. Turcotte (eds.) (1993), *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*, Geodynamics Series, Vol. 24, American Geophysical Union, Washington DC.
- Sośnica, K. (2014), *Determination of Precise Satellite Orbits and Geodetic Parameters using Satellite Laser Ranging*, Ph.D. Thesis, Astronomical Institute, Faculty of Science of the University of Bern, Bern, Switzerland, 253 pp.
- Sośnica, K., D. Thaller, A. Jäggi, R. Dach, and G. Beutler (2012), Sensitivity of Lageos orbits to global gravity field models, *Artif. Sat.* **47**, 2, 47-65, DOI: 10.2478/v10018-012-0013-y.
- Sośnica, K., D. Thaller, R. Dach, A. Jäggi, and G. Beutler (2013), Impact of loading displacements on SLR-derived parameters and on the consistency between GNSS and SLR results, *J. Geod.* **87**, 8, 751-769, DOI: 10.1007/s00190-013-0644-1.
- Sośnica, K., A. Jäggi, D. Thaller, G. Beutler, and R. Dach (2014), Contribution of Starlette, Stella, and AJISAI to the SLR-derived global reference frame, *J. Geod.* **88**, 8, 789-804, DOI: 10.1007/s00190-014-0722-z.
- Tapley, B.D., B.E. Schutz, R.J. Eanes, J.C. Ries, and M.M. Watkins (1993), Lageos laser ranging contributions to geodynamics, geodesy, and orbital dynamics. **In:** D.E. Smith and D.L. Turcotte (eds.), *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*, Geodynamics Series, Vol. 24, American Geophysical Union, Washington DC, 147-173, DOI: 10.1029/GD024p0147.

-
- Wünsch J., P. Schwintzer, and S. Petrovic (2005), Comparison of two different ocean tide models especially with respect to the GRACE satellite mission, Scientific Technical Rep. STR05/08, GeoForschungsZentrum Potsdam, Germany.
- Zahran K.H., G. Jentzsch, and G. Seeber (2006), Accuracy assessment of ocean tide loading computations for precise geodetic observations, *J. Geodyn.* **42**, 4-5, 159-174, DOI: 10.1016/j.jog.2006.07.002.

Received 5 August 2014

Received in revised form 19 October 2014

Accepted 15 December 2014