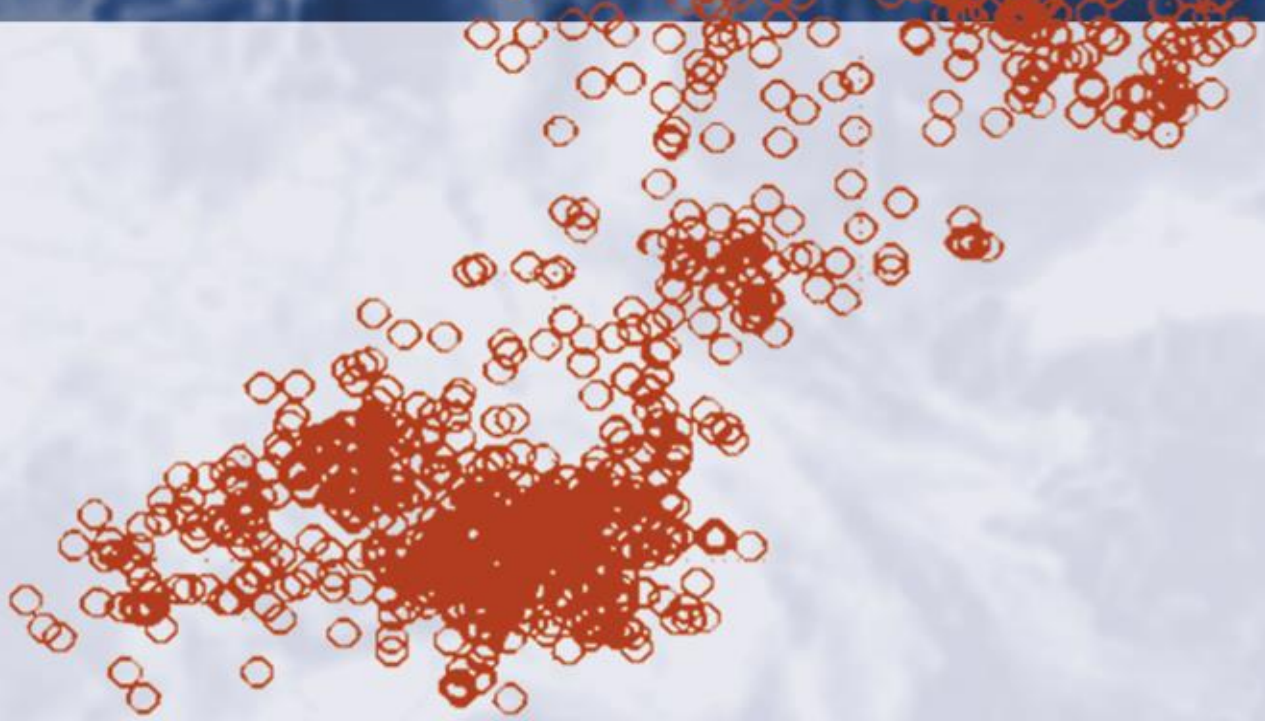
ISSN 1895-6572

Volume 58 • Number 6 • December 2010

Acta Geophysica



VERSITA

Institute of Geophysics Polish Academy of Sciences

Springer

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 963-972
DOI: 10.2478/s11600-010-0014-9

A New Magnetic Index Based on the External Part of Vertical Geomagnetic Variations

Tomasz ERNST, Jerzy JANKOWSKI, and Krzysztof NOWOŻYNSKI

Institute of Geophysics, Polish Academy of Sciences, Warszawa, Poland
e-mails: ternst@igf.edu.pl (corresponding author),
jerzy@igf.edu.pl, kn@igf.edu.pl

A b s t r a c t

Variations of geomagnetic components X, Y, and Z recorded in 19 Intermagnet European observatories in 2004 were analysed. The original data from all observatories were preliminarily processed. In the first step, periods longer than three hours were filtered out. In the second step, variations of vertical geomagnetic component Z were separated into external and internal parts. We introduced a non-dimensional index η defined as the square root of a ratio of the energy of the external part of the vertical component to that of the horizontal components.

Maps of the surface distribution of a new magnetic index η for the area of Europe at selected time periods were created, and their time changes are presented. The time changes of η for selected observatories are also shown. Moreover, we discuss a very interesting phenomenon we discovered, that has never been described in geophysical literature. Namely, in the recordings of all the observatories we noticed the presence of very regular variations, observed almost exclusively in the vertical component Z, which is quite unusual. These regular variations occur in the form of sinusoidal "wave packets". The amplitudes of these variations do not depend on the geomagnetic latitude and appear in the records of all the observatories we analyzed. They occur in quiet days, which suggests that their source is in the ionosphere.

Key words: geomagnetic variations, ionosphere, magnetic activity, new magnetic index η.

1. INTRODUCTION

An analysis of magnetic field variations of periods from few minutes to some hours often involves several simplifications. The most popular is the assumption that the magnetic activity manifests itself by variations of horizontal components alone (index *K*); the argument is that the variations of the vertical component *Z* are not only contaminated by the induction effect, but also their amplitudes are much lower than those of horizontal components. In magnetotelluric surveys, it is assumed that in the geomagnetic field variations in the middle geographical latitudes the external field appears exclusively on the horizontal components.

In the earlier work (Ernst and Jankowski 2005) we analysed the justification of this latter assumption and it turned out that it is not always true. In the present paper, making use of the ample Intermagnet data basis, we decided to verify the former assumption by analyzing the separated external part of the vertical geomagnetic field component.

The systematic analysis was made with the use of a non-dimensional index *E* defined as the ratio of energies of this external part of vertical component to the energy of horizontal components. The data at our disposal were the year-long, one-minute time series from 18 European observatories belonging to the Intermagnet. In the present paper, we discuss the results obtained, in some cases unexpected. We think, they may be of importance for ionospheric research.

2. SEPARATION OF THE EXTERNAL PART OF VERTICAL GEOMAGNETIC COMPONENT VARIATIONS

The variations of geomagnetic components *X*, *Y*, and *Z* recorded in 19 Intermagnet European observatories in 2004 were analysed. Location of the selected observatories is shown in Fig. 1. The list of the observatories is presented in Table 1.

The original data from all the observatories were preliminarily processed. In the first step, the periods longer than three hours were filtered out. We used symmetric optimal filters calculated according to Parks and McClellan's (1972) algorithm. In the second step, the vertical geomagnetic component Z was separated into external and internal parts.

The method of separation is the same as that in our cited paper (Ernst and Jankowski 2005). Below we describe it shortly. When the external magnetic field is represented by horizontal components only, the vertical component fulfils, in the time domain, the linear relation (Wiel̨adek and Ernst 1977)

$$Z(t) = a(t) * X(t) + b(t) * Y(t)\,, \tag{1}$$

where $X(t)$, $Y(t)$, and $Z(t)$ are the recorded values of three components varia-

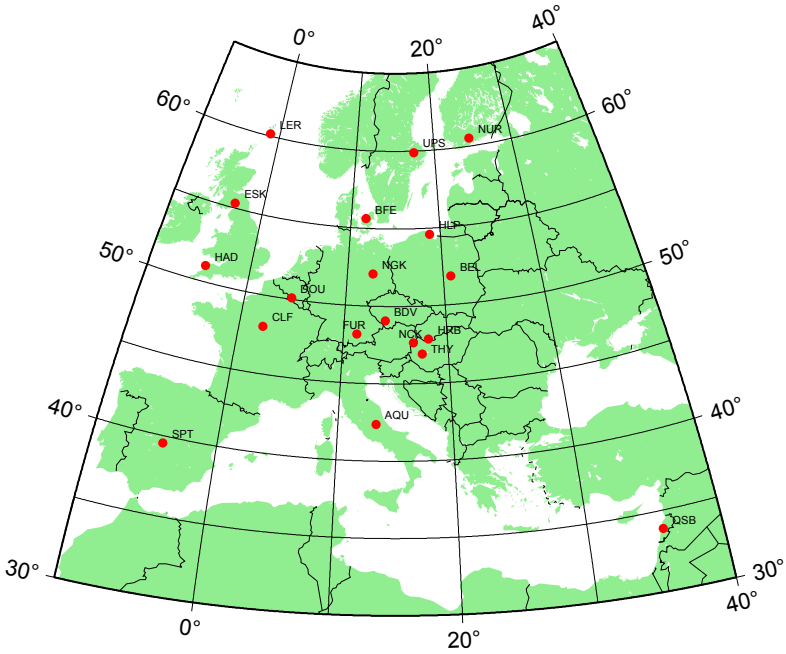


Fig. 1. Location of the selected magnetic observatories. Colour version of this figure is available in electronic edition only.

Table 1

List of selected Intermagnet Magnetic Observatories

Codes	Observatories	Coordinates	
		latitude	longitude
AQU	L'Aquila, Italy	42°23.0′N	13°19.0′E
BDV	Budkov, Czech Republic	49°04.8′N	14°00.9′E
BEL	Belsk, Poland	51°50.2′N	20°47.5′E
BFE	Brorfelde, Denmark	55°37.8′N	11°40.2′E
CLF	Chambon la Foret, France	48°01.0′N	02°16.0′E
DOU	Dourbes, Belgium	50°06.0′N	04°35.9′E
EBR	Ebro, Spain	40°49.3′N	00°29.6′E
ESK	Eskdalemuir, Scotland, UK	55°18.0′N	03°12.0′W
FUR	Furstenfeldbruck, Germany	48°09.9′N	11°16.6′E
HAD	Hartland, England	51°00.0′N	04°30.0′W
HLP	Hel, Poland	54°36.5′N	18°49.0′E
HRB	Hurbanovo, Slovakia	47°52.5′N	18°11.4′E
LER	Lerwick, Shetland Islands, UK	60°06.0′N	01°12.0′W
NCK	Nagycenk, Hungary	47°37.8′N	16°43.2′E
NGK	Niemegk, Germany	52°04.3′N	12°40.5′E
NUR	Nurmijarvi, Finland	60°30.5′N	24°39.3′E
QSB	Qsaybeh, Libanon	33°52.3′N	35°38.6′E
THY	Tihany, Hungary	46°54.0′N	17°53.6′E
UPS	Uppsala, Sweden	59°54.2′N	17°21.2′E

tions of the total magnetic field, $a(t)$ and $b(t)$ represent the impulse responses, $*$ denotes the convolution.

In such a case, $Z(t)$ has the internal (inductive) origin, connected with internal currents system. Using well selected data sets from the Intermagnet Data Base we can estimate, using eq. (1), the impulse responses $\tilde{a}(t)$ and $\tilde{b}(t)$ for each observatory.

In a general case, the vertical component $Z(t)$ resulting from the convolution of impulse responses $\tilde{a}(t)$ and $\tilde{b}(t)$ with the recorded horizontal components variations $X(t)$ and $Y(t)$ has the internal origin only. Hence,

$$Z_i(t) = \tilde{a}(t) * X(t) + \tilde{b}(t) * Y(t) . \tag{2}$$

Finally, we can estimate the external part of the vertical field variations, basing on the formula:

$$Z_e(t) = Z(t) - Z_i(t) . \tag{3}$$

The external part Z_e calculated in this manner may also contain non-inductive variations of internal origin (artificial disturbances). However,

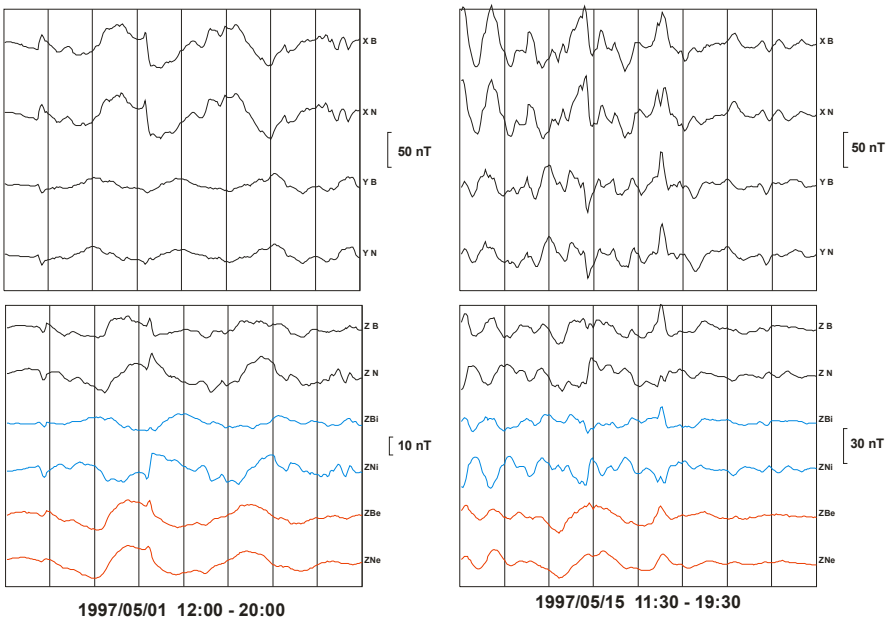


Fig. 2. Examples of separation of vertical component Z into internal (inductive) and external parts for two observatories: Belsk (BEL) and Niemegk (NGK). X_B, Y_B, Z_B – variations recorded at Belsk; X_N, Y_N, Z_N – variations recorded at Niemegk; Z_{Bi}, Z_{Ni} – internal parts of the Z component for Belsk and Niemegk; Z_{Be}, Z_{Ne} – external parts of the Z component for Belsk and Niemegk. Note that the registered vertical components Z_B and Z_N differ very much, while the external parts are very similar.

since the magnetic observatories are usually located in low-noise places, we can make the assumption that they are relatively small. This simple method is very efficient, as illustrated by an example of two observatories, Belsk and Niemegk, located close to each other (Fig. 2). We see that the Z component records in these two observatories (blue) differ a lot, while their external parts after separation (red) are practically the same.

3. AN ANALYSIS OF CHANGES OF INDEX η

The separation of the vertical component was made, using the procedure described above, for all selected observatories. In further analysis we will apply the index η defined as follows:

$$\eta = \sqrt{\frac{P_{Z\text{ext}}}{P_X + P_Y}},$$

where $P_{Z\text{ext}}$, P_X, and P_Y are the energies of components X, Y and $Z\text{ext}$ in the period range from 240 to 10 800 s. Energies $P_{Z\text{ext}}$, P_X, and P_Y were calculated in the time domain according to Parceval theorem. Index η was calculated for 2-hour intervals. Although the calculations themselves were not much complicated, the amount of material was so huge that we are unable to present all the results, showing just some typical examples. In Fig. 3 we show variations of parameter η for two months from some selected observatories. The values of η were calculated every hour for 2-hour intervals. It is seen from Fig. 3 that:

- The variations in all the observatories are alike.
- The variations exhibit large dynamics. Sometimes index η is in excess of 2, which means that the $Z\text{ext}$ component variations are greater than the horizontal components ones.
- The maximum amplitudes occur around noon.
- Changes of parameter η show distinct daily periodicity.

After this analysis of variations of parameter η we attempted an overall analysis of space-time changes. To this end, we compiled two sets of surface distribution maps of this parameter over Europe for various instants of time. One set was constructed for a day with weak magnetic activity, $\Sigma K = 6$ (Fig. 4), and the other for a very disturbed day, $\Sigma K = 51$ (Fig. 5), where K is the index of magnetic activity that is usually applied.

The features observed on these two map sets strongly differ. For large activity period, index η is large for the area of polar cusp and quickly decreases in Central Europe. This is understandable, since the perturbations are mainly due to the currents coming from the magnetosphere to the cusps and flowing therein. Quite different pattern is visible from the maps made for the

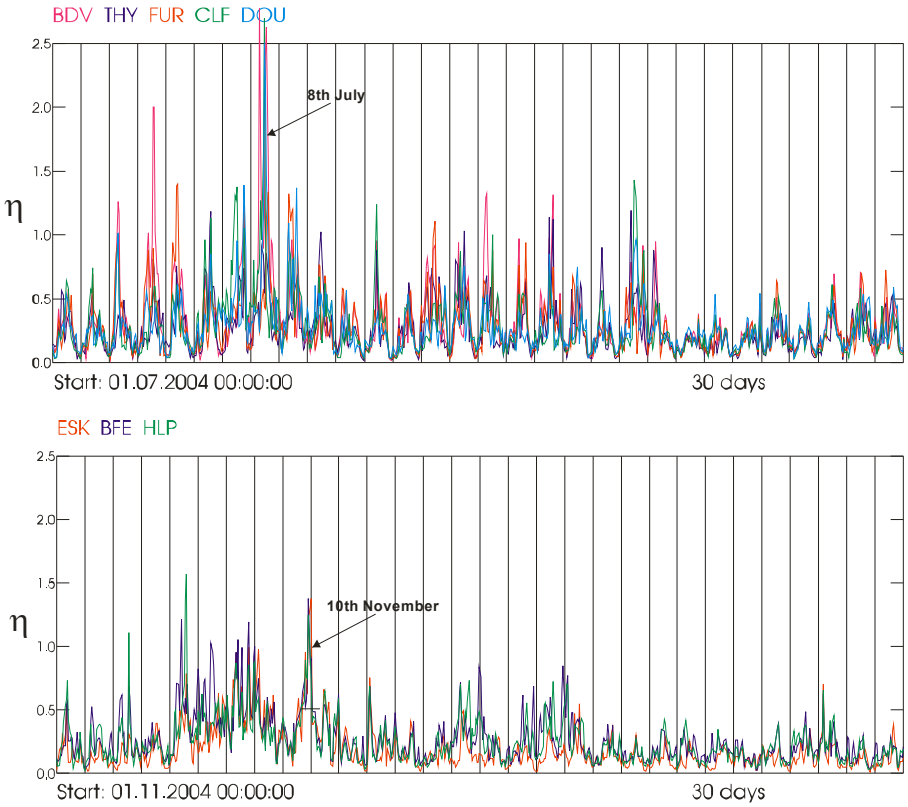


Fig. 3. Two examples of time changes of parameter η for selected observatories over one month. The changes show a clear daily periodicity.

very weak activity periods. In this case, the larger values occur in central Europe, their maximum value appearing around noon (see Fig. 4). Additionally it may be noticed that for magnetically quiet periods the values of η are the greatest. We also observe considerable dynamics of the changes in the maximum η values distribution. The maximum values seem to be displaced westwards. At the end of our analysis (Fig. 6), we plotted the recorded Z_{ext} component values for the selected 16 observatories for a quiet day of 8 July 2004. This diagram is, in our opinion, the most important part of the paper.

In the recordings of all the observatories (Fig. 6) we noticed the presence of very regular variations, observed almost exclusively in the vertical component Z, which is quite unusual. These regular variations occur in the form of sinusoidal ***wave packets***, whose amplitude does not exceed 10 nT and the period is of about 2 hours.

The amplitude of these variations is similar in various observatories, and does not depend on geomagnetic latitude, which suggests that their source is relatively close, not in the polar cusp. The amplitudes of these variations are the smallest at night, and the greatest around noon. Small differences in the external part of the vertical geomagnetic component point to existence of local inhomogeneities in the ionosphere.

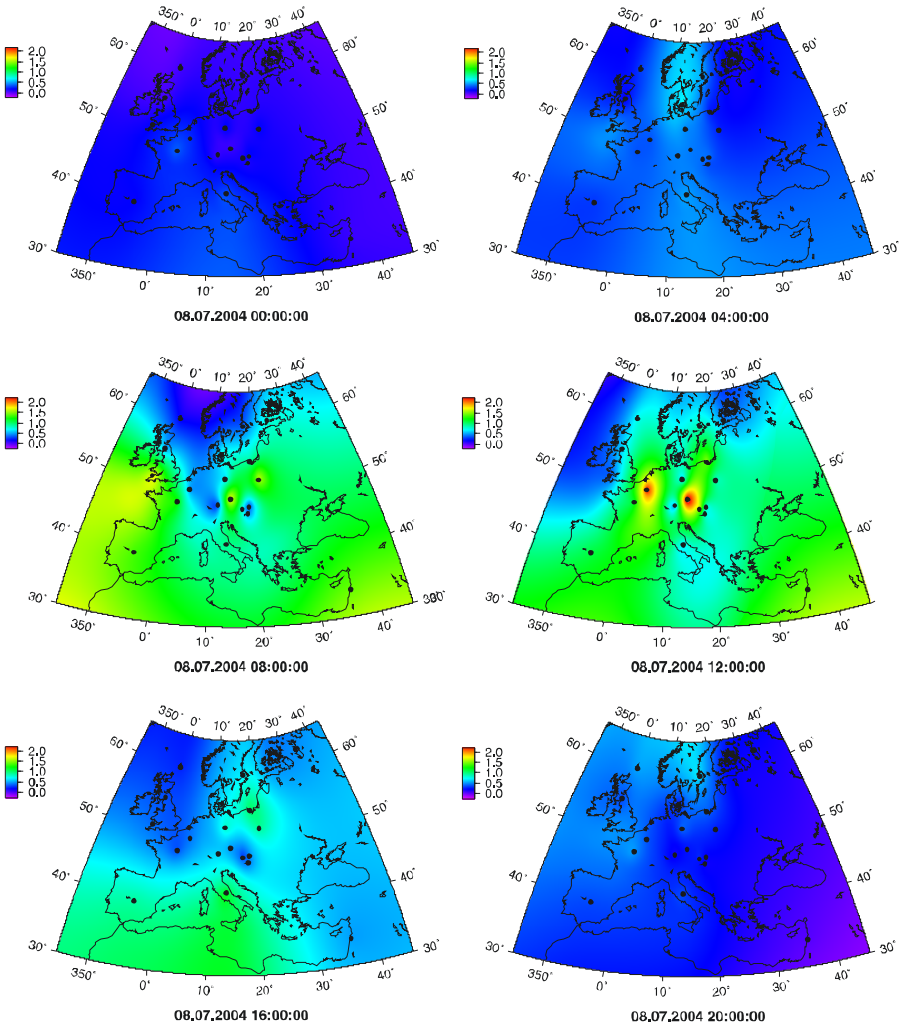


Fig. 4. Maps of surface distribution of parameter η over Europe for a selected, very quiet day of 8 July 2004 ($\Sigma K = 6$). The maps indicate a large dynamics of changes. The greatest values of η occur in the 40-50° latitudes and appear around noon.

The values of index η vary within rather large range. In extreme cases it is even in excess of two, which means that the amplitude of variations of the vertical component external part is much greater than the amplitudes of horizontal components. Interestingly, the large values of index η appear in quiet days.

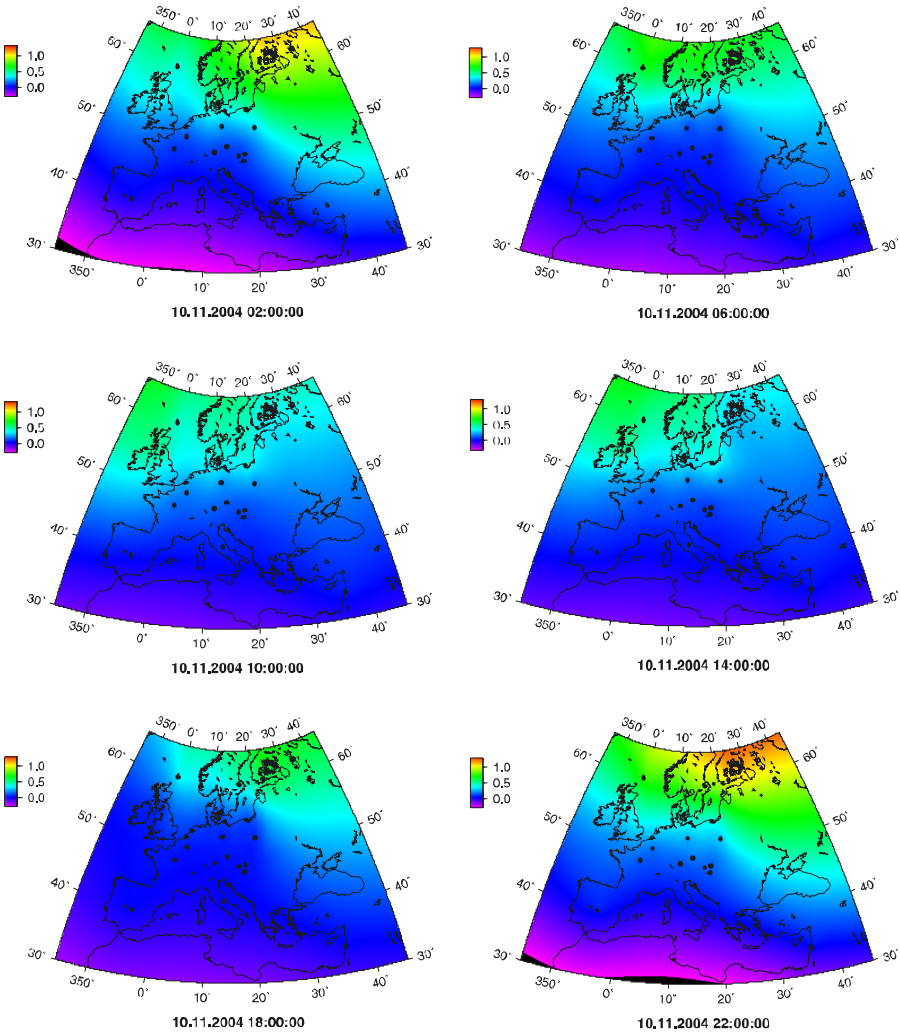


Fig. 5. Maps of surface distributions of parameter η over Europe for a selected, very strongly disturbed day of 11 November 2004 ($\Sigma K = 51$). In this case, it is seen that when the sources are within the polar cusp, the dynamics of changes is much smaller, and the values of η grow with growing latitude.

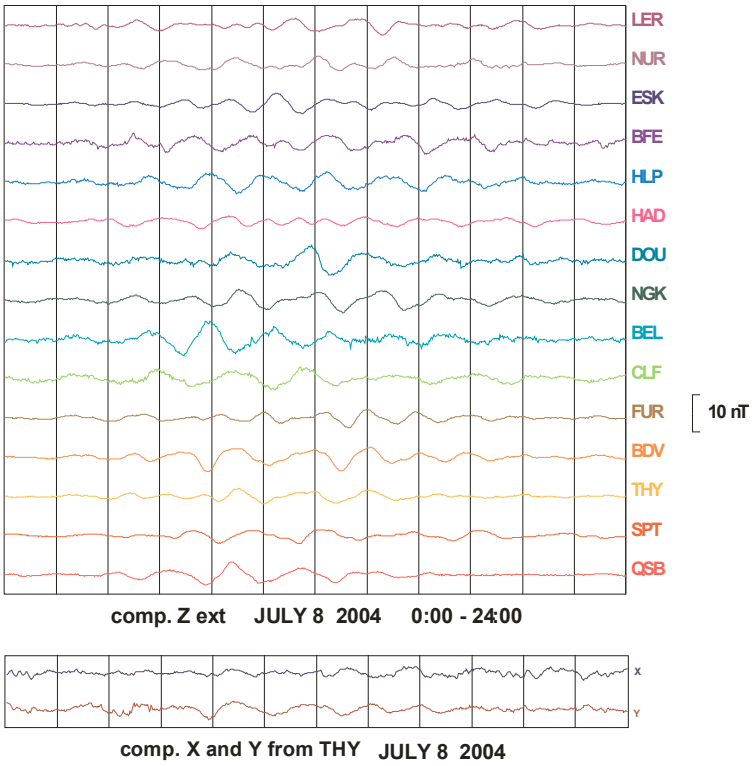


Fig. 6. An example of simultaneous variations of the external part of component Z for several European observatories on 8 July 2004 (24 hours). Note that in all the observatories we can see the presence of very regular variations. At the bottom, the variations of the horizontal magnetic components from Tihany Observatory are presented.

To assess the magnetic activity, routine procedure in the observatories is to use the recordings of horizontal component variations only. However, it turns out that, after eliminating the effect of currents induced in the Earth, the vertical component variations bring valuable, independent information about the system of currents in the ionosphere.

It seems therefore desirable, especially with a view to quiet-day ionosphere monitoring, to include the calculation of index η into the routine procedures in geomagnetic observatories.

We have not found in the literature any reference to variations of this type, so we treat them as our own finding. We hope, the observed feature will draw attention of other researchers. We do not attempt to create a model of generation of the observed variations. We think, though, that several facts give grounds for believing that the sources of these variations are located in the ionosphere, maybe in the form of electric current vortices.

4. CONCLUSIONS

The following conclusions may be drawn from our study:

- We discovered sequences of geomagnetic field variations, 2 hours in period, observed most often in the vertical component on quiet days, seems to be a novelty in the analysis of morphology of the natural geomagnetic field variations.
- While monitoring the state of the ionosphere, it seems unjustified to use only the horizontal component variations. The vertical component variations, after eliminating the effect of currents induced within the Earth, bring independent and meaningful information on the system of currents in the ionosphere (ionosphere state). Hence, the index η could be an important element of the ionosphere monitoring. Routine calculation of index η at the observatories, and its availability *via* Intermagnet, would be very useful for ionospheric studies.
- We suggest to eliminate Z_e or make a detailed initial selection of data before the estimation of tippers in geomagnetic soundings.

Acknowledgment. This research was supported by the Ministry of Science and Higher Education under Grant N N307 2558 33. The results presented in this paper rely on data collected at Magnetic Observatories. We thank the national institutes that support them and the INTERMAGNET for promoting high standards of magnetic observatory practice.

References

Ernst, T., and J. Jankowski (2005), On the plane wave approximation of the external geomagnetic field in the regional induction study, *Izvestiya, Physics Solid Earth* **41**, 5, 363-370.

Parks, T.W., and J.H. McClellan (1972), Chebyshev approximation for nonrecursive digital filters with linear phase, *IEEE Trans. Circuit Theory*, **CT-19**, 189-194,

Wielądek, R., and T. Ernst (1977), Application of the method of least squares to determining impulse responses and transfer functions, *Publs. Inst. Geophys. Pol. Acad. Sc.* **G-1**, 110, 3-12

Received 13 November 2009
Accepted 26 January 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 973-987
DOI: 10.2478/s11600-010-0004-y

Gabriel Graph of Geomagnetic Sq Variations

Muthusamy SRIDHARAN[1] and Ayyathurai M.S. RAMASAMY[2]

[1]Magnetic Observatory, Indian Institute of Geomagnetism,
Pondicherry University Campus, Puducherry, India, e-mail: mopondy@bsnl.in

[2]Ramanujan School of Mathematics, Pondicherry University, Puducherry, India
e-mail: amsramasamy@gmail.com

Abstract

This paper describes the pattern of geomagnetic solar quiet day, *Sq*, variations recorded at the Indian geomagnetic observatories. The extent to which the high and low latitude ionosphere is interlinked electromagnetically during periods of quiet geomagnetic conditions is a point of debate. The concept of Gabriel graph is applied to derive the boundaries for the variations of horizontal, vertical, and declination components of the earth's magnetic field during geomagnetically quiet periods. Data of the six Indian geomagnetic observatories (Alibag, Hyderabad, Nagpur, Pondicherry, Visakapatnam, and Trivandrum) are considered for this analysis. This graph theoretical model is complementary to the classical data analysis techniques. Analytical method and the results of the analysis are presented in the paper.

Key words: geomagnetic *Sq* variation, Gabriel graph, Euclidean distance.

1. INTRODUCTION

The daily variation of the magnetic field at the Earth's surface during geomagnetically quiet periods, *Sq*, is known to be associated with the dynamo currents driven by winds and tidal motions in the E-region of the ionosphere, known as atmospheric dynamo. Winds in the upper atmosphere are blowing ionized gas across the geomagnetic field thereby creating an electric field and flow of currents where the magnetic field is fixed and independent of the flow pattern. The diurnal variation of the geomagnetic field observed at the

ground is a result of these currents. The daily pattern of the horizontal component of the field goes through a transition in shape at mid latitudes whereas the declination components of the field are reversed in direction for the northern and southern hemisphere. The solar quiet day variations, *Sq*, show up best on geomagnetic records during magnetically quiet conditions at low latitudes. Besides the atmospheric dynamo, other sources of electric field at the equatorial region contribute to the *Sq* variations on different components of the geomagnetic fields observed at the ground level. Daily range of the geomagnetic field is an important parameter measuring the magnitude of the diurnal variation. On days that are not excessively disturbed, the daily maximum should be predominantly controlled by the ionosphere dynamo *Sq* current intensity, so that the daily range and the maximum values may be expected to fluctuate in unison.

Visualization is an important technology for analyzing large amount of data. Graph retrieval as a means of recognizing 2D shapes from data-bases is an emerging technique to detect misleading projections of the data. Computed decompositions and presentation of time-series data by Gabriel graph offers substantive reductions in the expertise and time required to understand complex data set. It is complementary to classical principal component analysis, curvilinear component analysis, self organizing maps, projection techniques and cluster analysis. It is highly useful for comparing different decompositions when the volume of data is very high in a higher dimensional data space. Proximity drawings that obey the neighborhood constraints are known in the literature as Gabriel drawings. Their formulation is rigid and not parametrizable; there is only a single solution for a given input set. It can efficiently detect the boarder points in proximity graphs and can measure the margin between clusters. It preserves the original boundaries between data points of different classes very well. Gabriel graphs are used to get information on the topology of different classes in a potentially high dimensional data set and provides geometric interpretation along with mathematically derived ideas behind the geometry (Zhang and King 2002). The earlier techniques – principal component analyses (PCA) – were used to analyze the equatorial geomagnetic variations to test whether they can separate the electrojet and counter electrojet variations (Rajaram 1980), to determine the latitude of *Sq* focus (Rajaram 1983), day-to-day variability in the equatorial electrojet strength in the Indian region (Alex *et al*. 1998), etc. Projection techniques – multi dimensional scaling for geomagnetic storm time ranges (Sridharan and Ramasamy 2002) and geomagnetic *Sq* variations (Sridharan and Selvaraj 2009), and fuzzy clustering analysis (Sridharan *et al*. 2005) – are well applied to geomagnetic data. In all the above techniques the output is presented in a textual form, where the information that can be derived is limited. An alternative approach is to display the data in a graphical way so

that large amount of information can be perceived and this technique is superior to the earlier methods for the analysis of higher dimensional data (Aupetit 2003). For a background of this technique one may refer to Heckel and Hamann (1998). Basic concepts and the mathematics are explained in this paper.

2. REPRESENTATION OF CLUSTERS USING GRAPHS

The basic concepts are presented below.

Graph

Graph G is a set $G = \{V, E\}$ where V is a set of vertices and E is a set of edges (Gross 1999).

Edge

Edge is an arc or line joining two vertices. If $V = \{v_1, v_2, ..., v_n\}$ is a set of vertices and $E = \{ e_1, e_2, ..., e_m\}$ is a set of edges in a graph, then the ordered pair (v_i, v_j) gives a link between v_i and v_j.

Euclidean distance

Consider two objects, S_1 and S_2, with n entities as follows:

$$S_1 = \left(X_1, X_2, ..., X_n\right),$$

$$S_2 = \left(Y_1, Y_2, ..., Y_n\right).$$

The Euclidean distance between these two points is denoted by $d(S_1, S_2)$ or simply d, and is defined by the relation

$$d = \left[\left(X_1 - Y_1\right)^2 + \left(X_2 - Y_2\right)^2 + ... + \left(X_n - Y_n\right)^2\right]^{1/2}.$$

This distance can be used for relatively robust object recognition and location.

Euclidean distance matrix

The matrix obtained by considering all possible distances between any two points in a given set of data is called the Euclidean distance matrix.

Complete graph

Let G be a graph with n vertices and m edges. If $m = n\,(n - 1)/2$ then G is said to be a complete graph. If the set E possesses an order relation $l_1 \le l_2 \le l_3 \le ..., l_m$, then G is said to be an ordered graph (Foulds 1992).

Proximity graph

Using ordinal relation among Euclidean distance values, one can construct a graph called proximity graph. A graph connecting the vertices according to the magnitude of the distances is called a proximity graph. A proximity graph $G = (V, E)$ is an ordered graph where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of objects to be clustered and $E = \{e_1, e_2, \ldots, e_m\}$ is a set of object pairs termed as links. The links in E reflect the ordinal selection of the proximity data on the object pairs. A complete proximity graph distinctly represents the ordinal proximity over the ordered pairs.

3. GABRIEL GRAPH

In the proximity graph provided in Fig. 1, the circle described on the line joining any two points as a diameter does not contain the remaining point. Such a graph is called a Gabriel graph (Gross and Yellen 2003). Using the concept of Gabriel graph, a proximity graph is drawn for a given set of data and analysis can be performed through a Gabriel graph.

4. GABRIEL PAIR

Consider a given set of points. A Gabriel pair is the one where the circle drawn on this pair as a diameter does not contain any other point of the given set. A given graph G need not be Gabriel. However, it may contain certain Gabriel pairs.

Criterion for a Gabriel pair

In Figure 2, let A, B, C be three given points. Consider the circle with AB as the diameter. If the angle C is acute, then the point C is not in the interior of the circle and so AB is a Gabriel pair with respect to C. If the angle C is obtuse, then the point C is in the interior of the circle and so AB is not a Gabriel pair with respect to C, as shown in the figure.

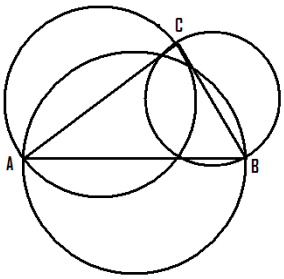


Fig. 1. Gabriel graph.

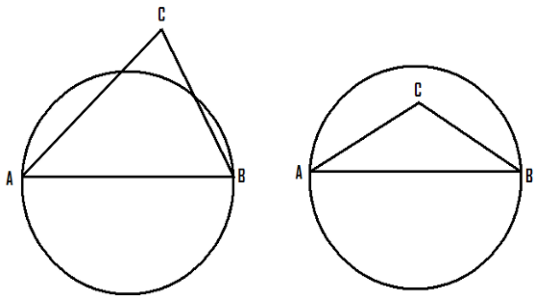


Fig. 2. Criterion for a Gabriel pair.

5. MAXIMAL GABRIEL SUBSYSTEM

Consider the Euclidean distances between any two points of this set. A circle is described with each pair of vertices as its diameter. Thus a system of circles is formed. From this system of circles a subsystem is derived which consists of the maximum number of circles such that no circle in the subsystem contains any other point in its interior. The subsystem so formed is called a maximal Gabriel subsystem.

6. DATA ANALYSIS

Data of six Indian geomagnetic observatories (Alibag, Hyderabad, Nagpur, Pondicherry, Visakapatnam, and Trivandrum) were analyzed; the respective station code is provided in Table 1. Daily range of the solar quiet day hourly mean values of declination D, horizontal H, and vertical Z components of the Earth's magnetic field for the common data for the period from January 1995 to September 1999 were taken up for the present study. With the view to minimizing the effect of random and stray fluctuations in unduly enhancing or diminishing the daily maximum and minimum values, the greatest hourly mean values and smallest hourly mean values were taken as the daily maximum and minimum values instead of usually considered instantaneous maximum and minimum values (Yacob and Sen 1974). The Trivandrum Observatory which was very nearer to dip equator was closed down in September 1999 due to technical reasons. During this period there were 285 international geomagnetically quiet days. The westerly declinations in minutes of arc of all the above six observatories have been converted to variations in nT for uniformity. The geographical and dipole coordinates of the observatories are provided in Table 1. Location maps are provided in Fig. 3. Diurnal patterns of variations of D, H and Z components are provided in Fig. 4. Euclidean distances for the D, H and Z variations of the six observatories are calculated and the Euclidean distance matrix is presented in Table 2.

Table 1

Location of observatories

Serial no.	Station	Station code	Geographic coordinate		Dipole coordinate
			latitude	longitude	
1	Alibag	ABG	18°37'N	72°52'E	10.0°N, 145.9°
2	Hyderabad	HYG	17°25'N	78°33'E	8.29°N, 151.3°
3	Nagpur	NGP	21°09'N	79° 05'E	11.96°N, 152.1°
4	Pondicherry	PON	11°55'N	79°55'E	2.7°N, 152.1°
5	Trivandrum	TRD	08°29'N	76°58'E	0.8°S, 148.5°
6	Visakhapatnam	VSK	17°41'N	83°19'E	08.2°N, 155.9°

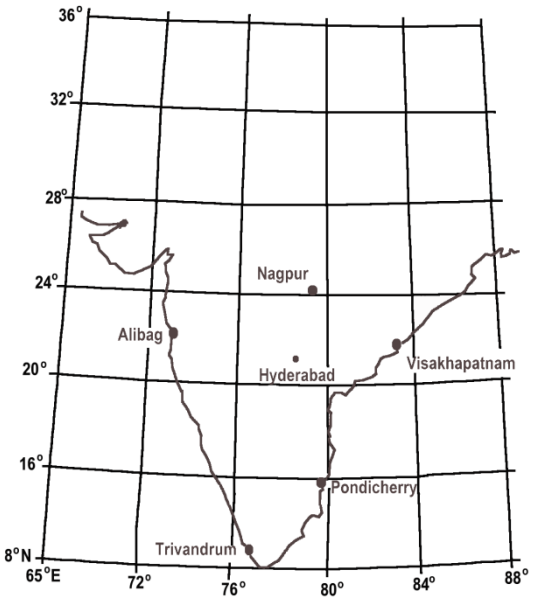


Fig. 3. Location maps of geomagnetic observatories.

Table 2

Euclidean distance matrix for *D*, *H*, and *Z* components

	HYB	NGP	PON	VSK	TRD
D component					
ABG	29.3	35.1	54.4	41.5	105.1
HYB		33.3	57.1	40.5	107.3
NGP			45.5	34.1	101.5
PON				64.7	63.7
VSK					124.4
H component					
ABG	41.5	92.3	105.1	37.6	124.2
HYB		61.7	73.3	23.4	153.1
NGP			63.1	74.8	200.9
PON				79.5	202.9
VSK					138.3
Z component					
ABG	79	74	58.4	94.8	85.9
HYB		24	76.7	32.1	97.0
NGP			78.1	36.4	95.9
PON				80.7	52.8
VSK					93.1

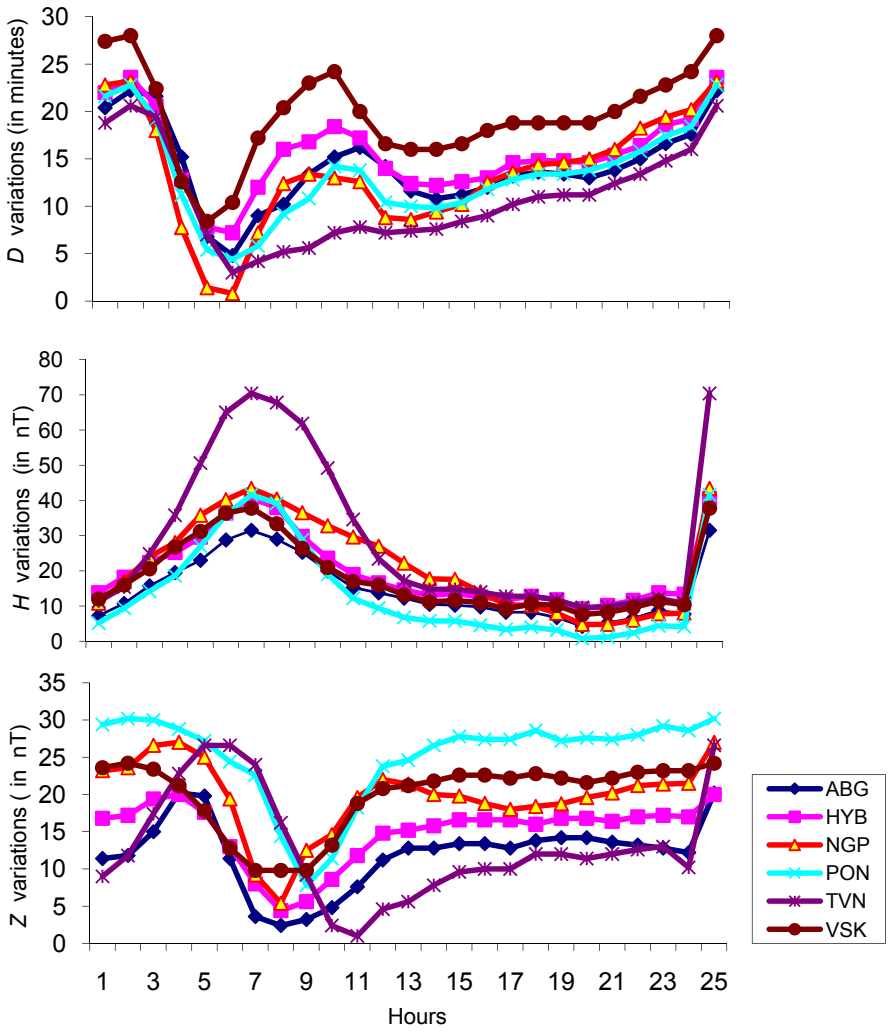


Fig. 4. Variations of D, H, and Z components. Mean diurnal variations of D, H, Z derived from hourly variations of internationally declared 5 continuous quiet days in January 1996. Colour version of this figure is available in electronic edition only.

7. APPLICATION OF GABRIEL GRAPH TO *Sq* VARIATIONS

Let V_1, V_2, V_3, V_4, V_5 and V_6 refer to Alibag, Hyderabad, Nagpur, Pondicherry, Visakhapatnam and Trivandrum, respectively. Let V denotes the vertex set $\{V_1, V_2, V_3, V_4, V_5, V_6\}$. Consider the Euclidean distances between each pair of vertices in G. A circle is described with each pair of vertices as its diameter. Thus a system of circles is formed. The system contains 20 circles.

Table 3

Gabriel triangulation for *D*, *H*, and *Z* components

Vertices of triangle			Sides of triangle			Angle [in degrees]		
A	B	C	BC(a)	CA(b)	AB(c)	*A*	*B*	*C*
D component								
ABG	HYB	NGP	33	35	29	61.4	67.9	50.7
ABG	HYB	PON	57	54	29	79.8	69.8	30.3
ABG	HYB	VSK	40	41	29	67.2	70.9	41.9
ABG	HYB	TRD	107	105	29	86.4	77.8	15.9
ABG	NGP	PON	46	54	35	56.3	83.8	40.0
ABG	NGP	VSK	34	41	35	52.1	73.6	54.4
ABG	NGP	TRD	101	105	35	74.5	86.0	19.5
ABG	PON	VSK	65	41	54	83.8	39.5	56.7
ABG	PON	TRD	64	105	54	29.7	125.3	25.0
ABG	VSK	TRD	124	105	41	107.7	53.7	18.5
HYB	NGP	PON	46	57	33	52.9	91.4	35.7
HYB	NGP	VSK	34	40	33	53.9	73.8	52.2
HYB	NGP	TRD	101	107	33	71.0	90.9	18.2
HYB	PON	VSK	65	40	57	81.3	38.2	60.6
HYB	PON	TRD	64	107	57	29.0	125.2	25.8
HYB	VSK	TRD	124	107	40	105.0	56.6	18.4
NGP	PON	VSK	65	34	46	107.9	30.0	42.0
NGP	PON	TRD	64	101	46	26.0	135.7	18.4
NGP	VSK	TRD	124	101	34	124.8	42.2	13.1
PON	VSK	TRD	124	64	65	150.0	14.9	15.2
H component								
ABG	HYB	NGP	62	92	41	32.8	125.8	21.4
ABG	HYB	PON	73	105	41	31.9	130.8	17.4
ABG	HYB	VSK	23	37	41	33.9	63.9	82.1
ABG	HYB	TRD	153	124	41	127.5	40.0	12.4
ABG	NGP	PON	63	105	92	36.6	82.9	60.6
ABG	NGP	VSK	74	37	92	51.4	23.1	105.5
ABG	NGP	TRD	201	124	92	135.7	25.6	18.7
ABG	PON	VSK	79	37	105	38.9	17.3	123.7
ABG	PON	TRD	203	124	105	124.2	30.6	25.2
ABG	VSK	TRD	138	124	37	104.0	60.7	15.3
HYB	NGP	PON	63	73	61	54.9	71.9	53.1
HYB	NGP	VSK	75	23	61	115.7	16.4	47.9
HYB	NGP	TRD	201	153	61	134.0	33.2	12.8
HYB	PON	VSK	79	23	73	96.8	16.9	66.2
HYB	PON	TRD	203	153	73	123.5	38.9	17.5
HYB	VSK	TRD	138	153	23	47.2	125.7	7.1
NGP	PON	VSK	79	75	63	69.8	62.0	48.1
NGP	PON	TRD	203	201	63	82.8	79.2	17.9
NGP	VSK	TRD	138	201	75	26.8	139.2	14.1
PON	VSK	TRD	138	203	79	28.4	135.8	15.9

to be continued

Table 3 (contd.)

Vertices of triangle			Sides of triangle			Angle [in degrees]		
A	B	C	BC(a)	CA(b)	AB(c)	*A*	*B*	*C*
Z component								
ABG	HYB	NGP	24	74	79	17.7	69.1	93.1
ABG	HYB	PON	77	58	79	65.8	44.0	70.1
ABG	HYB	VSK	32	95	79	18.6	109.8	51.6
ABG	HYB	TRD	97	86	79	71.9	57.3	50.7
ABG	NGP	PON	78	58	74	71.2	45.0	63.7
ABG	NGP	VSK	36	95	74	20.5	114.0	45.4
ABG	NGP	TRD	96	86	74	73.3	59.1	47.6
ABG	PON	VSK	81	95	58	57.8	84.4	37.7
ABG	PON	TRD	53	86	58	37.1	101.0	41.8
ABG	VSK	TRD	93	86	95	61.7	54.4	63.8
HYB	NGP	PON	78	77	24	84.5	77.7	17.8
HYB	NGP	VSK	36	32	24	79.5	60.0	40.4
HYB	NGP	TRD	96	97	24	80.2	85.5	14.3
HYB	PON	VSK	81	32	77	85.4	23.4	71.2
HYB	PON	TRD	53	97	77	32.8	95.3	51.8
HYB	VSK	TRD	93	97	32	73.5	87.2	19.3
NGP	PON	VSK	81	36	78	80.8	26.5	72.7
NGP	PON	TRD	53	96	78	33.4	92.1	54.4
NGP	VSK	TRD	93	96	36	74.6	83.2	22.1
PON	VSK	TRD	93	53	81	85.8	34.4	59.8

In triangle ABC, let a, b, c denote the length of the sides BC, CA and AB, respectively. Let the angles of the triangle ABC be denoted by A, B, C. Their measurement can be calculated by the following formula:

$$C = \cos^{-1}\left(\frac{a^2 + b^2 - c^2}{2ab}\right).$$

To derive the maximal Gabriel subsystem from the above system of 20 circles, the angles subtended by the pairs of observatories at another observatory are calculated for the D, H and Z components and the results are presented in Table 3.

By referring to the angles of the triangles provided in Table 3 one can verify whether a vertex is an interior point of a circle or not. In a triangulation, the pair of stations opposite to an obtuse angle cannot be Gabriel pairs. For example, from Table 3 (H component), ABG-HYB-NGP, the angle 125.8° is an obtuse angle. So the circle with diameter as ABG-NAG can not be formed as per the criterion for a Gabriel pair. Hence the circle was not drawn. Only 7 out of 20 circles provided in Fig. 5 fulfill the condition for the

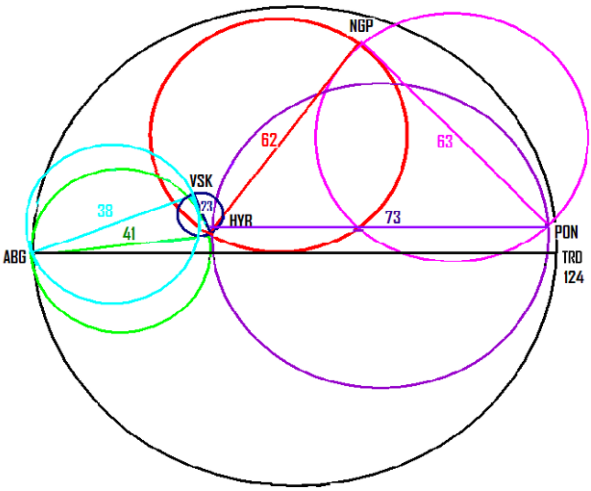


Fig. 5. Gabriel graph for *H* component – stage 1. Colour version of this figure is available in electronic edition only.

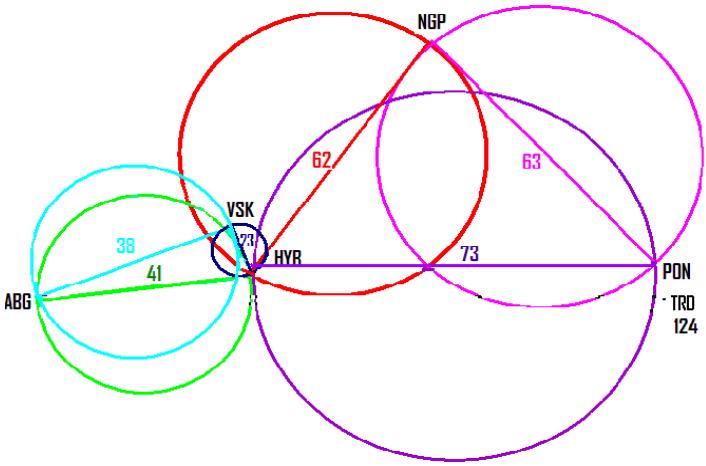


Fig. 6. Gabriel graph for *H* component – stage 2. Colour version of this figure is available in electronic edition only.

criterion of a Gabriel pair for the horizontal *H* component. From the system of 20 circles, a maximal Gabriel subsystem is obtained by the step-by-step procedure for each component. At each step, a circle with the largest diameter containing another vertex in its interior is removed from the system. This procedure is repeated until a subsystem with maximum number of circles is reached such that no circle in the subsystem contains another vertex in its interior.

8. MAXIMAL GABRIEL SYSTEM OF CIRCLES FOR THE HORIZONTAL *H* COMPONENT

The maximal Gabriel subsystem associated with the horizontal *H* component of the six observatories is presented in Fig. 5. The system consists of the following Gabriel pairs: (1) ABG-TRD, (2) ABG-VSK, (3) ABG-HYB, (4) HYB-VSK, (5) HYB-NGP, (6) HYB-PON, (7) NGP-PON.

In the system given by Fig. 5, the largest circle has ABG and TRD as the extremes of its diameter. This circle contains HYB, VSK, NGP and PON in its interior.

Alex *et al*. (1998) has estimated the effect of equatorial electrojet (EEJ) as the difference in the daily range of variation of *H* between ABG and TRD by applying principal component analysis. The equatorial enhancement of transient variation in the geomagnetic field is well known and has been attributed to increased conductivity of the *Sq* current layer in the equatorial region Trivandrum to that at Alibag which is outside the influence of the EEJ current. The abnormal variations of *H* at Trivandrum (Fig. 4, middle panel) are due to the effect of Petersen conductivity that supports the east-west dynamo currents augmented due to the inhibition of the vertical Hall current (Baker and Martyn 1953). As reported by Yacob (1975), the analysis of data from observatories in the vicinity of dip equator has shown that the ranges there and at a station outside the influence of the equatorial electrojet are not well established. The four stations, namely HYB, VSK, NGP, and PON, are away from the influence of day time equatorial electrojet current system. Therefore, the system is not a Gabriel system. This indicates that each of the four stations has an appreciable influence with the pair of stations ABG and TRD. Consequently, this circle is removed from the system. This results in a subsystem of six circles. This system at stage 2 is shown in Fig. 6.

In the system given by Fig. 6, the circle described with HYB and NGP contains VSK in its interior. The similarity and nearness in pattern of variations at ABG and VSK during geomagnetically quiet days for the *H* component has been revealed in the geomagnetic coastal studies (Sridharan *et al*. 2005). Again the three stations, HYB, NGP, and VSK, are non-equatorial stations and the latitudinal separations between NGP-VSK and NGP-HYB are the least, VSK has the influence of *Sq* over NGP and HYB. It follows that VSK has an appreciable interaction with the pair HYB and NGP. In view of this fact, it is seen that the system is not a Gabriel system. Therefore, this circle is eliminated from the system. The resulting system is shown in Fig. 7.

In the system given by Fig. 7, no circle contains any other vertex in its interior. Hence, the condition of optimality is reached. Now the maximal Gabriel subsystem has been attained. This system has 5 Gabriel pairs, namely,

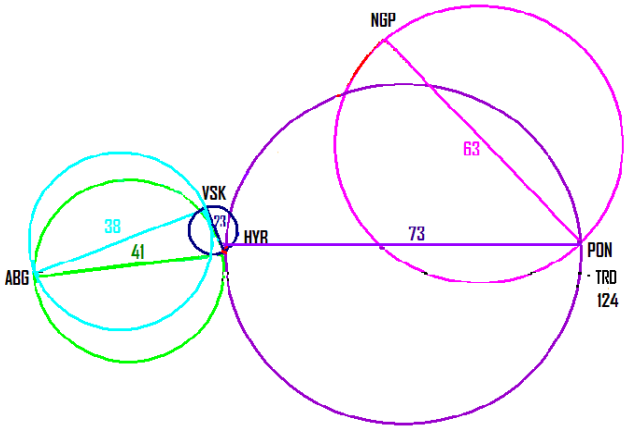


Fig. 7. Gabriel graph for *H* component – stage 3. Colour version of this figure is available in electronic edition only.

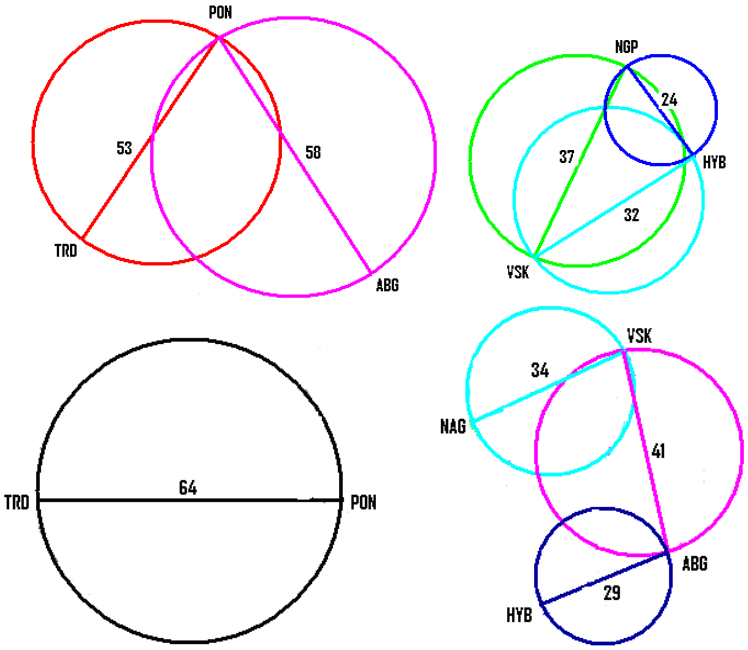


Fig. 8. Gabriel graph for *D* (upper) and *Z* (bottom) components. Colour version of this figure is available in electronic edition only.

HYB-VSK, HYB-ABG, ABG-VSK, HYB-PON, PON-NGP. Station Trivandrum is totally deviating from all the Gabriel pairs.

By applying the same procedure, Gabriel graph for the declination *D* and vertical *Z* components are obtained and they are provided in Fig. 8.

9. RESULTS AND DISCUSSION

Horizontal *H* component

The Gabriel graph in Fig. 7 shows the pattern of variations of the horizontal *H* component. Horizontal intensity variation on quiet days, *Sq*, is a composite effect of ionospheric dynamo current increasing the field during daytime and of eccentric and partial magnetospheric ring current decreasing the field in the evening-night sector. The relative importance of the two effects depends on the degree of disturbance. In several studies of the equatorial electrojet it has been shown that the difference field TRD-ABG is a true indicator of the equatorial electrojet strength, not augmented by planetary *Sq* not contaminated by magnetospheric disturbances (Bhardwaj and Rangarajan 1994). The equatorial station TRD is totally isolated, having no link with other stations. Multidimensional scaling for the analysis of geomagnetic storms (Sridharan and Ramasamy 2002) and *Sq* variations (Sridharan and Selvaraj 2009) shows that increase in latitudinal separation between a set of observatories corresponds to the increase in the spanning distances except at the equatorial stations. The Gabriel pair ABG-VSK circle is smaller than NGP-PON pair of circles. The similarity in variations of *Sq*(*H*) between ABG-VSK and NGP-PON in respect of latitudinal dependence is in agreement with the Gabriel graph provided in Fig. 7. The stations ABG, VSK and HYB form a Gabriel system connected to NGP-PON Gabriel pair. This graph provides the boundaries of the *Sq*(*H*) variations at the observatories described in the fuzzy clustering analysis to study the geomagnetic coastal effects (Sridharan *et al*. 2005). The Gabriel graph for the *H* component in Fig. 7 represents the geomagnetic *Sq*(*H*) variations at Indian observatories.

Declination *D* component

Information about the local magnetic declination is in common use throughout the world. Almost all topographic maps of small areas contain, along with the date of map production, a graphical indication of the magnetic declination for the area. (Campbell 1997). Some of the abnormal features of declination component based on the hourly mean values on international quiet days for different latitudes during different seasons at the Indian region have been studied by Alex *et al*. (1992). Ranges of *Sq*(*D*) at ABG and HYB are comparable and those from VSK are only marginally enhanced in all the three seasons (Srivastava *et al*. 2001). The pattern of variations at the Indian region for the declination component studied in geomagnetic coastal effects (Sridharan 2005) is again established in the present study. Gabriel graph in Fig. 8 (upper) shows the pattern of variations of declination component at the Indian region.

Vertical *Z* component

The diurnal variation at all the northern hemisphere observatories, south of *Sq* focal latitude, records a noon-time maximum in *H* and a depression of the field in *Z*. This is the case with all the observatories in the Indian region, irrespective of their location either on the east coast or west coast, or far from inland. The role of Deccan Traps in the induction problem between Alibag and Hyderabad and the anomalous variations of *Z* component is not fully understood. The TRD observatory which is also a coastal region is largely affected by anomalously induced currents. The Gabriel system described in Fig. 8 (bottom) represents the inter relationship of the variations in the vertical Z component at the Indian region.

10. CONCLUSIONS

The graphical representation of geomagnetic field variations at the Indian observatories has been explained in terms of the Gabriel graph. The Gabriel graph can be applied for studies on classification as a way to edit and condense large data sets by keeping only the relevant information to establish the interrelationship of one observatory with the others. This method can be applied for ground geomagnetic surveys to estimate the local anomalies, identification of boundaries for equatorial, auroral and sub-auroral regions, latitudinal movement of dip equator, storm time ranges, secular variations etc., to study the pattern of geomagnetic variations based on different mechanisms.

Acknowledgement. The authors are very grateful to the two referees for their comments and suggestions for the improvement of the paper. The first author is very grateful to Prof. Archana Bhattacharyya, the Director of the Indian Institute of Geomagnetism; Prof. S. Alex, Area Chairman, Observatory and Data Analysis, and Prof. B.M. Pathan, Head of Observatories, for their kind support in carrying out this research.

References

Alex, S., L. Jadhav, and D.R.K. Rao (1992), Complexities in the variation of declination component (D) of the geomagnetic field in the Indian region, *Mem. Geol. Soc. India* **24**, 263-274.

Alex, S., B.D. Kadam, and D.R.K. Rao (1998), Ionospheric current systems on days of low equatorial ΔH, *J. Atmos. Sol.-Terr. Phys.* **60**, 3, 371-379, DOI: 10.1016/S1364-6826(97)00075-8.

Aupetit, M. (2003), High-dimensional labeled data analysis with Gabriel graphs, *Proc. Intl. European Symp. on Artificial Neural Networks, Bruges, Belgium*, 21-26.

Baker, W.G., and D.F. Martyn (1953), Electric currents in the ionosphere. I. The conductivity, *Phil. Trans. Roy. Soc. Lond.* **A 246**, 913, 281-294, DOI: 10.1098/rsta.1953.0016.

Bhardwaj, S.K., and G.K. Rangarajan (1994), Solar control of the day-to-day variability in the equatorial geomagnetic field, *Indian J. Radio Space Phys*. **23**, 326-333.

Campbell, H. (ed.) (1997), *Introduction to Geomagnetic Fields*, Cambridge University Press, Cambridge, UK.

Foulds, L.R. (1992), *Graph Theory Applications*, Springer-Verlag, New York, 385 pp.

Gross, J.L., and J. Yellen (1999*), Graph Theory and its Applications*, CRC Press, Boca Raton-London-New York.

Gross, J.L., and J. Yellen (2003), *Handbook of Graph Theory*, CRC Press, New York.

Heckel, B., and B. Hamann (1998), Visualization of cluster hierarchies, *Proc. SPIE – The International Society for Optical Engineering* 3298, 162-171.

Rajaram, M. (1980), Method of natural orthogonal components applied to equatorial geomagnetic variations, *Ann. Geophys.* **36**, 599-603.

Rajaram, M. (1983), Determination of the latitude of Sq focus and its relation to the electrojet variations, *J. Atmos. Sol.-Terr. Phys*. **45**, 8/9, 573-578.

Sridharan, M., and A.M.S. Ramasamy (2002), Multidimensional scaling technique for analysis of magnetic storms at Indian observatories, *J. Earth System Science* **111**, 4, 459-465, DOI: 10.1007/BF02702058.

Sridharan, M., and S. Selvaraj (2009), Multi dimensional scaling of geomagnetic Sq(H) variations, *Indian J. Radio Space Phys.* **38**, 165-173.

Sridharan, M., N. Gururajan, and A.M.S. Ramasamy (2005), Fuzzy clustering analysis to study geomagnetic coastal effects, *Ann. Geophys.* **23**, 1157-1163.

Srivastava, B.J., H. Abbas, T. Rama Gopal, D.R.K. Rao, and B.M. Pathan (2001), Geomagnetic coast and other effects deduced from the new observatory at Visakhapatnam, India, *Geophys. J. Int.* **146**, 3, 827-832, DOI: 10.1046/ j.0956-540x.2001.01499.x.

Yacob, A. (1975), Latitudinal profile in India of Sq(H) range and of its prominent periodicities, *Pure Appl. Geophys***. 113**, 1, 601-609, DOI: 10.1007/ BF01592945.

Yacob, A., and A.K. Sen (1974), On quiet-day maximum minimum and range in H, *Pure Appl. Geophys*. **112**, 2, 464-471, DOI: 10.1007/BF00876154.

Zhang, W., and I. King (2002), A study of relationship between support vector machine and Gabriel graph, *Proc. IEEE Intl. Joint Conf. on Neural Networks* **1,** 239-244.

Received 22 September 2009
Received in revised form 16 December 2009
Accepted 21 December 2009

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 988-994
DOI: 10.2478/s11600-010-0030-9

Seismic Event of January 22, 2010 near Bełchatów, Poland

SHORT COMMUNICATION

Paweł WIEJACZ and Łukasz RUDZIŃSKI

Institute of Geophysics, Polish Academy of Sciences, Warszawa, Poland
e-mails: pwiejacz@igf.edu.pl, rudzin@igf.edu.pl

A b s t r a c t

The paper summarizes the findings about a seismic event on January 22, 2010, near Bełchatów, Poland, in terms of data, event location, magnitude and source parameters.

Key words: local seismicity, Bełchatów brown coal mine.

1. INTRODUCTION

On January 22, 2010, at 04:05 UTC (05:05 local time), a relatively strong seismic event has occurred in central Poland, causing widespread public concern. The event was quickly reported by several seismological centers with magnitude 4.3, some centers assigning it as 4.5 or even higher. The event was neither the first nor the biggest that has taken place in the region, however considerable time has passed since the previous significant event of comparable size in 1996.

The seismicity of the area near Bełchatów is considered to be associated with the open pit brown coal mine that was opened in 1977 (Gibowicz *et al*. 1982). The historical earthquake catalogue of Poland (Pagaczewski 1982) does not evidence any seismic events in the Bełchatów area while the updated version (Guterch and Lewandowska-Marciniak 2002) shows only

events that have taken place after the mine was opened. The event was the third largest that has ever happened in the area. The biggest event in the Bełchatów area took place on November 29, 1980 (Table 2).

2. INSTRUMENTAL DATA

The seismic event of January 22, 2010, was widely recorded by many seismic stations in Europe. In all, the event has been reported to the International Seismological Centre by 69 stations; this is however only the number of stations that have considered the event big enough to report it. Of the 10 seismic stations of the Polish national network, 2 stations have not recorded the event. Station PHL, relatively distant and located just by the sea, had the event lost in the high local noise. OJC observatory was down at the time due to lack of electricity. The loss of OJC is somewhat bearing since it was a sort of a reference station in studying past seismic events from Bełchatów area. Figure 1 shows the map of the stations in and around Poland that have reported the event. Figure 2 shows closeup of the epicentral area. Seismic recordings at distances up to several hundred kilometers are generally good. Sample seismograms from the broadband stations in Poland are shown in Fig. 3.

The seismic event was located at the Institute of Geophysics, Polish Academy of Sciences (IGF) with the Hyposat method (Schweitzer 2001) using data from 18 stations of the virtual network. The obtained result was 51.219N, 19.031E, 4.1 km depth and the origin time of 04:05:46.16, with horizontal location error of 2.9 km, 0.33 s error of origin time and a 9.2 km error on depth. In practice, this means that the depth is unresolved. The European-Mediterranean Seismological Centre (EMSC) has located the event using 64 stations – the same 17 of the 18 used by IGF (except CHZP), augmented by 47 distant stations (the most distant being YKA in Canada). EMSC has fixed the depth of the event at zero and developed the location at 51.250N, 19.060E and the origin time of 04:05:43.80, with possible 3.8 km error in latitude, 2.5 km in longitude and 0.12 s in origin time. The two locations are shown in Fig. 2. Of the two locations, the EMSC falls almost exactly at the western pit of the Bełchatów brown coal mine, named Szczerców, while the IGF location falls about 4 km away. This seems to favor the EMSC location, however this location has been resolved for fixed zero depth. The 4.1 km depth found by IGF, even though burdened with a big error, seems to be more likely in view of the large area over which the event was felt as far away as Łódź (about 40 km) and Częstochowa (50 km away).

The magnitude of the event may be calculated on several ways, the most popular are the local magnitude *Ml* and the spectral magnitude *Mw*. These may be calculated using data from different stations, yielding somewhat different results. Polish seismic stations yield from 4.2 to 4.6 for *Ml*, with

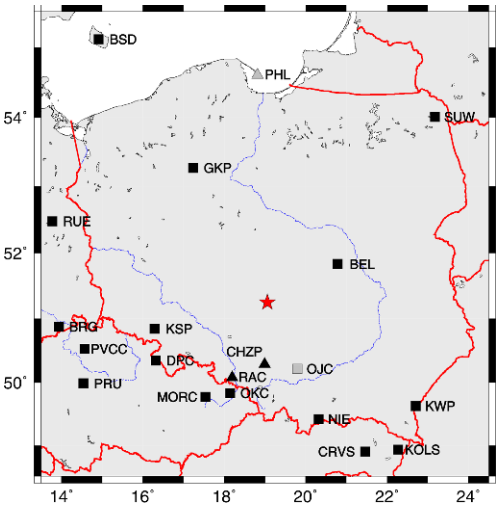


Fig. 1. Location of the January 22, 2010, seismic event at Bełchatów on the map of Poland, showing broadband (squares) and short period (triangles) seismic stations, that have recorded the event and were used for location. The two stations in Poland that were not used are shown in gray. Colour version of this figure is available in electronic edition only.

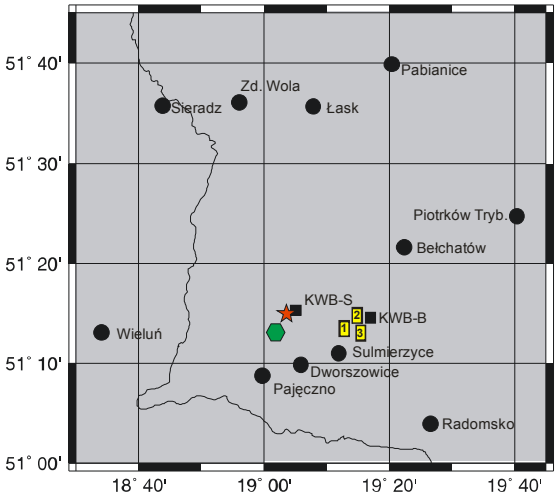


Fig. 2. Location of the January 22, 2010, seismic event at Bełchatów on local map, showing local municipalities. Location by the European-Mediterranean Seismological Centre is shown by star, the location by the Institute of Geophysics, Polish Academy of Sciences is denoted by hexagon (on scale of the map of Poland they are both at the same place). The two pits of the Bełchatów brown coal mine are shown as KWB-B and KWB-S. The locations of past $M > 4.0$ events in the area are denoted by rectangles with numbers, see Table 2. Colour version of this figure is available in electronic edition only.

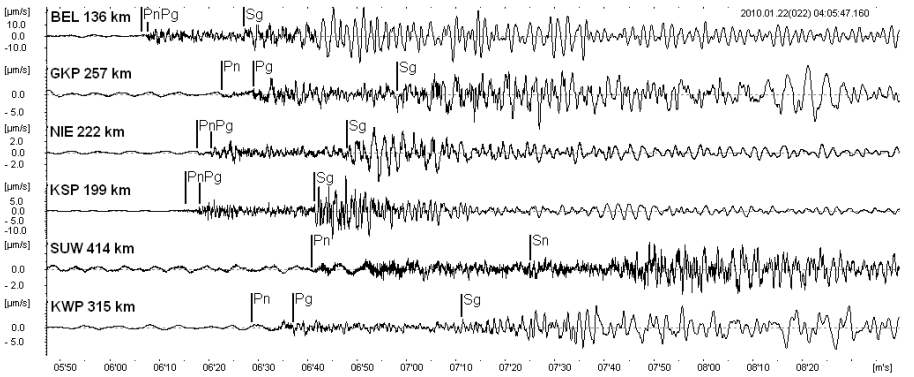


Fig. 3. Seismograms of the January 22, 2010, seismic event near Bełchatów. The figure shows unfiltered vertical component seismograms from BEL, GKP, NIE, KSP, SUW, and KWP.

average of 4.3, although some foreign seismological data centers and the EMSC have settled on $Ml = 4.5$ for the event. As for *Mw*, the results are somewhat smaller (Table 1), with an average of 3.95 (average of results from *S* wave, CHZP not considered in the average).

3. SOURCE PARAMETERS

Source parameters of the event have been calculated by the method described by Domański (2007), using the formalism described by Gibowicz and Kijko (1994). The method uses the approximation of the wave impulse proposed by Brune (1970), while the determination of corner frequency f_c and spectral level Ω_0 has been performed according to Andrews (1986) and Snoke (1987). The calculations of source radius r_0, seismic moment M_0, seismic energy E, total seismic energy ET, stress drop $\Delta\sigma$ and apparent stress σ_a were performed using both *P*- and *S*-wave records from the Polish seismic stations that have recorded the event. Results are given in Table 1.

The results obtained for CHZP seem erroneous, possibly because it is a short period station and the event was too distant for it – the spectral level and, consequently, corner frequency are erroneous because the frequencies of the signal pertaining to the spectral level are outside the frequency band range of the seismometer. Results obtained for *S* waves seem more reliable, as the *S*-wave amplitudes are generally greater, especially in case of KSP and NIE. It looks like these two stations might be near the direction of *P*-wave source mechanism nodal plane. Average values obtained from *S* waves are thus 431 m for source radius r_0, 1.1×10^{15} N·m for seismic moment M_0, seismic energy of 3.2×10^{11} J, stress drop $\Delta\sigma$ of 9.1 MPa, and apparent stress σ_a of 9.3 MPa.

Table 1

Source parameters of the January 22, 2010, Bełchatów seismic event, derived from records of the seismic stations in Poland

Station	*P* and *S*	f_c [Hz]	r_0 [m]	$M_0 \times 10^{14}$ [N·m]	M_W	E [J]	E_T [J]	$\Delta\sigma$ [MPa]	σ_a [MPa]
BEL	*P*	2.5	412	2.3	3.52	1.1×10^{11}	7.5×10^{11}	1.4	13.5
	S	2.5	411	12.7	4.02	6.4×10^{11}		8.0	13.2
KSP	*P*	7.0	147	0.3	2.93	0.3×10^{11}	3.4×10^{11}	4.4	25.9
	S	5.2	200	4.4	3.72	3.1×10^{11}		24.2	19.1
NIE	*P*	5.1	202	0.2	2.83	0.1×10^{11}	1.9×10^{11}	1.0	16.9
	S	2.4	432	6.0	3.80	1.8×10^{11}		3.3	7.9
GKP	*P*	1.7	606	5.4	3.79	1.0×10^{11}	4.8×10^{11}	1.0	4.9
	S	2.9	456	15.6	4.08	3.8×10^{11}		7.2	6.5
KWP	*P*	1.3	793	6.4	3.82	0.7×10^{11}	3.1×10^{11}	0.6	3.0
	S	3.8	342	9.2	3.93	2.4×10^{11}		10.0	7.1
SUW	*P*	1.1	938	7.4	3.86	0.5×10^{11}	2.3×10^{11}	0.4	2.0
	S	1.4	746	19.8	4.15	1.8×10^{11}		2.1	2.5
CHZP	*P*	7.2	143	0.5	3.08	0.8×10^{11}	4.5×10^{11}	8.0	42.3
	S	9.2	112	1.8	3.45	3.7×10^{11}		56.3	56.7

4. DICUSSION

The event of January 22, 2010, was one of the four $M > 4.0$ events that have happened in the Bełchatów area so far and is the third on list according to size. The big events are listed in Table 2.

Apart from these, there has been a $M = 3.8$ event on April 17, 2001, further 16 events in the 3.2-3.6 magnitude range recorded by seismic stations above 100 km distance and – possibly – a number of events small enough to be picked up only by the local seismic station of the Bełchatów brown coal mine.

Event No.1 has been subject to a thorough study (Gibowicz *et al.* 1982, Gibowicz 1984). Event No. 3 has been also subject to a study (Draber *et al.* 2002). Event no. 2 seemingly has not had such a study. Locations of all four events within their accuracy point to the mine openings. However, the event of January 22, 2010, points to the western, "Szczerców" pit of the mine while the earlier three events seem to be connected with the eastern, "Bełchatów" pit. The "Bełchatów" pit has been started in 1977 and the big event No. 1 has taken place three years later. The "Szczerców" pit has been started in 2007. This is by no means a scientific finding, only a supposition of a possible parallel. In any case, the starting of seismicity in vicinity of the newly opened "Szczerców" mine pit is a fact which is disturbing all by itself.

Table 2

Current list of seismic events above $M = 4.0$ from Bełchatów area

Event number	Date	Time [h m]	Magnitude
1	November 29, 1980	20:42	4.6
2	November 28, 1992	02:02	4.4
3	November 06, 1996	20:08	4.2
4 (this event)	January 22, 2010	04:05	4.3

As there are not known seismic events in the area from the time before the mining was started, it seems natural to assume that the seismicity is triggered. The big events, however, cannot be associated directly with the caved out pits. The source radii have been found to be of the order of hundreds of meters and there have not been such large-scale surface effects in the mine. One possible explanation could be a high level of tectonic stress in the area. Mining removes the overburden and changes the local stress pattern, thus enabling release of the tectonic stress.

References

Andrews, D.J. (1986), Objective determination of source parameters and similarity of earthquakes of different size. **In:** S. Das, J. Boatwright, and C.H. Scholz (eds.), *Earthquake Source Mechanics*, 259-267, Geophysical Monograph Series 37 (Maurice Ewing Volume 6), Am. Geophys. Union, Washington, D.C.

Brune, J.N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *Bull. Seism. Soc. Am*. **74**, 1615-1621.

Domański, B.M. (2007), Source parameters of the 2004 Kaliningrad earthquakes, *Acta Geophys.* **55**, 3, 267-287, DOI: 10.2478/s11600-007-0021-7.

Draber, D., B. Guterch, and H, Lewandowska-Marciniak (2002), Local earthquakes recorded by Polish seismological stations 1995-1997, *Publs. Inst. Geophys. Pol. Acad. Sci.* **B-24**, 3-13.

Gibowicz, S.J. (1984), The mechanism of large mining tremors in Poland. **In**: Proceedings of the 1st International Congress on "Rockbursts and Seismicity in Mines", Johannesburg 1982, N.C. Gay and E.H. Wainwright (eds.), SAIMM, Johannesburg, 17-28

Gibowicz, S.J., and A. Kijko (1994), *An Introduction to Mining Seismology*, Academic Press, San Diego, 399 pp.

Gibowicz, S.J., B. Guterch, H. Lewandowska-Marciniak, and L. Wysokiński (1982), Seismicity induced by surface mining: the Bełchatów, Poland, earthquake of 29 November 1980, *Acta Geophys. Pol*. **30**, 193-219.

Guterch, B., and H. Lewandowska-Marciniak (2002), Seismicity and seismic hazard in Poland, *Folia Quaternaria* **73**, 85-99.

Pagaczewski, J. (1982), Catalogue of earthquakes in Poland in 1000-1970 years, *Publs. Inst. Geophys. Pol. Acad. Sci*. **51**, 3-36.

Schweitzer, J. (2001), HYPOSAT – an enhanced routine to locate seismic events, *Pure Appl. Geophys*. **158**, 277-289, DOI: 10.1007/PL00001160.

Snoke, J.A. (1987), Stable determination of (Brune) stress drops, *Bull. Seism. Soc. Am*. **77**, 530-538.

Received 22 March 2010
Received in revised form 20 May 2010
Accepted 21 May 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 995-1020
DOI: 10.2478/s11600-010-0032-7

Microseismicity Induced During Fluid-Injection: A Case Study from the Geothermal Site at Groß Schönebeck, North German Basin

Grzegorz KWIATEK, Marco BOHNHOFF, Georg DRESEN, Ali SCHULZE, Thomas SCHULTE, Günter ZIMMERMANN, and Ernst HUENGES

Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, e-mails: kwiatek@gfz-potsdam.de (corresponding author), bohnhoff@gfz-potsdam.de, dre@gfz-potsdam.de, robert@gfz-potsdam.de, tschulte@gfz-potsdam.de, huenges@gfz-potsdam.de

A b s t r a c t

The technical feasibility of geothermal power production in a low enthalpy environment will be investigated in the geothermal site at Groß Schönebeck, North German Basin, where a borehole doublet was completed in 2007. In order to complete the Enhanced Geothermal System, three massive hydraulic stimulations were performed. A seismic network was deployed including a single 3-component downhole seismic sensor at only 500 m distance to the injection point. Injection rates reached up to 9 m^3/min and the maximum injection well-head pressure was as high as ~60 MPa. A total of 80 very small ($-1.8 < M_W < -1.0$) induced seismic events were detected. The hypocenters were determined for 29 events. The events show a strong spatial and temporal clustering and a maximum seismicity rate of 22 events per day. Spectral parameters were estimated from the downhole seismometer and related to those from other types of induced seismicity. The majority of events occurred towards the end of stimulation phases indicating a similar behavior as observed at similar treatments in crystalline environments but in our case at a smaller level of seismic activity and at lower magnitudes.

Key words: induced seismicity, spectral analysis, hydraulic stimulation, enhanced geothermal systems.

1. INTRODUCTION

Passive seismic monitoring (PSM) of microseismicity induced during hydraulic fracturing through massive fluid-injection is a well-established method to map the fracture growth, reservoir extent and permeability enhancements in hydrocarbon or geothermal reservoirs. Systematic fluid-injection was pioneered at the Rangely Oil Field, Colorado, confirming the hypothesis that earthquakes may be triggered by an increase of fluid pressure (Raleigh *et al*. 1972). Since the Rangely experiment, many of PSM campaigns have been pursued in the petroleum industry confirming the direct correlation between injection flow rate and pressure and rate of induced seismicity (*e.g.*, Kovach 1974, Albright and Pearson 1982, McGarr 1991, Zoback and Harjes 1997, Phillips *et al*. 1998, 2002). On the other hand, Tosha *et al*. (1998) studied the seismicity recorded at Kakkonda Geothermal Field in Japan and found no clear relationship between the well operations and microearthquake swarms during the injections. However, when production was shut down, swarm activity occurred in the nearby area. They attributed this to a change in the characteristics of the geothermal reservoir due to multiple injections. The changes in spatial and temporal distribution of seismic events were also reported by Nagano *et al*. (1994) who pointed out that the seismicity did not occur in previously stimulated zones (the so-called Kaiser effect, also reported in, *e.g.*, Baisch *et al.* 2002). The changes in seismic event rate during multiple injections were also reported by Simiyu (1999) who noted the time lag between consecutive stimulations in the same reservoir. All these observations led to the development of downhole seismic recording during reservoir stimulation as a key method to monitor the migration of the injected fluid at depth and thus develop quantitative models of the reservoirs.

In recent years, there has been a considerable improvement in seismic data acquisition: *e.g.*, the use of sensitive sensors with high sampling rate, installation of extensive seismic networks composed of both surface and borehole seismometers, located close to injection areas (*e.g.*, Evans *et al*. 2005, Bourouis and Bernard 2007). It enables recording of high-quality seismograms of very small events and allows for more sophisticated analysis schemes of small-scale brittle failure processes related to hydraulic stimulation. It improved primarily the precision of standard source parameter determination such as hypocenter location and event magnitude. It also allowed for calculation of other source parameters such as fault plane solutions, seismic moment tensors, moment magnitudes, source sizes as well as stress drop estimates.

Zoback and Harjes (1997) analyzed almost 400 microearthquakes ($M <$ 1.2) induced by injection of 200 m^3 of heavy brine at almost 9 km depth at

the KTB deep drill hole in Germany in 1994. They reported microearthquakes grouped spatially into clusters that were induced by very small pore pressure perturbations. The highly similar waveforms suggested successive movement of adjacent fault patches. Jost *et al*. (1998) used the same waveform dataset to calculate compound fault plane solutions and invert the waveforms for the relative source time function and calculate the rupture directivity using the empirical Green's function technique (Hartzell 1978, Mueller 1985). They also performed the relative moment tensor inversion. They found pure double-couple events supporting the results previously obtained by Zoback and Harjes (1997). They also found that the static stress drop is increasing with seismic moment. In a similar experiment at the same site six years later (Baisch *et al*. 2002, Bohnhoff *et al*. 2004), the seismic network detected 2799 events ($-1.2 < M_L < 1.1$). 237 events had a sufficient signal-to-noise ratio at the surface stations to determine reliable hypocenters. A strong spatiotemporal clustering was observed with clouds of seismicity moving away from the borehole with time. 125 focal mechanisms were calculated and used in the following stress tensor inversion. The study confirmed the strike-slip regime observed at KTB. Shapiro *et al*. (2006) showed that only positive pore pressure perturbations induced seismicity and they confirmed that fluid pressure diffusion is a dominant mechanism of seismicity triggered by fluid injections in KTB. Kümpel *et al*. (2006) analyzed the seismicity related to injection test performed at KTB in 2004-2005. They found that the seismicity was much weaker than in previous experiments. They attributed it to the lower pressure build-up and increased transmissivity of the fault system. Jahr *et al*. (2008) investigated the deformation caused by water injected in KTB. They found the maximum deformation of ~3 cm correlated with the region of the observed induced seismicity.

Extensive downhole seismic monitoring of induced seismicity was also performed at the European Deep Geothermal Energy Program site in Soultz-sous-Forêts, France. The sequence of fluid injection/production experiments was performed in 1993, 2000 and 2003-2005 (see Evans *et al*. 2005, Cuenot *et al*. 2006, Darnet *et al*. 2006, Charléty *et al*. 2007, for a comprehensive review). The maximum flow rate was reached in 2003 with 5.4 m^3/min when a total of 37 000 m^3 of water was injected. High seismicity rate was observed almost instantaneously with the start of injection, except when the target area had been stimulated during previous injections (Charléty *et al*. 2007). In general, occurrence rates of events were sensitive to changes in the flow rate and the microseismic activity decayed exponentially after shut-in of the boreholes. The stronger microearthquakes occurred just after the shut-in phase. Charléty *et al*. (2007) also observed a slight increase in average magnitude values during shut-in in comparison with injection periods and a slow decrease in the number of earthquakes from one year to another. Various

spatial structures of seismicity were observed, including planes with sizes much smaller than the source dimension of individual events, suggesting a multiple breakdown of the asperities. Hundreds of double-couple fault plane solutions were calculated (Cuenot *et al*. 2006, Charléty *et al*. 2007) indicating a normal faulting environment in the proximity of injection wells with strike-slip components at greater distances. Indications for non-double-couple seismic events were observed and considered to be related neither to numerical artifacts, nor to possible curvature of the faults (*i.e.*, complex source processes) nor to the opening and shearing mode of failure (Cuenot *et al*. 2006). They were attributed to either cooling around the injection or large increase in overpressure. No evidence for tensile faulting was observed (Charléty *et al*. 2007). No clear evidence was found for the breakdown in self-similarity of seismic events (*i.e.*, moment dependent stress drop).

In this study we analyze microseismicity induced during a massive (injection rate reaching 9 m^3/min) fluid injection experiment performed at the geothermal site in Groß Schönebeck, Germany, in August 2007. Seismic monitoring was achieved by the seismic network including a borehole geophone as close as 500 m to the injection point. We investigate the spatial and temporal distribution of the events to monitor the propagation of the injection. We determine the source characteristics such as seismic moments and source radii in order to assess the earthquakes' strength and extent of ruptured faults. We investigate whether there are any signatures of the type of event (shear/extensional failure) by comparing the energy radiated from *P* and *S* waves. We try to assess the damaging potential of recorded seismicity by comparing the seismic moment and radiated energy.

2. SITE DESCRIPTION

The Groß Schönebeck Research wells E GrSk 3/90 (here abbreviated as EG) and Gt GrSk 4/05 (GG) are located in the eastern part of the North German Basin, which is characterized by sedimentary deposits of several km thickness and absence of recent tectonic activity. Low enthalpy geothermal reservoir rocks (see Zimmermann *et al*. 2010, Zimmermann and Reinicke 2010, for details) are to be found here as siliciclastic sediments and volcanics of the lower Permian at an average depth of about 4000 m and at formation temperatures of up to 150°C. The existing well EG was used for repeated stimulation treatments to investigate scenarios of enhancing productivity of thermal fluid recovery from the underground (Zimmermann *et al*. 2009). Subsequently, the doublet has been completed with a second well GG with a total depth of 4400 m. In order to maximize the zone over which the subsurface heat exchanger can be created, this new well is inclined at 47° in the reservoir section. It was drilled in the direction of the minimum horizontal stress $\sigma_h = 288°$ (Moeck *et al*. 2007) for optimum hydraulic fracture align-

ment. Hence the orientation of any hydraulically induced fractures is expected to be 18°N, *i.e.*, in the direction of the maximum horizontal stress. While the deep well GG was used to perform fluid injection to locally enhance permeability and thus fluid circulation between both wells, a downhole seismic sensor was deployed in well EG at 3800 m depth. The sensor is located at only 500 m distance to the injection point, providing the opportunity to monitor the induced seismicity as low as $M_W \approx -2.0$.

3. INJECTION EXPERIMENT

Three hydraulic treatments were performed in well GG in summer 2007: One in the Lower Rotliegend volcanic section and two in the sandstone section of the Upper Rotliegend Dethlingen formation (see Fig. 1). The entire well was cased and cemented except for the lowermost 40 m, where an uncemented perforated casing was installed over an open hole section. The

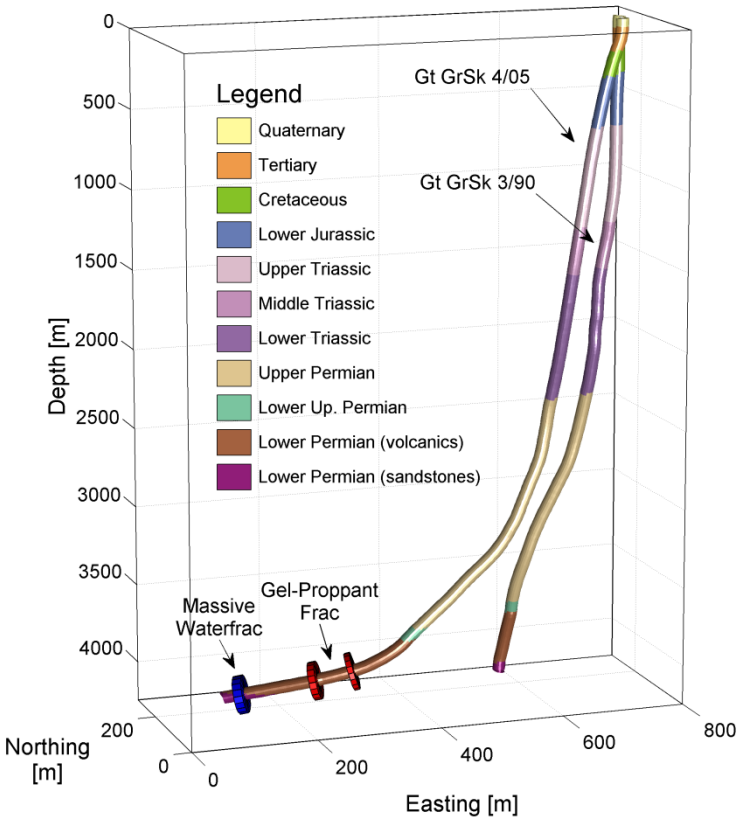


Fig. 1. Alignment of the well paths and the fracture treatments of the doublet system at the Groß Schönebeck drill site, after Zimmermann *et al*. (2008).

permeable volcanic rocks were stimulated through a massive cyclic water-frac treatment (Zimmermann *et al*. 2008). A cyclic injection procedure was chosen following technical constraints such as availability of fresh water and the expectation that a cyclic, high flow rate injection (up to 9 m^3/min) will enhance fracture propagation and performance compared to a constant and low (3 m^3/min) stimulation. The first and major injection was performed over a period of 6 days, between 9 and 14 August 2007. A total amount of 13 000 m³ of water was injected. The maximum injection well-head pressure reached 58.6 MPa. Two other treatments were carried out in the porous and permeable Upper Rotliegend sandstone formations over the intervals of 4122 to 4118 m and 4204 to 4208 m, respectively. 500 m³ of crosslinked gel were injected in each of the treatments (Zimmermann *et al*. 2010, Zimmermann and Reinicke 2010), at maximum well-head pressures of 49.5 MPa and 38.0 MPa, respectively. In the following analysis we primarily focus on the major injection in volcanic rocks.

4. SEISMIC NETWORK

The site conditions for a near-surface based seismic network are unfavorable due to the thick sedimentary formations and evaporite sequences at depth. Both results in low signal-to-noise ratios of seismic waves decreasing with distance from the source (see Weber *et al*. 2005). Nevertheless, we deployed a seismic network consisting of seven three-component seismometers including one downhole seismometer operated at 3800 m depth in EG well at only ~500 m distance to the injection point (see Fig. 2). The additional six instruments were installed both at the surface and in shallow (~60 m deep) boreholes, at about 3 km distance from the well head. The deep borehole sensor was of type GEOSPACE HS-1 with natural frequency of 15 Hz and sampling rate of 1000 Hz. The sensors located at the surface and in shallow boreholes were equipped with MARK SERCEL L4-3C seismometer (1 Hz) and SENSOR SM6-B geophones (4.5 Hz), respectively, and all sampled at a rate of 200 Hz. The acquisition system was continuously recording between 2 and 22 August 2007, framing the injections into the volcanic and sandstone layers. Data recovery rate of the network was 95%. We used regional seismic events from the Legnicko-Głogowski Copper District in Poland to calibrate the network (8 August 2007, M_L = 3.7, and 15 August 2007, M_L = 4.3, R = ~230 km according to SZGRF network). Those recordings and a calibration shot fired at 4000 m depth in the injection well (about ~500 m from the deep borehole sensor; see Fig. 3), close to the perforation point, were used to determine the orientation of the deep downhole sensor with an uncertainty of 15°. Unfortunately, the calibration shot could not be used to estimate the *P*- and *S*-wave velocities, as its origin time was not synchronized to the acquisition system clock due to technical problems.

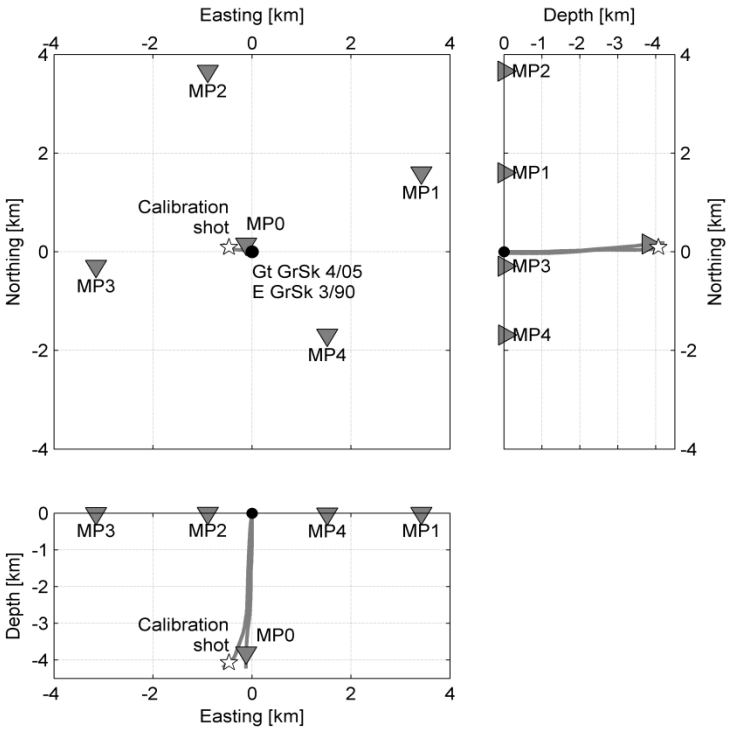


Fig. 2. Temporary seismic network installed at the Groß Schönebeck Geothermal Laboratory during the injection experiment. Seismometer locations are denoted with triangles. The location of the calibration shot is indicated by a star. MP1 and MP3 sites were composed of both surface and shallow borehole stations (60 m depth), MP2 and MP4 sites have only seismometers in shallow boreholes. The black dot is the well head.

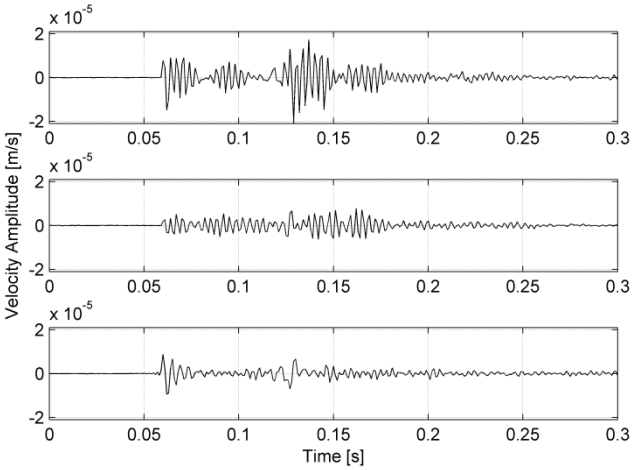


Fig. 3. Sample waveforms of recorded calibration shot (components Z, N, E from top to bottom) fired at ~500 m distance from the deep borehole seismometer.

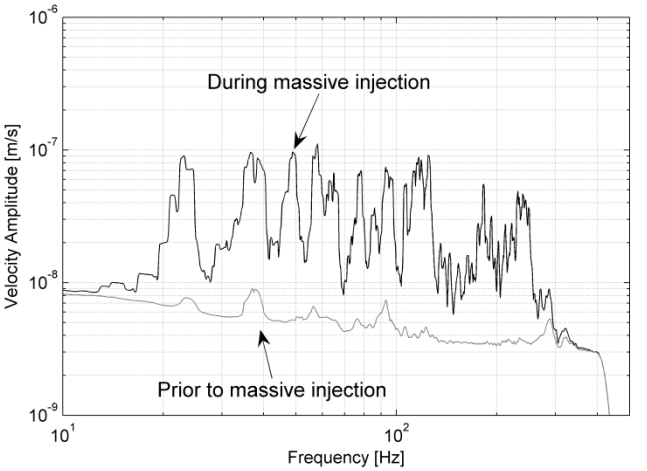


Fig. 4. Average noise levels recorded at the deep borehole seismometer during injection experiments. The black line shows the noise level at high injection rates, grayed – at very low injection rates.

Noise levels at the seismic sensors were sufficiently low prior to injection and during relatively low injection rates (not exceeding 2 m^3/min). However, during high injection rates, the deep downhole sensor was seriously contaminated by noise generated by the water pumps extending almost over the entire frequency range of the recorded seismic energy (Fig. 4). As a result, the recording and detection conditions were significantly limited during the periods of larger injection rates.

5. INDUCED SEISMICITY

A total number of only 80 microearthquakes were detected by the downhole seismometer using a LTA/STA (long-term average/short-term average) detection algorithm supported by an autoregressive AIC picker (Yokota *et al.* 1981, Leonard and Kennett 1999) to increase the detection accuracy. The *P* and *S* onsets were measured with uncertainty not exceeding 3 samples (*i.e.*, 3 ms). The estimated magnitudes of induced microseismic events were unexpectedly low and ranged $-1.8 < M_W < -1.0$. In consequence, because of source-receiver distances exceeding 5 km, high frequency content of the recorded seismic events (> 200 Hz) and attenuation of seismic signals, the events were only recorded by the three-component deep borehole sensor and were too small to be detected by the surface stations. Therefore, the entire analysis had to be focused on recordings from the one three-component deep borehole sensor. The analysis using one sensor limited the reliability of obtained source parameters; nevertheless, we decided to calculate the source parameters in order to get an insight in seismicity induced by the stimulation.

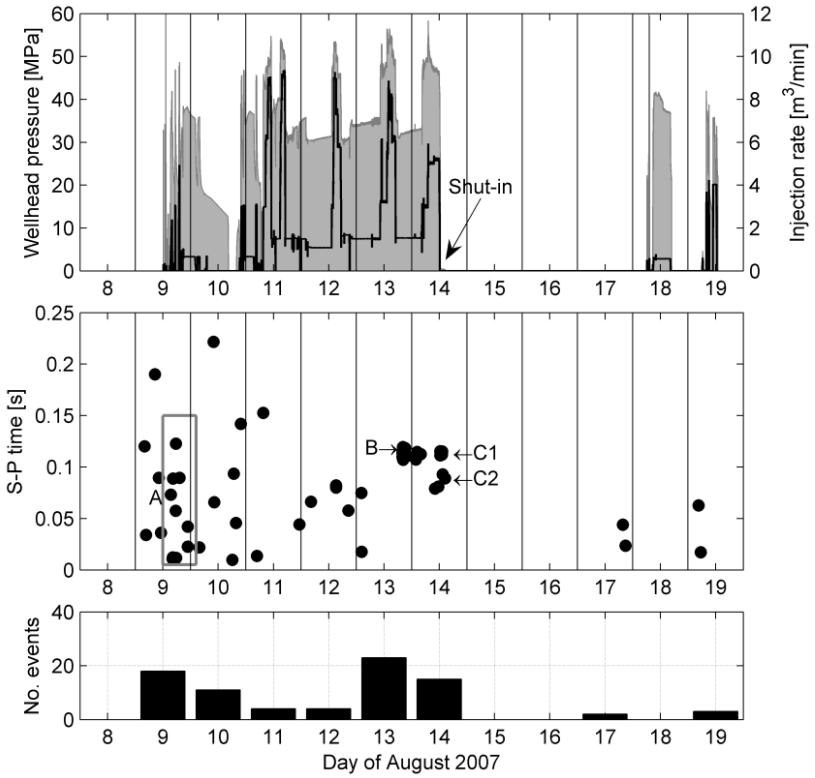


Fig. 5. Top panel: Well-head pressure (shaded area) and injection rate (black line) during the major (9-14 Aug) and the first minor (18-19 Aug) injection experiments. Middle panel: *S*-*P*-wave differential travel times of seismic events, detected at the deep borehole sensor. The arrows and rectangle mark the A, B, C1 and C2 clusters analyzed in this study (see Fig. 9 and Table 1). Bottom panel: Daily rate of detected seismic events.

Only a few seismic events were recorded before the first injection test that started on 9 August. During the stimulation test (9-10 Aug), the injection rate varied between 0 and 5 m^3/min, whereas the well-head pressure ranged between 0 and 50 MPa (Fig. 5). A relatively large number (20) of seismic events occurred almost instantaneously (~20 min) during the injection test (Fig. 5, sequence A). For some of the events, the *S*-*P* times could be determined and we were able to estimate at least the hypocentral distance. The majority of *S*-*P* times did not exceed 0.1 s. Assuming median *P*-wave (V_P = 4388±450 m/s) and *S*-wave velocities (V_S = 2575±250 m/s) as determined from 3 core sample measurements of sandstones (Trautwein and Huenges 2005), it corresponds to hypocentral distances smaller than 600 m. The hypocentral distances coincide with the distance between the deep sensor

location and injection area. Since we do not have information on the incident angle of the recorded seismicity, we speculate that the events from sequence A occurred in sandstones, in the direct proximity of the injection point.

No significant increase in seismic event rate was observed after the start of the longer injection phase on 11 August contradicting earlier findings during comparable experiments. This lack of seismicity may be only partially explained by the increase in pumping noise level (on 11 August the injection rate reached two times its peak value, *i.e.*, 9.38 m^3/min) as there were also a few episodes with much lower injection rate (and thus lower noise level) where almost no seismicity was detected. On 13 August, about two hours after the flow rate dropped from 5 m^3/min to less than 1 m^3/min (decrease in well-head pressure from 50 to 30 MPa), the most prominent seismic sequence occurred (sequence B in Fig. 5). The sequence lasted about 1.5 hours and consisted of more than 20 events with clear *S*-wave onsets whereas *P*-wave onsets were difficult to be identified in some cases. Interestingly, six events of this sequence formed three pairs in time with ~200, ~600, and ~700 ms difference between *P*-wave onsets, respectively. The *S*-*P* time appeared to be consistent for all events (110-120 ms, 683-751 m). The hypocenter determination performed with the three-component deep borehole sensor, described in the following section, confirms the common spatial origin of these events as shown in Fig. 6. The cross-correlation analysis of cluster B events also reveals similarity of waveforms (Fig. 7) with correlation coefficients ranging between 0.61 and 0.84.

The last sequence (C) occurred after the end of injection on 14 August. This sequence contains 12 events (9 located) with high signal-to-noise ratio for both *P* and *S* arrivals. The first events from sequence C occurred 20 minutes after the total shut-in of the well. Sequence C could be divided into two subgroups. The first subgroup (C1) consisted of six events with waveforms similar to that from cluster B (cross-correlation coefficients 0.65-0.95, 0.8-0.98 within C1 cluster). The almost identical *S*-*P* times of 113-115 ms, azimuths and incidence angles correspond to the same parameters estimated for cluster B (*cf*. Figs. 5 and 6). The second subgroup (C2) comprises at least three events (cross-correlation coefficients 0.61-0.71) with different waveforms and azimuths. *S*-*P* times range from 86 to 93 ms. Three examples of recorded seismic events from groups B, C1 and C2 are shown in Fig. 7. The detailed information on seismic parameters is listed in Table 1.

After shut-in, the number of recorded seismic events dropped to a few events per day (only two could be located) while the well remained unchanged for four days until the next experiment was performed on 18/19 August. The injection into a more porous and permeable sandstone formation induced only two very weak events, hardly recorded by the deep borehole sensor (additional two preceded the injection).

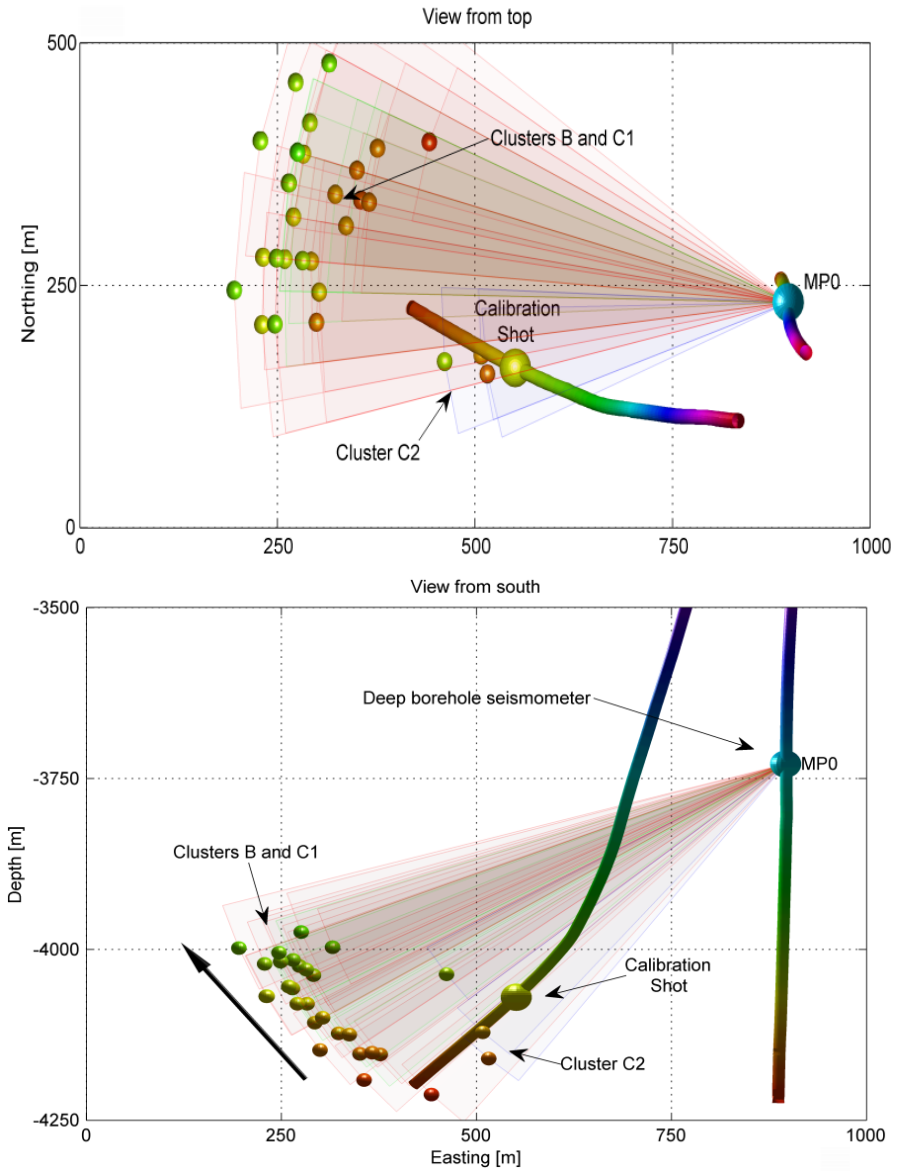


Fig. 6a. Hypocenter distribution of induced seismic events at the Groß Schönebeck geothermal site as determined from three-component recordings of the deep borehole seismometer. Color reflects the hypocentral depth of events plotted in accordance with the borehole trajectory for comparison. Top: Map view of hypocenters. Semi-transparent fans denote maximum horizontal error as discussed in the text. Bottom: Depth section seen from the south. Semi-transparent fans denote maximum vertical error as discussed in the text. The thick black arrow shows the migration of seismic events with time for cluster B and C1.

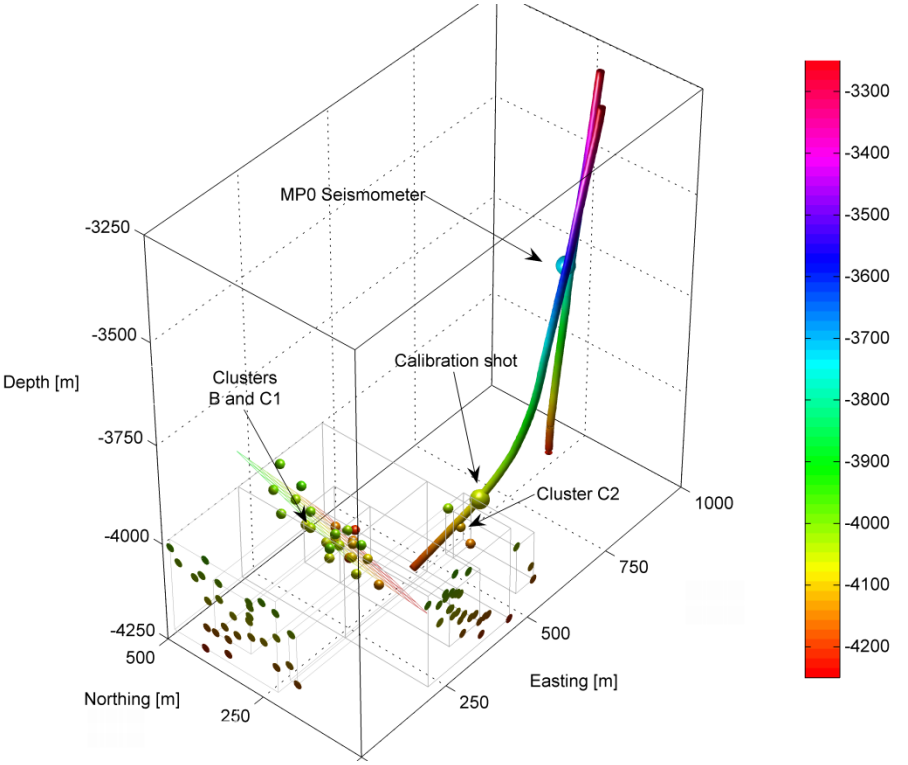


Fig. 6b. Perspective view of the hypocenter distribution of induced seismic events at the Groß Schönebeck seen from the southwest: The colored plane reflects the least squares fit to the locations of events from clusters B and C1. The events are also projected to the western and southern side of the bounding box.

6. HYPOCENTER DETERMINATION

In the first step, we estimated the spatial distribution of the recorded seismicity using the 3 component borehole sensor. The hypocentral distances ranged between 680 and 750 m and between 536 and 580 m for clusters B/C1 and C2, respectively. The picking accuracy for *P* and *S* onsets for the noisiest events analyzed does not exceed three samples (corresponding to the maximum difference of ±19 m in hypocentral distance). Its contribution to the overall error in hypocentral distance is thus smaller than the one resulting from uncertainties in the velocity model used, which are of the order of ±70 m (the estimations based on ±10% variation of *P* and *S* velocity of core samples measured). We assumed overall uncertainty (3σ) of distance estimation to be ±89 m. To determine the direction of incoming waves at the downhole sensor we applied a polarization analysis (*e.g.*, Plesĭnger *et al.* 1986). In addition, the results were tested by a manual rotation of the wave-

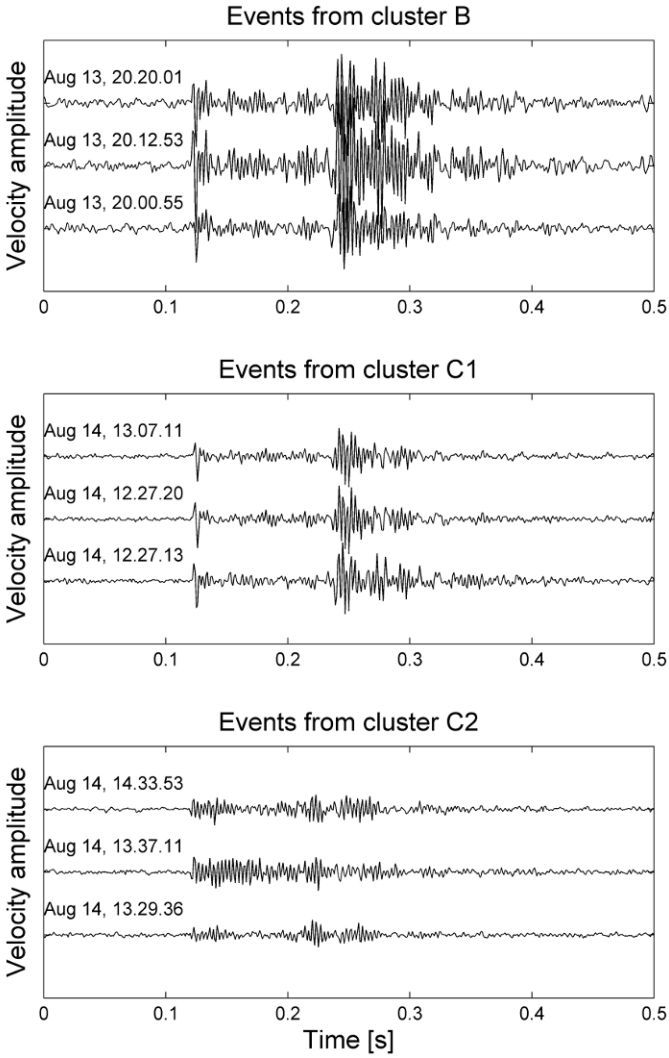


Fig. 7. Examples of waveforms recorded at the downhole sensor during the injection experiment. Each panel presents the vertical component for three recorded events from clusters B, C1 and C2, respectively, from top to bottom. Amplitudes are uniformly scaled.

forms towards the maximum polarization of the *P* wave. Assuming isotropic medium and straight seismic ray paths between seismic source and seismometer, a total number of 29 events could be located using *S*-*P* times as a measure of distance (Fig. 6). We compared results from automatic and manual polarization analysis, and concluded that the uncertainty for the azimuth

Table 1

Source parameters of analyzed seismic events on 13 August 2008 (sequence B) and 14 August 2008 (sequences C1 and C2) from clusters B, C1 and C2 (see text for a detailed explanation)

No.	Time	Distance [m]	Moment magnitude	Corner frequency [Hz]	Radiated energy [J]	Source radius [m]	Stress drop [MPa]	Apparent stress [MPa]
Sequence B								
1	$19^h56^m08.4^s$	700	–1.3	234	20.6	3.0±2.2	0.17	0.04
2	20 00 56	683	–1.3	214	59.2	2.5±1.8	0.32	0.10
3	20 01 37	719	–1.3	222	24.9	2.3±1.7	0.40	0.05
4	20 04 14	690	–1.2	263	89.2	2.9±1.8	0.26	0.12
5	20 06 40	714	–1.3	284	90.8	1.7±1.0	1.21	0.13
6	20 11 05.3	729	–1.4	292	39.6	1.6±1.0	0.87	0.09
7	20 11 49	690	–1.3	237	22.0	1.9±1.5	0.65	0.04
8	20 12 53	748	–1.0	244	449.2	2.9±2.0	0.60	0.26
9	20 14 26	686	–1.3	247	59.3	2.8±1.9	0.21	0.11
10	20 15 55	686	–1.1	250	187.5	2.9±1.9	0.40	0.16
11	20 20 01.9	710	–1.1	221	166.5	3.4±2.5	0.28	0.13
12	20 23 51	686	–1.3	235	32.0	3.1±2.2	0.18	0.05
13	20 29 55	706	–1.3	241	26.2	2.9±2.0	0.21	0.04
14	20 29 55	704	–1.2	219	49.1	3.2±2.5	0.21	0.06
15	20 42 35	723	–1.2	214	29.5	2.2±1.9	0.57	0.04
16	20 42 35	726	–1.3	230	30.2	2.1±1.6	0.57	0.05
17	20 55 48	751	–1.4	235	11.0	1.9±1.6	0.45	0.03
18	20 55 48	748	–1.4	327	102.9	1.8±0.8	0.66	0.24
19	21 07 59	748	–1.2	199	28.9	2.2±2.2	0.72	0.03
20	21 42 57	686	–1.3	279	50.6	1.8±1.1	0.79	0.10
Sequence C1								
21	$12^h27^m13^s$	717	–1.2	255	129.0	2.8±1.8	0.32	0.16
22	12 27 20	704	–1.2	213	37.6	3.3±2.6	0.20	0.04
23	12 38 11	704	–1.5	242	10.8	2.9±2.0	0.10	0.04
24	12 5 4 32	723	–1.4	212	10.5	3.4±2.7	0.09	0.03
25	13 07 11	704	–1.3	201	24.8	3.5±2.9	0.13	0.04
26	13 17 36	723	–1.3	202	26.8	3.5±2.9	0.12	0.04
Sequence C2								
27	$13^h29^m36^s$	555	–1.8	221	0.9	3.1±2.4	0.04	0.01
28	13 37 11	580	–1.6	239	1.8	2.8±2.1	0.08	0.01
29	14 33 53	536	–1.7	245	3.2	2.8±2.0	0.07	0.02

angle and angles of incidence is ±10º and ±5º, respectively. This corresponds to maximum horizontal and vertical errors of ±125 m and ±63 m for clusters B and C1, respectively. Accordingly, the location uncertainties for cluster C2 are ±97 and ±49 m, respectively. The spatial distribution of seismic events from cluster B and C1 and mutual similarity of their waveforms suggests that all events occurred on the same (fault) plane and that they might reflect earthquakes that repeatedly re-ruptured the same fault patch, as noted by, *e.g.*, Baisch and Harjes (2003). We fitted location coordinates, using least-squares technique to the plane surface. The strike and dip of the resulting plane (marked in Fig. 6) were found to be 17º and 52º, respectively, which is in good agreement with the stress regime in the studied area (see *e.g.*, Moeck *et al*. 2009a). We found events from clusters B and C1 moving outwards from the injection area with progressing time (*i.e.*, the shallower events occurred in the later stages of injection, see Fig. 6). This possibly reflects the migration of fluids through volcanic rocks and sandstones.

7. ASSESSMENT OF SOURCE PARAMETERS

We made an attempt to assess the source parameters using the borehole sensor to gain some insight into the seismic moment release caused by injection. We also wanted to achieve a rough estimation of the fault sizes' and the damaging potential of seismic events induced by the injection.

The preprocessing started with rotation of 3-component waveform data into a local ray coordinate system (radial, *SV* and *SH*) of the maximum *P*-wave polarization. Then, the ground velocity records were integrated to obtain the displacement waveforms. Selected parts of *P* and *S* phases (typically not longer than 80-90 ms) were tapered using a 10% von Hann's window. A Fast Fourier Transformation (FFT) was then applied to both the ground velocity and displacement waveforms. The resulting spectra were multiplied by a factor $\exp(\pi f R/V_C Q_C)$, where R is the distance and V_C is the *P*- or *S*-wave velocity, to correct for frequency-independent attenuation. We assumed $Q_P = 300$ and $Q_S = 150$ (these values seemed to produce the best fit between obtained spectra and Ω^{-2} source model). As the exact values of the attenuation factor are in fact not known in the direct vicinity of the borehole and injection area, we examined the influence of applied Q correction on calculated source parameters. We assumed that the quality factor may vary in a range of 200-400 and 100-200 for *P* and *S* waves, respectively, and then estimated the uncertainty of derived source parameters by the rule of uncertainty propagation. The attenuation correction had no visible effect on spectral level (therefore, *e.g.*, seismic moment) but it strongly affected the estimation of energy flux and corner frequency. An example of amplitude spectrum is shown in Fig. 8.

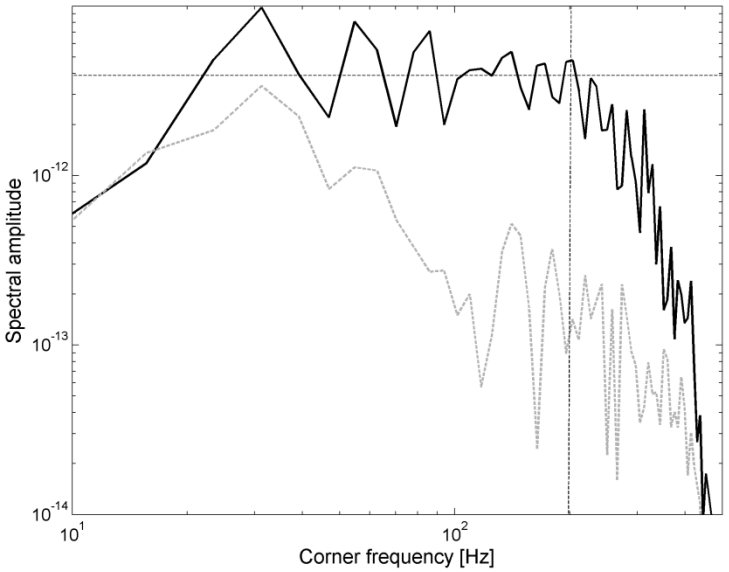


Fig. 8. Example of displacement spectrum calculated from radial component (sequence C1, event 25, see Table 1 for details). The noise level is shown with grayed line. The dashed lines mark spectral level and corner frequency selection.

The spectral parameters were estimated from horizontal components for events from clusters B and C, assuming Ω^{-2} Brune's model (Brune 1970) and the methodology developed by Andrews (1986). The bandwidth used in this study ranged from 50 to 300 Hz. The reliable estimation of source parameters from radial component was possible only for cluster C (no noise from water pumps) and selected events from cluster B. Source parameters were corrected for the limited frequency band considered, according to Di Bona and Rovelli (1988) and Ide and Beroza (2001). The seismic moment was calculated using:

$$M_0 = \frac{4\pi\rho V_C^3 R}{R_C}\Omega_C \ , \tag{1}$$

where $\rho = 2900$ kg/m^3 is the medium density, V_C is either *P*- or *S*-wave velocity, *R* is the source-receiver distance calculated from the *S*-*P* propagation time difference and Ω is the spectral level investigated manually on seismograms (for *S* waves, $\Omega_S = (\Omega_{SV}^2 + \Omega_{SH}^2)^{0.5}$). The additional term R_C is the correction accounting for radiation pattern. In this study, we used Monte-Carlo methodology provided in Boore and Boatwright (1984) to calculate R_C. We assumed that the most probable focal mechanism ranges from normal to strike slip faulting (rake from –90º to 0°), as suggested by Moeck *et al.*

(2009b). We assumed fault strike equal to 17°±15° and dip equal to 52°±5°. We found R_P and R_S ranged 0.35-0.49 and 0.86-0.90, respectively. We used median values of $R_P = 0.42 \pm 0.07$ and $R_S = 0.88 \pm 0.02$ to correct for radiation pattern. The corrections for a free surface effect and site were not applied due to the downhole location of the sensor (*e.g.*, Gibowicz and Kijko 1994). We assumed that the seismic moment may be biased by a factor of 2 due to manual picking. The moment magnitude was computed using the standard relationship (Hanks and Kanamori 1979): $M_W = 0.66 \log M_0 - 6.07$. The overall uncertainty of M_W estimation, affected primarily by the uncertainty of the picking process and velocity model, did not exceed ±0.30.

The radiated energy for either *P* or *S* wave is calculated following Boatwright and Fletcher (1984):

$$E_C = 4\pi\rho V_C \langle R_C \rangle^2 \left(\frac{R}{R_C} \right)^2 J_C \ , \tag{2}$$

where $J_C = 2\int |V(f)|^2 \mathrm{d}f$ is the measure of the seismic energy flux (Snoke 1987) and $V(f)$ is the ground velocity FFT spectrum of either *P* or *S* phase, corrected for attenuation. The average radiation coefficients equaled $\langle R_P \rangle = 0.52$ and $\langle R_S \rangle = 0.63$ (Boore and Boatwright 1984). The radiated energy was corrected for the limited frequency band according to Ide and Beroza (2001). The *P* and *S* energy could only be calculated for events from cluster C. The main source of uncertainties (98% of total uncertainty) originates from assumed variations in the quality factor and the energy flux can vary by a factor of 0.66-2.52 and 0.61-3.00 for *P* and *S* waves, respectively. However, persistent pumping noise and high-pass filtering of signals make the quantification of radiated energy difficult and the error may be even higher.

The corner frequency f_C was investigated manually on-screen and then averaged over *P* and *S* waves. Similarly to radiated energy, the estimations of corner frequency suffered from assumed variations in the quality factor. The calculated values may be biased by a factor of 0.87-1.36 and 0.84-1.43 for *P* and *S* wave, respectively.

The remaining source parameters assessed in this study were calculated as follows:

$$r = \frac{K_C V_S}{2\pi f_C} \ , \tag{4}$$

$$\sigma_a = \mu \frac{E}{M_0} \ , \tag{5}$$

$$\Delta\sigma = \frac{7}{16} \left(\frac{M_0}{r^3} \right) , \tag{6}$$

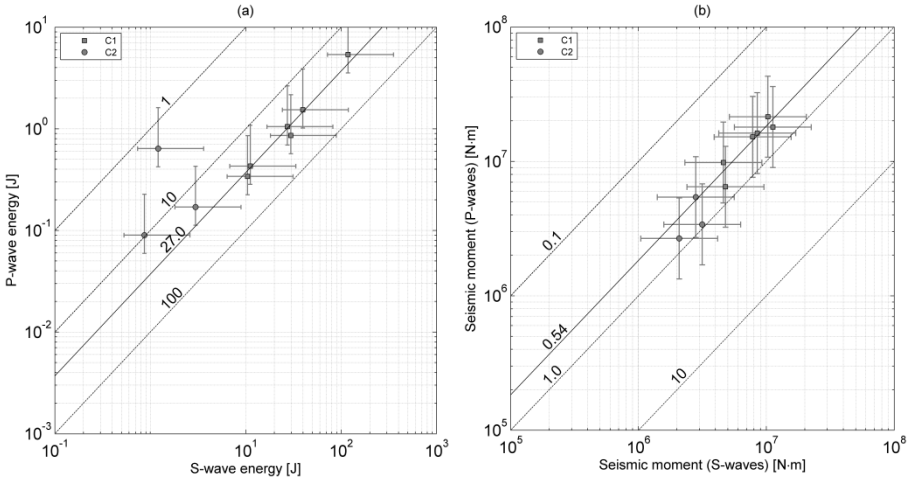


Fig. 9. The dependence between: (a) radiated energy, E_S and E_P , from *S* and *P* waves, and (b) seismic moment, M_0^S and M_0^P, calculated from *S* and *P* waves for events from clusters C1 and C2. The values of constant E_S/E_P and M_0^S/M_0^P equal to 0.1, 1.0, and 10 are also shown. The median value of corresponding ratios for events from cluster C1 is marked with solid line. The error bars shows the uncertainty of radiated energy (a) or seismic moment (b) estimation.

where r is the source radius, σ_a is the apparent stress and $\Delta\sigma$ is the stress drop (*e.g.*, Snoke 1987). K_C is the constant depending on the wave type and the source model used. We used the quasidynamic circular fault model of Madariaga (1976) to calculate the fault size and assumed that $K_P = 2.01$ and $K_S = 1.32$, appropriate values when focal mechanisms are not known. These correction coefficients provide a reasonable source size in case of small and induced seismic events (Gibowicz and Kijko 1994). For the apparent stress we assumed the rigidity coefficient μ equal to ρV_S^2. Source parameters for 29 events with $-1.8 < M_W < -1.0$ are presented in Table 1.

Figure 9a presents the relationship between radiated energy E_S and E_P calculated from *P* and *S* phases for clusters C1 and C2. A similar relationship for seismic moments is shown in Fig. 9b. For cluster C1, the E_S/E_P ratio ranges from 21.9 to 34.9 (mean value 27.0). The obtained values for cluster C1 are in agreement with other studies, where the energy radiated in *P* waves commonly tends to be a small fraction (0.05-0.3) of that radiated in *S* waves (Boatwright and Fletcher 1984). The values of the E_S/E_P ratio for cluster C2 are lower, in the range of 1.87-17.31. However, the magnitudes are small and the signal-to-noise ratio is rather poor and we cannot draw any conclusions. The ratio M_0^S/M_0^P ranges from 0.47 to 0.74 with the mean value equal to 0.54. The values of seismic moment depend mainly on the ac-

curacy in determination of the spectral level by manual picking and uncertainties in the wave velocity (*cf.* eq. 1). As the quality of determination of seismic moment is essentially same for *P* and *S* waves, the difference in seismic moment estimation reflects possibly the uncertainty of the velocity model used in our study and can be fully suppressed by decreasing the *S*-wave velocity by 150 m/s. This is much below the assumed uncertainty in *S*-wave velocity (±250 m/s). The change in velocity is supported by the calibration shot data.

The corner frequencies calculated from *S* waves (averaged over *SH* and *SV* components) and *P* waves for clusters C1 and C2 are consistent with f_S/f_P ratios equal to ~1.2. The difference may be an intrinsic source effect as well as the result of inappropriate correction for *P* or *S* attenuation. The *S*-wave corner frequency, ranging from 199 to 327 Hz, was subsequently used to

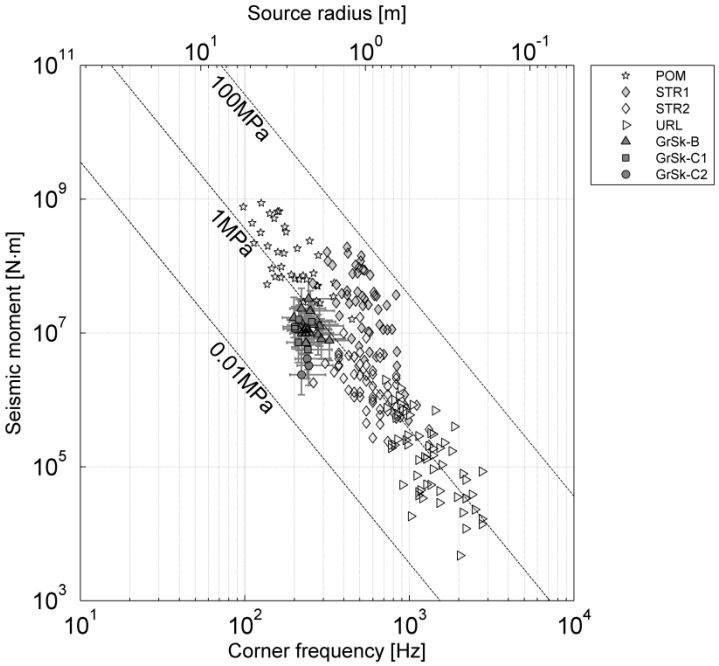


Fig. 10. The dependence between seismic moment and corner frequency and source radius for three analyzed clusters: B (black triangles), C1 (black squares) and C2 (black circles) in comparison with other studies in similar magnitude range: POM (start) – Pyhasalmi ore mine, Finland (Oye *et al.* 2005), URL (rotated triangles) – Underground Research Laboratory (Gibowicz *et al*. 1991), STR1, STR2 (light and dark grey rhombs) – Strathcona mine (Urbancic and Young 1993, Urbancic and Trifu 1996). The values of constant static stress drop are shown as lines. The error bars show uncertainty of seismic moment and corner frequency estimates (see text for details).

calculate the source radii. It is well known that calculation of the source radius is model-dependent (*e.g.*, Gibowicz and Kijko 1994). The resulting source radii ranged from 1.6 to 3.5 m and are affected by the uncertainty of the corner frequency estimation (*i.e.*, variations in attenuation). The uncertainties of the source radius estimation did not exceed 3 m (see Table 1 for details). The relation between seismic moment and source radius and corner frequency is shown in Fig. 10, together with uncertainties of seismic moment and corner frequency estimations and the lines of constant static stress drop ranging from 0.01 to 100 MPa. The values of static stress drop range from 0.04 to 1.21 MPa. They are, however, strongly affected by the uncertainty coming from source radius (78% of total error) and seismic moment (22%). As a result, static stress drop can vary by a factor of max. 6.7.

The calculated energy estimates generally provide an intermediate energy release values comparable to other studies within the same magnitude range, with values ranging from 0.9 to 449 J. Interestingly, cluster C2 presents the lowest values of radiated energy and this also applies to seismic moment or static stress drop estimates.

The uncertainty of apparent stress estimation is directly related to radiated energy (73% of total uncertainty originates from radiated energy). The other source of uncertainty is the seismic moment estimation (24% of the total error). As a result, the values of apparent stress are biased by a factor of 3.2. The values of apparent stress range from 0.01 to 0.26 MPa.

8. DISCUSSION AND CONCLUSIONS

The level of seismicity observed at the Groß Schönebeck geothermal site in the North German Basin is very low. The massive fluid injection into volcanic and sandstone formations at > 4 km depth in the sedimentary basin triggers induced microseismicity at magnitude levels $1.8 < M_W < 1.0$. The seismic signals do not penetrate thick sediments likely due to the high source-receiver distances, high frequency content of recorded waveforms and attenuation in the evaporite sequence above the reservoir. The seismicity rate is unexpectedly low considering the high injection rates of up to 9 m^3/min and maximum injection pressures of nearly 60 MPa at the well-head. This is in contrast to induced seismicity observed in granitic environment, *e.g.*, at Soultz-sous-Forêts (France) and KTB (Germany). There, thousands of seismic events were recorded (with significantly larger event magnitudes) during injection experiments at much lower injection rates and well-head pressure, but with similar source-receiver geometries. This cannot be explained by the substantially higher noise level during injection as no major increase of the seismicity rate was observed during injection phases with low surrounding noise. It is obvious that the local geologic and tectonic

settings are causing the difference in seismicity rate, and that comparable pore-pressure perturbations generated during fluid injection are not the only criterion to cause brittle failure during rock fracturing. The volcanic rocks at the Groß Schönebeck site might provide only limited shear strength accumulated on existing fractures compared to crystalline rock environments in Soultz or at the KTB. The fact that only two individual microseismic events were induced during the sandstone stimulation may be attributed to the possible different travel paths of fluids which were injected directly into sandstone deposits, characterized by even more limited shear strength compared to the volcanics (Moeck *et al*. 2009b).

Besides the low seismicity rate during the stimulation phase, the activity stopped after the stimulation episodes and shut-in of the well (clusters B and C). This confirms findings from other injection tests independent of the local rock formation, indicating that especially larger (and thus easier detectable) events seem to occur predominantly after injection phases related to a sudden drop in injection rate and well-head pressure. It is still not clear, however, why the events have not occurred after the earlier episodes of high-injection rate followed by the drop down in injection rate and well-head pressure, *e.g.*, on August 12 and 13, as the injection and pressure history were similar to the later episodes with recorded seismicity.

The majority of analyzed events form a plane structure striking ~17° and dipping ~52°ESE. Despite of uncertainties in the location accuracy described earlier, the orientation of the planar structure seems to be in very good agreement with the recent findings of Moeck *et al.* (2009a, b), where the highly shear-stressed minor fault system located in the vicinity of injection wells is striking and dipping in the similar direction. Also, the waveforms from clusters B and C1 represent a significant degree of similarity as determined from cross-correlation analysis, suggesting a similar focal mechanism. As a conclusion, the spatial distribution of events in comparison to their source radii not exceeding 3.5±3.0 m may indicate repeated slip on individual patches. Such a behavior was observed in a number of data sets of induced seismicity (*e.g.*, Baisch and Harjes 2003) as well as from natural microseismicity, *e.g.*, at the Parkfield segment of the San Andreas Fault in central California (Waldhauser *et al*. 2004). Moreover, we observe the migration of events outward from the injection area. Therefore, we think that almost all events considered in this paper (clusters B and C1) are attributed to the reactivation of fault plane(s) due to the increased pore pressure in the vicinity of the injection point.

We found seismic moment release with moment magnitudes extending over $-1.8 < M_W < -1.0$ and source radii not exceeding 3.5±3.0 m. Seismic moments of the events are relatively smaller than events analyzed in other studies in the same frequency range (Gibowicz *et al*. 1991, Urbancic and

Young 1993, Urbancic and Trifu 1996, Oye *et al*. 2005) (*cf.* Fig. 10). This cannot be explained by the uncertainties coming from velocity model, radiation pattern, attenuation or spectral level picking accuracy. The total effect of the aforementioned factors on the seismic moment is shown in Fig. 10 as error bars. The calculated values of seismic moment are situated at a lower part of the static stress drop interval with static stress drop smaller then 1.21 MPa, similar to data from Strathcona mine and Underground Research Laboratory in Canada. These values of static stress drop, even if we consider all the uncertainties, suggest a partial stress drop environment (*i.e.*, not a uniform and coherent release of stress over the fault plane) and complex source processes (*e.g.*, Gibowicz and Kijko 1994).

Unfortunately, both energy-related estimates (total radiated energy and apparent stress) are strongly affected by assumptions on attenuation factors. The estimations also suffer from theoretical assumption on radiated energy coming from the finiteness of the frequency band and persistent pumping noise (the effect has not been taken into account). However, one may compare the ratio between radiated energy from *P* and *S* waves which is better constrained. The high amount of energy released as *S* wave in comparison to *P*-wave energy suggests dominantly shear-type events. This again supports the idea that the faulting process favors the slip over the (pre-existing?) fault planes rather than the opening of the cracks due to the increase of the pore pressure in which case the amount of energy released as *P* waves should be larger.

The estimated apparent stress is comparable to other studies and appear to be dependent on seismic moment. In agreement with other studies (*e.g.*, McGarr 1999, Ide and Beroza 2001), the scaling relationship between seismic moment and apparent stress is not seen when all individually scaled data sets are compared over a broad magnitude range. It is possible that the dependence may be related either to the different source processes or differences in travel path/attenuation between events from clusters B/C1 and C2. However, the most likely scenario is that we were simply unable to efficiently record and analyze events with higher corner frequencies (thus higher energies) within the same seismic moment range (Ide and Beroza 2001), which would make the scaling between apparent stress and seismic moment less significant or even not visible. Finally, the average value of the Savage-Wood efficiency (the ratio between apparent stress and static stress drop) for analyzed events is ~0.25. This indicates friction-dominated events, according to nomenclature provided by Richardson and Jordan (2002). These events are similar to tectonic earthquakes and occur on pre-existing planar zones of weakness, such as bedding planes, dikes, and reactivated faults. This is again in agreement with the idea for the mechanism of induced microseismicity recorded in Groß Schönebeck.

References

Albright, J.N., and C.F. Pearson (1982), Acoustic emissions as a tool for hydraulic fracture location: Experience at the Fenton Hill hot dry rock site, *SPE J.* **22**, 4, 523-530.

Andrews, D.J. (1986), Objective determination of source parameters and similarity of earthquakes of different size. **In**: S. Das, J. Boatwright, and C.H. Scholz (eds.), *Earthquake Source Mechanics*, American Geophysical Union, Washington, DC, 259-268.

Baisch, S., and H.-P. Harjes (2003), A model for fluid-injection-induced seismicity at the KTB, Germany, *Geophys. J. Int*. **152**, 1, 160-170, DOI: 10.1046/j.1365-246X.2003.01837.x.

Baisch, S., M. Bohnhoff, L. Ceranna, Y. Tu, and H.-P. Harjes (2002), Probing the crust to 9-km depth: Fluid-injection experiments and induced seismicity at the KTB super deep drilling hole, Germany, *Bull. Seism. Soc. Am*. **92**, 6, 2369-2380, DOI: 10.1785/0120010236.

Boatwright, J., and J.B. Fletcher (1984), The partition of radiated energy between P and S waves, *Bull. Seism. Soc. Am*. **74**, 2, 361-376.

Bohnhoff, M., S. Baisch, and H.-P. Harjes (2004), Fault mechanisms of induced seismicity at the superdeep German Continental Deep Drilling Program (KTB) borehole and their relation to fault structure and stress field, *J. Geophys. Res*. **109**, B02309, DOI: 10.1029/2003JB002528.

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

Bourouis, S., and P. Bernard (2007), Evidence for coupled seismic and aseismic fault slip during water injection in the geothermal site of Soultz (France), and implications for seismogenic transients, *Geophys. J. Int*. **169**, 2, 723-732, DOI: 10.1111/j.1365-246X.2006.03325.x.

Brune, J.N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res*. **75**, 26, 4997-5009, DOI: 10.1029/JB075i026p04997; Correction, *J. Geophys. Res*. **76**, (1971), 20, 5002, DOI: 10.1029/JB076i020p05002.

Charléty, J., N. Cuenot, L. Dorbath, C. Dorbath, H. Haessler, and M. Frogneux (2007), Large earthquakes during hydraulic stimulations at the geothermal site of Soultz-sous-Forêts, *Int. J. Rock Mech. Min. Sci*. **44**, 8, 1091-1105, DOI: 10.1016/j.ijrmms.2007.06.003.

Cuenot, N., J. Charléty, L. Dorbath, and H. Haessler (2006), Faulting mechanisms and stress regime at the European HDR site of Soultz-sous-Forêts, France, *Geothermics* **35**, 5-6, 561-575, DOI: 10.1016/j.geothermics.2006.11.007.

Darnet, M., G. Marquis, and P. Sailhac (2006), Hydraulic stimulation of geothermal reservoirs: fluid flow, electric potential and microseismicity relationships, *Geophys. J. Int*. **166**, 1, 438-444, DOI: 10.1111/j.1365-246X.2006.03026.x.

Di Bona, M., and A. Rovelli (1988), Effects of the bandwidth limitation on stress drops estimated from integrals of the ground motion, *Bull. Seism. Soc. Am*. **78**, 5, 1818-1825.

Evans, K.F., H. Moriya, H. Niitsuma, R.H. Jones, W.S. Phillips, A. Genter, J. Sausse, R. Jung, and R. Baria (2005), Microseismicity and permeability enhancement of hydrogeologic structures during massive fluid injections into granite at 3 km depth at the Soultz HDR site, *Geophys. J. Int.* **160**, 388-412, DOI: 10.1111/j.1365-246X.2004.02474.x.

Gibowicz, S.J., and A. Kijko (1994), *An Introduction to Mining Seismology*, Academic Press, San Diego.

Gibowicz, S.J., R.P. Young, S. Talebi, and D.J. Rawlence (1991), Source parameters of seismic events at the Underground Research Laboratory in Manitoba, Canada: Scaling relations for events with moment magnitude smaller than –2, *Bull. Seism. Soc. Am.* **81**, 4, 1157-1182.

Hanks, T.C., and H. Kanamori (1979), A moment magnitude scale, *J. Geophys. Res*. **84**, B5, 2348-2350, DOI: 10.1029/JB084iB05p02348.

Hartzell, S.H. (1978), Earthquake aftershocks as Green's functions, *Geophys. Res. Lett*. **5**, 1, 1-4, DOI: 10.1029/GL005i001p00001.

Ide, S., and G.C. Beroza (2001), Does apparent stress vary with earthquake size? *Geophys. Res. Lett.* **28**, 17, 3349-3352, DOI: 10.1029/2001GL013106.

Jahr, T., G. Jentzsch, A. Gebauer, and T. Lau (2008), Deformation, seismicity, and fluids: Results of the 2004/2005 water injection experiment at the KTB/Germany, *J. Geophys. Res.* **113**, B11410, DOI: 10.1029/2008JB 005610.

Jost, M.L., T. Büsselberg, Ö. Jost, and H.-P. Harjes (1998), Source parameters of injection-induced microearthquakes at 9 km depth at the KTB DEEP Drilling site, Germany, *Bull. Seism. Soc. Am*. **88**, 3, 815-832.

Kovach, R.L. (1974), Source mechanisms for Wilmington Oil Field, California, subsidence earthquakes, *Bull. Seism. Soc. Am.* **64**, 3, 699-711.

Kümpel, H.-J., J. Erzinger, and S.A. Shapiro (2006), Two massive hydraulic tests completed in deep KTB pilot hole, *Scientific Drilling* **3**, 40-42, DOI: 10.2204/iodp.sd.3.05.2006.

Leonard, M., and B.L.N. Kennett (1999), Multi-component autoregressive techniques for the analysis of seismograms, *Phys. Earth Planet. Int.* **113**, 1-4, 247-263, DOI: 10.1016/S0031-9201(99)00054-0.

Madariaga, R. (1976), Dynamics of an expanding circular fault, *Bull. Seism. Soc. Am.* **66**, 3, 639-666.

McGarr, A. (1991), On a possible connection between three major earthquakes in California and oil production, *Bull. Seism. Soc. Am.* **81**, 3, 948-970.

McGarr, A. (1999), On relating apparent stress to the stress causing earthquake fault slip, *J. Geophys. Res.* **104**, B2, 3003-3011, DOI: 10.1029/1998JB900083.

Moeck, I., T. Backers, and H. Schandelmeier (2007), Assessment of mechanical wellbore assessment by numerical analysis of fracture growth, EAGE 69th Conference and Exhibition, 11-14 June 2007, Extended abstracts volume, D047, London, UK.

Moeck, I., H. Schandelmeier, and H.-G. Holl (2009a), The stress regime in a Rotliegend reservoir of the Northeast German Basin, *Int. J. Earth Sci*. **98**, 7, 1643-1654, DOI: 10.1007/s00531-008-0316-1.

Moeck, I., G. Kwiatek, and G. Zimmermann (2009b), Slip tendency analysis, fault reactivation potential and induced seismicity in a deep geothermal reservoir, *J. Struct. Geol*. **31**, 10, 1174-1182, DOI: 10.1016/j.jsg.2009.06.012.

Mueller, C.S. (1985), Source pulse enhancement by deconvolution of an empirical Green's function, *Geophys. Res. Lett*. **12**, 1, 33-36.

Nagano, K., H. Moriya, H. Asanuma, M. Sato, H. Niitsuma, and H. Kaieda (1994), Downhole AE measurement of hydraulic fracturing in Ogachi HDR model field, *J. Geotherm. Res. Soc*. *Japan* **16**, 85-108 (in Japanese).

Oye, V., H. Bungum, and M. Roth (2005), Source parameters and scaling relations for mining-related seismicity within the Pyhäsalmi ore mine, Finland, *Bull. Seism. Soc. Am.* **95**, 3, 1011-1026, DOI: 10.1785/0120040170.

Phillips, W.S., T.D. Fairbanks, J.T. Rutledge, and D.W. Anderson (1998), Induced microearthquake patterns and oil-producing fracture systems in the Austin chalk, *Tectonophysics* **289**, 1-3, 153-169, DOI: 10.1016/S0040-1951(97) 00313-2.

Phillips, W.S., J.T. Rutledge, L.S. House, and M.C. Fehler (2002), Induced microearthquake patterns in hydrocarbon and geothermal reservoirs: Six case studies, *Pure Appl. Geophys*. **159**, 1-3, 345-369.

Plesĭnger, A., M. Hellweg, and D. Seidl (1986), Interactive high-resolution polarization analysis of broadband seismograms, *J. Geophys*. **75**, 129-139.

Raleigh, C.B., J.H. Healy, and J.D. Bredehoeft (1972), Faulting and crustal stress at Rangely, Colorado. **In:** H.C. Heard et al. (eds.), *Flow and Fracture of Rocks*, *Geophysical Monograph Series* **16**, AGU, Washington, D.C., 275-284.

Richardson, E., and T.H. Jordan (2002), Seismicity in deep gold mines of South Africa: Implications for tectonic earthquakes, *Bull. Seism. Soc. Am.* **92**, 5, 1766-1782, DOI: 10.1785/0120000226.

Shapiro, S.A., J. Kummerow, C. Dinske, G. Asch, E. Rothert, J. Erzinger, H.-J. Kümpel, and R. Kind (2006), Fluid induced seismicity guided by a continental fault: Injection experiment of 2004/2005 at the German Deep Drilling Site (KTB), *Geophys. Res. Lett*. **33**, L01309, DOI: 10.1029/2005 GL024659.

Simiyu, S.M. (1999), Induced micro-seismicity during well discharge: OW-719, Olkaria, Kenya rift, *Geothermics* **28**, 6, 785-802, DOI: 10.1016/S0375-6505 (99)00043-7.

Snoke, J.A. (1987), Stable determination of (Brune) stress drop, *Bull. Seism. Soc. Am*. **77**, 2, 530-538.

Tosha, T., M. Sugihara, and Y. Nishi (1998), Revised hypocenter solutions for microearthquakes in the Kakkonda geothermal field, Japan, *Geothermics* **27**, 5-6, 553-571, DOI: 10.1016/S0375-6505(98)00033-9.

Trautwein, U., and E .Huenges (2005), Poroelastic behaviour of physical properties in Rotliegend sandstones under uniaxial strain, *Int. J. Rock Mech. Min. Sci*. **42**, 7-8, 924-932, DOI: 10.1016/j.ijrmms.2005.05.016.

Urbancic, T.I., and R.P. Young (1993), Space-time variations in source parameters of mining-induced seismic events with M < 0, *Bull. Seism. Soc. Am*. **83**, 2, 378-397.

Urbancic, T.I., and C.-I. Trifu (1996), Effects of rupture complexity and stress regime on scaling relations of induced microseismic events, *Pure Appl. Geophys*. **147**, 2, 319-343, DOI: 10.1007/BF00877486.

Waldhauser, F., W.L. Ellsworth, D.P. Schaff, and A. Cole (2004), Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity, *Geophys. Res. Lett.* **31**, L18608, DOI: 10.1029/2004GL020649.

Weber, M., F. Zetsche, T. Ryberg, A. Schulze, E. Spangenberg, and E. Huenges (2005), Seismic detection limits of small, man-made reflectors: A test at a geothermal site in Northern Germany, *Bull. Seismol. Soc. Am*. **95**, 4, 1567-1573, DOI: 10.1785/0120040124.

Yokota, T., S. Zhou, M. Mizoue, and I. Nakamura (1981), An automatic measurement of arrival time of seismic waves and its application to an on-line processing system, *Bull. Earthq. Res. Inst. Univ. Tokyo* **55**, 449-484.

Zimmermann, G., and A. Reinicke (2010), Hydraulic stimulation of a deep sandstone reservoir to develop an Enhanced Geothermal System: Laboratory and field experiments, *Geothermics* **39**, 1, 70-77, DOI: 10.1016/j.geothermics.2009.12.003.

Zimmermann, G., A. Reinicke, W. Brandt, G. Blöcher, H. Milsch, H-G. Holl, I. Moeck, T. Schulte, A. Saadat, and E. Huenges (2008), Results of stimulation treatments at the geothermal research wells in Groß Schönebeck/Germany, Proc. Thirty-Third Workshop on Geothermal Reservoir Engineering, January 28-30, Stanford University, Stanford, CA, SGP-TR-185.

Zimmermann, G., T. Tischner, B. Legarth, and E. Huenges (2009), Pressure-dependent production efficiency of an Enhanced Geothermal System (EGS): Stimulation results and implications for hydraulic fracture treatments, *Pure Appl. Geophys*. **166**, 5-7, 1089-1106, DOI: 10.1007/s00024-009-0482-5.

Zimmermann, G., I. Moeck, and G. Blöcher (2010), Cyclic waterfrac stimulation to develop an enhanced geothermal system (EGS) – Conceptual design and experimental results, *Geothermics* **39**, 1, 59-69, DOI: 10.1016/j.geothermics.2009.10.003.

Zoback, M.D., and H.-P. Harjes (1997), Injection-induced earthquakes and crustal stress at 9 km depth at the KTB deep drilling site, Germany, *J. Geophys. Res*. **102**, B8, 18477-18491, DOI: 10.1029/96JB02814.

Received 16 October 2009
Received in revised form 12 March 2010
Accepted 31 May 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1021-1039
DOI: 10.2478/s11600-010-0025-6

Application of Seismo-Acoustic Signals to the Study of Local Site Effects

Petru T. NEGRARU

Southern Methodist University, Department of Earth Sciences, Dallas, TX, USA
e-mail: pnegraru@smu.edu

Abstract

The Nevada Seismic Array (NVAR) is a small-aperture seismic array designed for monitoring an eventual nuclear test ban treaty. In spite of the 4 km aperture, large amplitude variations are recorded due to the complicated local geology. This study takes advantage of the collocated infrasound and seismic sensors to discuss the use of air-to-ground coupled waves to characterize the shallow geological structure existing beneath the array.

Complex transfer functions between the infrasound and the corresponding seismic signals are computed as the quotient of the cross-spectrum of the infrasound signal and the seismic signal and the power spectrum of the infrasound signal. Then the median of the transfer functions for the sites where shallow geologic information is available is compared to a theoretical model. In the theoretical approach, the signals are modeled as instantaneous pressure loads propagating at sound speed velocities (330 m/s). Both theory and observations are in agreement which suggests that inverting the transfer functions to determine elastic properties of the medium, and eventually computing site effects, is possible.

Key words: seismo-acoustic signals, infrasound area, seismic arrays.

1. INTRODUCTION

Ground motion peak acceleration maps are used by engineers to design earthquake resistant structures and to predict patterns of strong shaking from large earthquakes. The damage from the 1994 Northridge earthquake showed

irregular patterns across the Los Angeles area. Pockets of severely damaged buildings were located within 1 km of largely undamaged areas. This difference in damage was attributed to differences in the local site responses. Although the idea of high site amplification over short distances is widely accepted, a significant work on the topic has started only recently.

Extensive work has been done on quantifying the local site effects, with direct applications in engineering studies (seismic microzonation or prediction of earthquake strong motion). The common practice is to find site amplification factors based on the average shear wave velocity for the top 30 m of underlying materials (Borcherdt 1994, Boore and Joyner 1997). Engineering site investigations are usually limited to 30 m, and this has become a standard depth for classifying the site characteristics. Borcherdt (1994) and Martin and Dobry (1994) recommended that the design of structures be based on properties in the upper 30 m.

There are alternate methods to derive site amplifications which were widely studied and analyzed recently. The amplification factors are computed from coda waves (Mayeda *et al*. 1991, Phillips and Aki 1986, Su and Aki 1995) or from spectral ratio (Theodulidis *et al.* 1996, Castro *et al*. 1997, Chen and Atkinson 2002, Siddiqqi and Atkinson 2002), while comparisons among various techniques are discussed by Bonilla *et al.* (1997), Field and Jacob (1995), and Riepl *et al*. (1998).

The focus of this paper is on using seismic and infrasound signals to constrain near surface material properties. The first step is to detect the air-to-ground signals, and this will be performed by employing a detection algorithm based on the correlation coefficients between the infrasound and seismic sensors. From the signals thus detected, an empirical transfer function will be calculated and it will be matched with a theoretical pressure/velocity relationship.

2. DESCRIPTION OF THE NEVADA SEISMIC ARRAY SITE EFFECTS

A small-aperture array, the Nevada Seismic Array (NVAR), was installed in December 1998 by the Southern Methodist University (for details see NVAR Certification Manual 1999). The location of the array was chosen following the negotiations between a team of experts from the USA and Russia. It was decided that one of the three-component stations be located in the Black Butte mine, on the setting of the old MNV station for which historical records of nuclear explosions are available. Although the array fulfills its mission very well, to provide coverage for the Nevada Test Site (NTS) (Tibuleac and Herrin 2001), the final location, in a very complex tectonic and geological setting led to unusual site effects for a short-aperture array.

The array consists of ten short-period GS13 vertical seismometers with an aperture of 4 km (NV01 to NV10) disposed approximately in two circular alignments around a central station. An experimental infrasound array with the designation NVIAR (Nevada Infrasound Array) is collocated with the central short-period elements (NV01, NV02, NV03 and NV04). Three broadband three-component accelerometers (KS54000) are disposed around the array, forming an approximate equilateral triangle with the edge of 20 km (NV31 to NV33). With the exception of the NV31 and NV32, all the elements of the array are located in approximately 13.3 m deep boreholes. NV31 is located in the Black Butte mine, while NV32 is located in a deep borehole, at a depth of 60 m. The present study will focus mainly on the inner array (NV01 to NV04), which has collocated infrasound sensors.

The large amplitude variations at NVAR are evident without any kind of processing. Figure 1 shows the configuration of the short-period elements of the array and raw data from a 5.7 magnitude event that occurred in California. All the traces are at the same scale. By far the largest amplitude is recorded on channel NV04 (red trace), and in this particular case the variations reach factor three. These variations can be attributed to the shallow geology within the array. More than half of the stations are located in the Triassic limestone (sites NV01, NV02, NV03, NV05, NV06 and NV10). Sites NV04

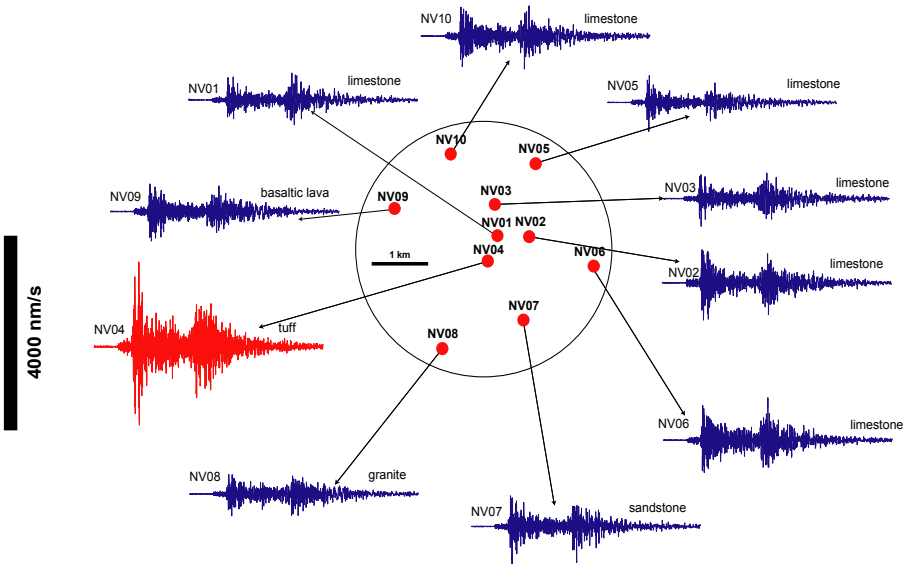


Fig. 1. Regional waveform recorded at NVAR, array configuration. Site NV32 is approximately 5 km south of site NV07. All the traces are shown at the same scale. The red trace is NV04 which shows the largest amplitude. In this case, the variations reach factor three. Colour version of this figure is available in electronic edition only.

and NV09 are located in volcanic rocks of Tertiary age (volcanic tuff, very variable in composition and texture and basaltic lava, respectively). Site NV07 is located in Jurassic sandstone (Dunlop Formation), while site NV08 is located in metamorphosed granite of Cretaceous age (for additional geological information see Stewart 1980).

Both NV04 and NV09 (located in younger geologic rocks) exhibit frequency dependent site amplifications, compared to any of the other sites of the array. At lower frequencies (up to 0.5 Hz), the signals are very similar and the amplitude variations are less than a factor of 2. However, above 1 Hz (up to 4 Hz) channel NV04 exhibits much larger amplitudes than the rest of the channels. In this frequency band, channels NV04 and NV01, only 500 m apart, have variations of up to a factor of 6. From 4 Hz to the antialias filter frequency (16 Hz), NV09 exhibits the largest amplitude, being sometimes more than one order of magnitude larger than the rest of the channels. No significant dependency on the source azimuth and distance has been found for these effects, which suggest the lack of any organized dipping structure beneath the array.

3. CHARACTERISTICS OF AIR-TO-GROUND COUPLED WAVES

Seismic detections of infrasound waves were reported several times (McDonald and Goforth 1969, Goforth and McDonald 1970, Sorrells *et al.* 1971a, b, 2002, Donn *et al.* 1971, Anglin and Haddon 1987, Langston 2004), even if sometimes no infrasound sensors were employed. It is uncommon to take advantage of the collocated infrasound and seismic arrays, though there are advantages in using seismic data in combination with infrasound, particularly in source characterization and discrimination at lower magnitudes, 2.5-4 m_b (Sorrells *et al.* 1997, Herrin *et al.* 1998, 1999, Chilo *et al.* 2006, Evers *et al.* 2007, Chilo 2008). In addition, for short-aperture seismic and infrasound arrays, the estimated azimuths of infrasound arrivals are thought to be more precise than the estimates based only on seismic arrivals, because the seismic wavefront can be highly distorted by the geologic structure. At the Texas Seismic Array (TXAR) located near Lajitas, Texas, the azimuths of seismic arrivals show differences up to ±20° in azimuth (Tibuleac and Herrin 1997), due to a dipping Moho beneath the array. In contrast, for the Columbia shuttle explosion which generated an infrasound signal, the estimated azimuths of the all the arrivals were all within ±1.5° (McKenna and Herrin 2006).

Early papers reported a correlation between changes in magnitude of the seismic and atmospheric pressure fields (Capon 1969a, b). Later it was determined that local fluctuations in the atmospheric pressure field can contribute significantly to the vertical seismic background noise, especially if the

seismic sensor is located in low-density, slow seismic velocity layers (Sorrells 1971).

McDonald and Goforth (1969) and Goforth and McDonald (1970) discuss the seismic effects of the sonic boom. In their study they employed collocated seismic and infrasound sensors. They found that the ground particle velocity is linearly related to the maximum pressure of the boom and to the geological properties of the ground. They also reported the presence of a precursor before the arrival of the maximum velocity which was always present when the velocity of the compressional wave in the ground was greater than the velocity of the aircraft (which gives the velocity of the *N* wave). Their interpretation suggested that the maximum particle velocity is associated with the passage of the *N* wave, and the smaller precursors are produced by air-coupled Rayleigh waves. Recently the air-coupled Rayleigh waves precursors were successful modeled as "bow waves" originating at the advancing tip of the shock front trace (Sorrells *et al.* 2002).

A different approach was taken by Langston (2004) who computed synthetic seismograms to model the seismic ground motion caused by a bolide shock wave and concluded that the ground motions are generally confined to a near-surface layer of approximately 10 m, and they behave like high frequencies "leaky" and "locked" mode Rayleigh waves. He also suggested that the ground motions are not sensitive to the deeper earth structure, at least not in the Mississippi embayment.

4. DATA AND DATA PROCESSING

The seismo-acoustic signals (Fig. 2) were detected by estimating the correlation coefficients at zero lag between seismic and acoustic envelope functions for a sliding 10-second window. Because a shorter window yields higher correlation coefficients for seismic and infrasound noise, a tradeoff between window length and correlation threshold was achieved. The time length of most signals, which in general do not exceed a few seconds, was also taken into account. Empirically, a 10-second window and a detection threshold at 0.5 gave good results (Fig. 3). Typically the noise for such a window length has correlation coefficients that range from –0.3 to 0.3 while signals with good signal-to-noise ratio (SNR) have correlation coefficients above 0.5. The maximum correlation coefficient was 0.99, for a signal recorded on channel NV04. A detection was declared if there were correlation coefficients above 0.5 for at least three stations. With three stations, the backazimuth and the phase velocity can be computed. Although this method works well for NVAR, it is not expected to work in the same way for all stations in the western US, but a similar approach can be used to make a detector at each particular array (or site). Most of the signals detected at NVAR have

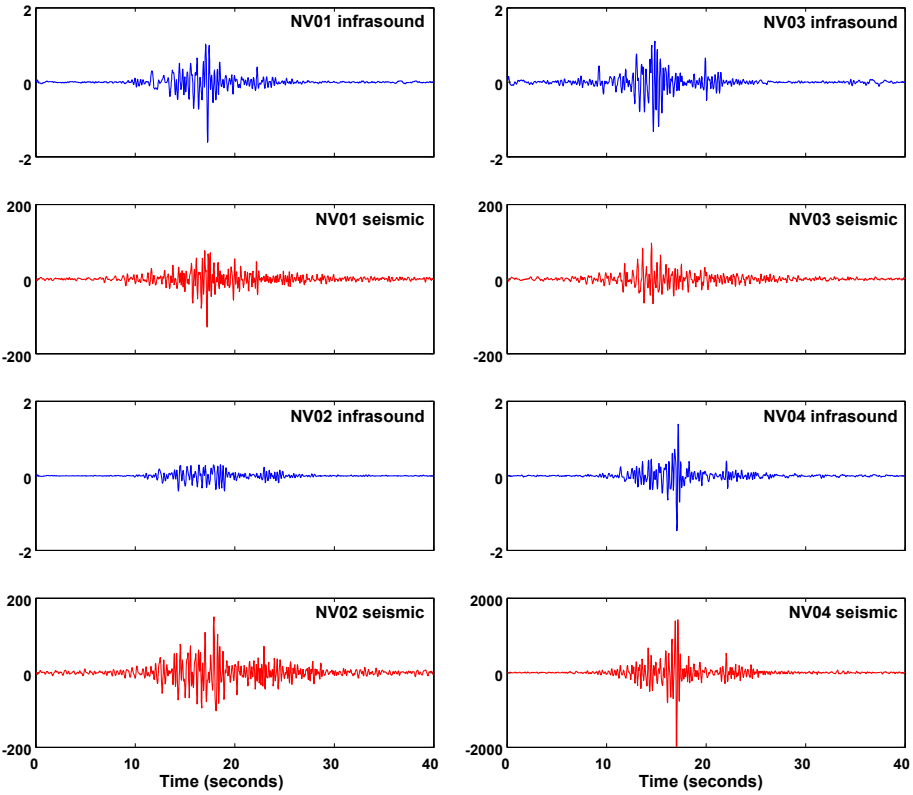


Fig. 2. Examples of air-to-ground coupled waves. The unit for seismic channels is nm/s, for infrasound is Pa. Shown here are the sites NV01 to NV04. Colour version of this figure is available in electronic edition only.

a dominant frequency around 2 Hz, well above the microbarom peak. Although a significant number of air-to-ground coupled signals were detected, due to low signal-to-noise ratio only 109 signals were selected for processing.

An important observation from Fig. 2 is that the infrasound arrivals exhibit significant phase and amplitude differences, in spite of the fact that the stations are only 500 m apart. Therefore, in the absence of infrasound sensors it is very difficult to extract quantitative information about the geologic structure. Infrasound signal decorrelation with distance is under investigation, but it appears that the infrasound signals are decorrelating faster than seismic signals.

For the signals detected and selected for processing, complex transfer functions were computed for the collocated infrasound and seismic channels. Transfer functions are the link between pressure signals and ground motions;

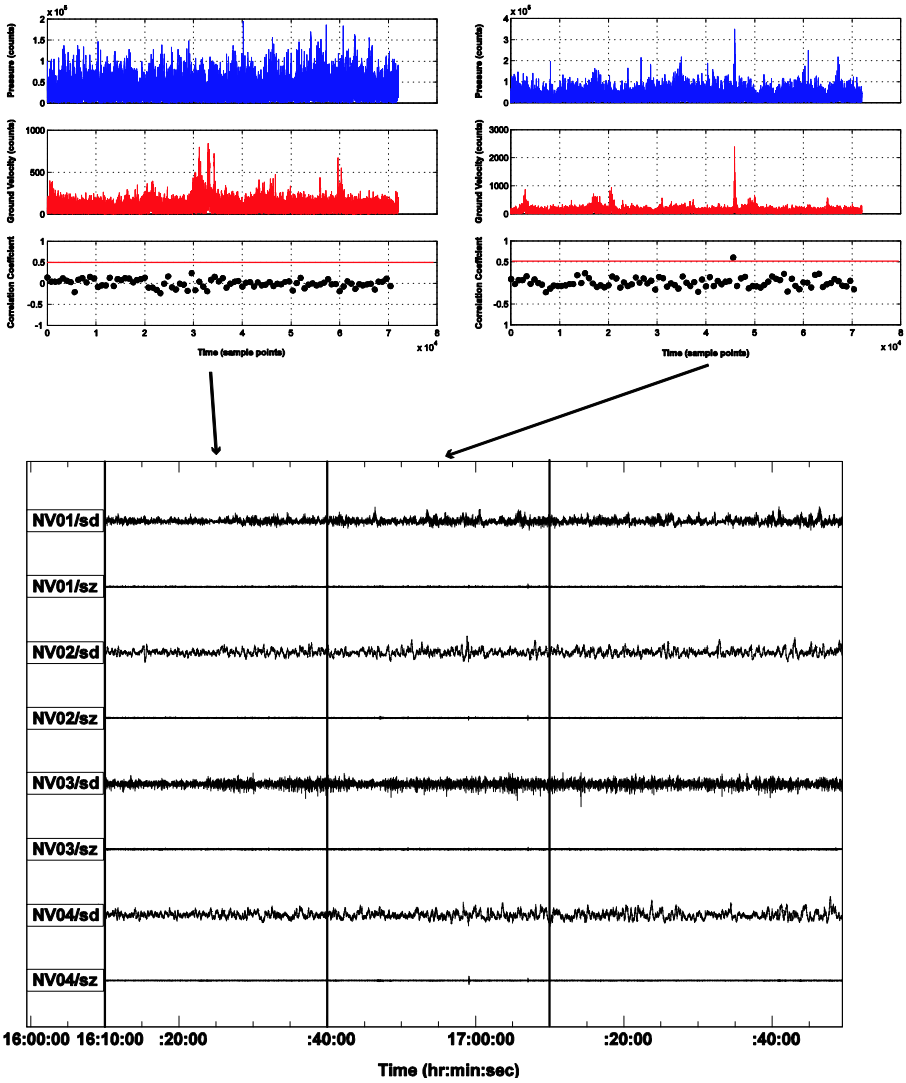


Fig. 3. Detection of seismo-acoustic signals. Colour version of this figure is available in electronic edition only.

they tell us how much ground motion is generated at a particular frequency for a given pressure signal. The whole air-to-ground coupling process could be regarded as filtering with a linear filter, where an input signal (the pressure signal) is filtered by the medium response to obtain an output signal (the seismic signal). What we need to find is the medium response (or the filter). The transfer functions were computed as the quotient of the cross-spectrum of the input signal (infrasound signal) and the output signal (seismic signal)

and the power spectrum of the input signal (infrasound signal). The approach makes use of Welch's averaged periodogram (Welch 1967) method. The length of each block was 64 points (1.6 seconds). The individual blocks in each sample record were overlapped by 50% and a Hanning window was used to reduce the variance of the estimate. Due to the short duration, the signals were also prewindowed using a Parzen window.

The magnitudes of the transfer functions have approximately similar shapes for the four channels (Fig. 4) with a certain amount of scatter for the

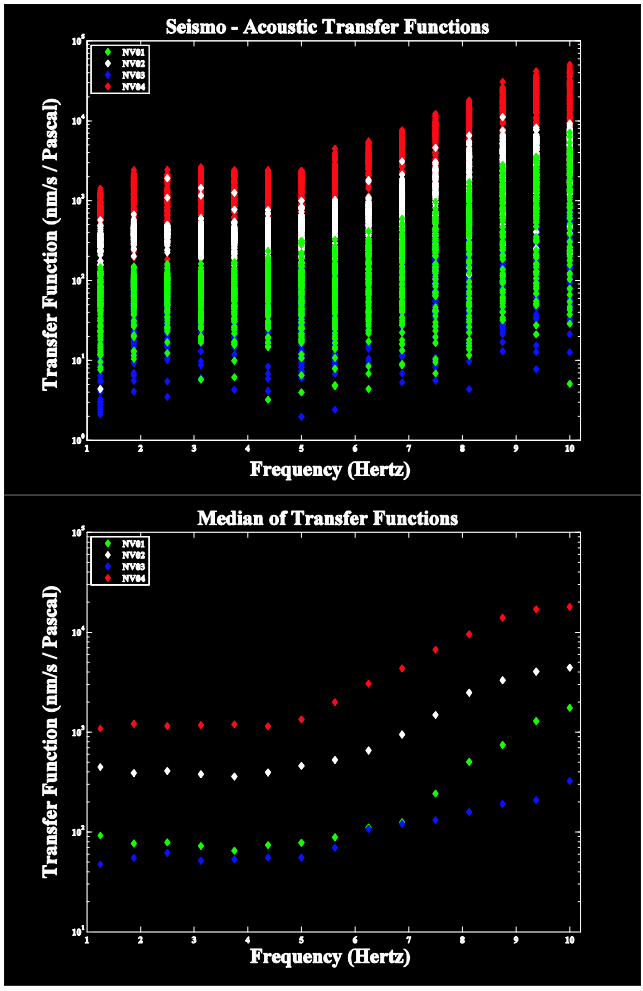


Fig. 4. Magnitudes of the transfer functions for the 109 signals selected for processing (upper plot), and the median of the transfer functions (lower plot). The increase above 5 Hz is due to *P* reverberations and *P* to *S* conversions in the very shallow layer. The approximately flat part of the spectrum up to 4 Hz is modeled as an atmospheric load. Colour version of this figure is available in electronic edition only.

low energy signals. The transfer functions are approximately flat to velocity/pressure up to 5 Hz, after which they start increasing. However, as expected, the magnitudes depend strongly on the site conditions. NV01 and NV03 are approximately at the same level, the median value for a 1 Pa pressure signal being 60 to 80 nm/s ground velocities for the flat part of the spectrum. At NV02, the ground velocities are more than 300 nm/s, while at NV04 the median ground velocities are on the order of 1200 nm/s, for the same 1 Pa pressure amplitude signal.

This increase above 5 Hz is represented by *P* reverberation and *P* to *SV* conversions in the shallow layers (approximately 10 m deep), with a mechanism similar to the one suggested by Langston (2004). However, in our case the short period vertical seismometers are located in approximately 13.3 m boreholes and coupled signals are recorded (though much attenuated) even by the broadband accelerometers located in competent granite at a depth of 60 m (station NV32). Therefore, the mechanism of coupling appears to be dominated by two effects: at low frequencies we model the amplitudes of the ground motion as instantaneous load, while at the high frequency the ground motions can be modeled with the approach suggested by Langston (2004). From a wavelength point of view, the lower frequency components of the signals have longer wavelength, and are insensitive to the finer structure, while the higher frequencies are sampling the shallower structure.

5. THEORETICAL ASPECTS

Sorrells (1971) was the first to derive an expression for the displacement ground response to a plane pressure wave, and a similar model slightly modified is adopted. The model is presented in detail in the Appendix, therefore only a brief discussion of it will be provided here.

The medium is approximated by a homogeneous and isotropic elastic half space. However, the velocity of the source wave (noted as c) is different. The original model (Sorrells 1971) was developed using wind speed velocities, much lower than the actual elastic velocities of the medium and the displacements were solved by approximation, finding the limit of the terms containing c^2. In our case, the speed of the plane wave is the velocity of sound in air (around 330 m/s), much higher than the wind speed velocities, and even exceeding the shear wave velocity of the low-speed materials. Therefore, it is necessary to take into account all the terms. The final expression for the displacement response to a plane pressure wave is given in relations (9) and (10) from Appendix. These two relationships were used to compare the observed and the calculated velocities. It is also worth noting that in the previous studies the theory was compared with long period obser-

vations, generated by nuclear explosions, while in our case the signals are mostly high frequency, above 2 Hz.

The displacement relationships (eqs. 9 and 10 in Appendix) are depending on the material properties of the medium, have a $1/\omega$ dependency and a $\pi/2$ phase shift between the horizontal and vertical displacements, similarly to Rayleigh wave displacements. Beside the trivial solution there are 3 zeros, one for the vertical displacements and two for the horizontal displacements. The zero for the vertical velocities is reached when the velocity of the pressure waves equals the compressional speed velocity of the medium. This is actually the point where the polarizations of the induced waves change. All the infrasound signals used in this study have normal sound speed, below the compressional wave velocity of the medium. Therefore, the polarizations were not tested here, but changes in polarizations were observed for the ground motions induced by Arkansas bolide (Langston 2004). The other two zeros are reached for very high c values, $c = \pm\alpha^2 \Big/ \sqrt{-\alpha^2 + 4\beta^2}$, which cannot be reached in a real physical model. Also note that when the velocity of the pressure wave exceeds the shear wave velocity of the medium, an additional phase shift is introduced.

6. DISCUSSION

To determine the physical properties of the medium necessary to compute the ground velocities, seismic refraction studies were conducted at sites NV01 and NV04 using a portable 12-channel EG&G seismograph and a BETSY seisgun seismic source. At NV01 the spacing of the geophones was 5 m, and the source was placed at 5 m offset. The medium increases gradually with depth in the upper 4 m from a very soft material (P-wave velocity of 457 m/s) to 2032 m/s at depth (Fig. 5). In addition, a shear geophone was lowered in the hole of the GS13 seismometer and an average velocity of 1520 m/s for the upper 12 m (the depth of the seismometer) was found. By weighting the velocities obtained through the seismic refraction survey with the layer thickness and then taking the average, an average value of 1507 m/s was obtained, close to the real result; therefore, the model appears to be accurate. The very soft nature of the upper layer could be observed on the first trace of the refraction profile, where the pressure wave arrives before the direct P wave.

A similar refraction survey at site NV04 (5 m spacing with 5 m offset) did not yield good results and the spacing of the geophones was increased to 10 m. The final model is composed of a layer over half space. The P-wave velocity of the upper layer is 570 m/s and has a thickness of approximately 30 m, and then the velocity increases to over 1800 m/s for the half space.

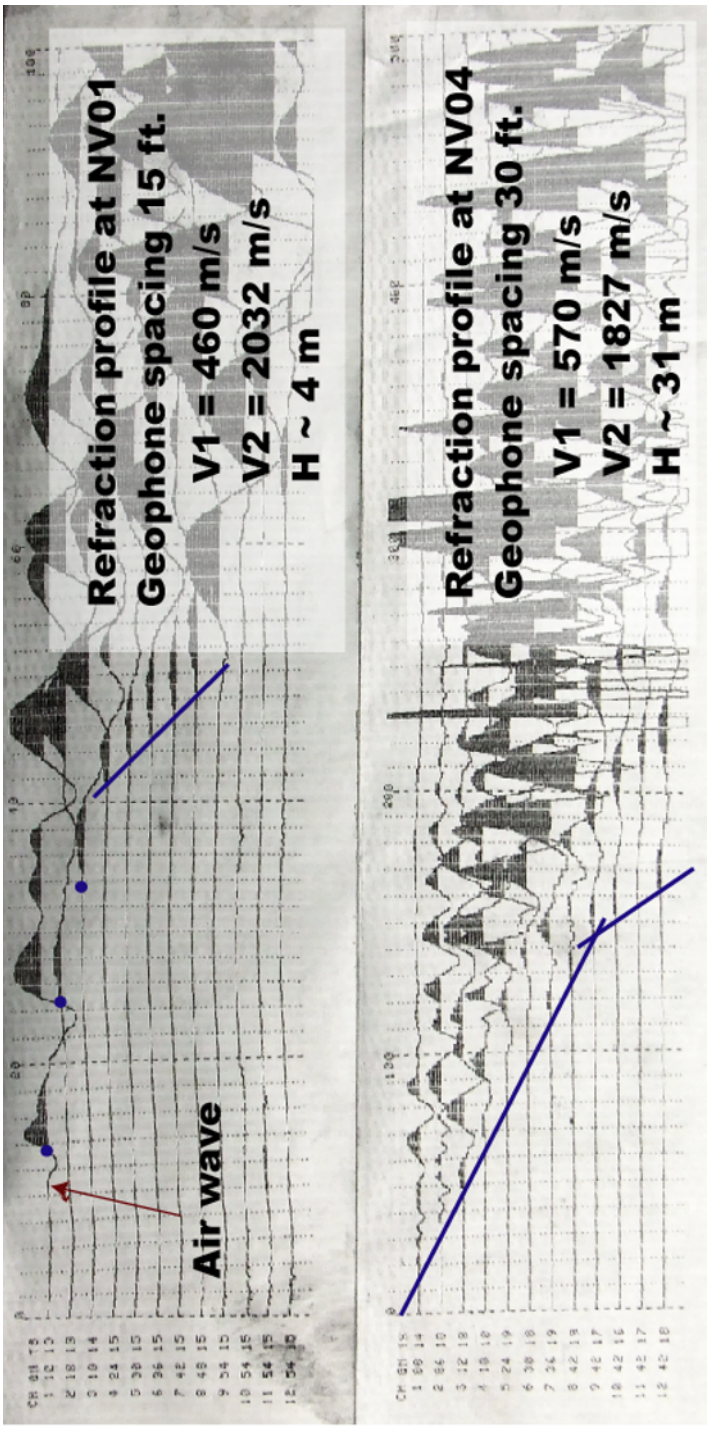


Fig. 5. Refraction profiles at NV01 and NV04 (photography of paper record).

The theoretical velocities were computed using relationships (9) (which includes the depth term, z) and (10) (which give the displacements at the surface) from Appendix multiplied with $i\omega$ to account for velocity and assuming values for the Poisson ratio necessary to compute shear velocities and elastic constants of the medium (Fig. 6). The shear wave velocity was computed assuming a 0.3 value for the Poisson ratio for both sites, and the density of the medium was assumed to be 2.75 g/cm^3 for NV01 (limestone) and 2.0 for NV04 (volcanic tuff). Good agreement with the observation was obtained for a *P*-wave velocity of 570 m/s for NV04 with no depth term and for 1975 m/s (close to the measured velocity), with a depth term of 12 m for NV01 (Fig. 6).

An interesting problem is how these signals attenuate with depth. The attenuation is frequency dependent, but it is also dependent on the medium properties (eqs. 9 from Appendix). The median of the transfer functions at NV01 shows a small decrease in the ground velocities from about 80 nm/s close 1 Hz to around 60 nm/s at 5 Hz. This decrease could be explained by introducing an attenuation term if we consider the depth of seismometer (approx. 12 m).

The attenuation with depth is best illustrated at site NV32 located in Garfield flats, in a homogeneous granite. Site NV32 is a KS54000 borehole accelerometer situated at a depth of 60 m. Because of the greater depth and

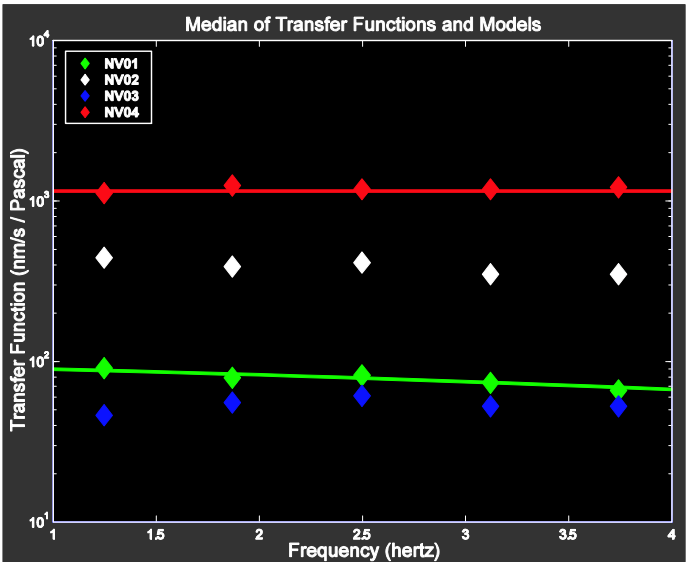


Fig. 6. Median of the transfer functions (filled diamonds) and computed velocities for sites NV01 and NV04 (solid lines). Colour version of this figure is available in electronic edition only.

the frequency content of signals, only a few of the large signals are observed on the seismometer. However, it is the only site which approaches the half space approximation. The Poisson ratio for such a medium is close to 2.5 and the *P*-wave velocity is at least 5800 m/s. Some infrasound signals with amplitudes of around 1 Pa on the infrasound sensors (amplitudes measured at NV01 to NV04) generated more than 20 nm/s^2 of ground motion on NV32. The computed ground accelerations were over 100 nm/s^2 at the surface for a 1 Pa amplitude pressure wave, while including an attenuation term the ground accelerations are 30 nm/s^2, which is well within the range of the observations. The most important disadvantage at this site is the lack of an infrasound sensor and there is no way of knowing whether the infrasound signal shown in Fig. 7 is weaker or stronger at site NV32, but even a weaker signal could not explain almost a factor of 4 in attenuation. Therefore, same decaying is due to the depths of the seismometer.

An application of these signals could be found by estimating site effects. An empirical approach would not work as the median spectral ratio

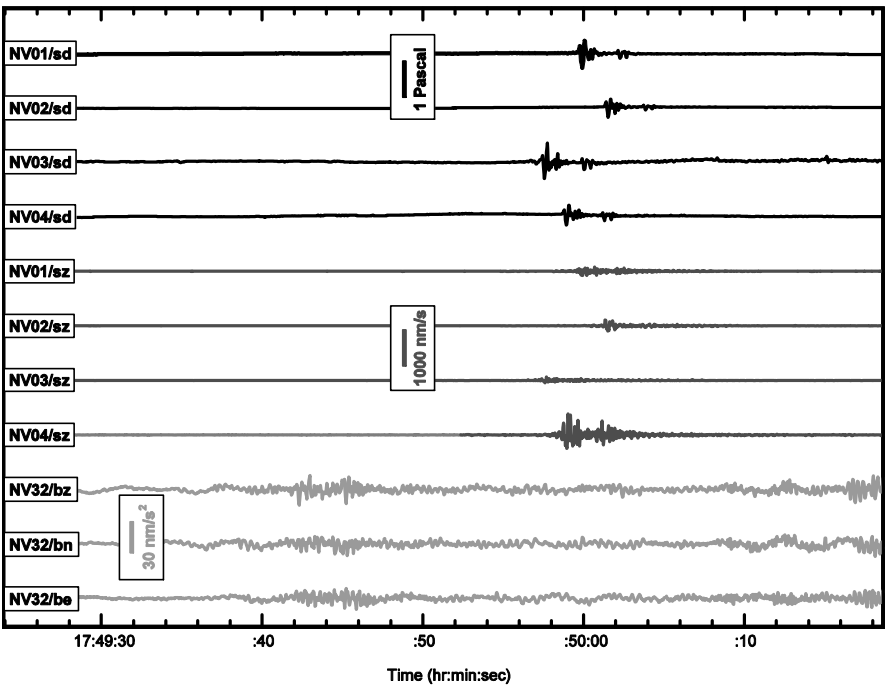


Fig. 7. Example of air-to-ground coupled waves at NV32. Site NV32 is located in granitic rocks at a depth of 60 m. The upper 4 traces (sd symbols) are the infrasound waveforms, the middle 4 traces (sz symbols) are the corresponding short-period waveforms, and the last three waveforms are observed vertical (bz) and horizontal (bn and be) ground motion components at site NV32.

NV04/NV01 of the purely seismic phases shows variations less than 4 in the 1-4 Hz frequency band, while the median ratio of the transfer functions shows variations more than one order of magnitude. In the standard approach proposed by USGS, site effects are estimated by using average shear wave velocity for the upper 30 m of the medium (Borcherdt 1994, Martin and Dobry 1994, and Boore and Joyner 1997). Therefore, accurate site effects could be obtained from air-to-ground coupled waves by inverting the transfer functions for average shear wave velocity (assuming values for the Poisson ratio) and then computing site effects. Both the observations and the theoretical approach suggest this is possible.

APPENDIX

In this Appendix we will derive an appropriate relationship for displacements in a isotropic and homogeneous half space due to a moving atmosphere load. The methodology is similar to the Sorrells (1971) approach, but certain modifications are required, and will be discussed in the text.

In a Cartesian coordinate system (x, y, z) we restrict movement only in x and z directions, therefore all the y components of the motion are null. Then the horizontal and vertical displacements can be written in terms of the potential functions $\Phi(\omega)$ and $\Psi(\omega)$ as

$$u = \frac{\partial \Phi}{\partial x} - \frac{\partial \Psi}{\partial z}, \qquad w = \frac{\partial \Phi}{\partial z} + \frac{\partial \Psi}{\partial x}, \tag{1}$$

where the Heimholtz potentials can be written as

$$\begin{aligned} \Phi(\omega) &= A \exp\left[\mathrm{i}(\omega t - kx) - z\sqrt{k^2 - \frac{\omega^2}{\alpha^2}} \right], \\ \Psi(\omega) &= B \exp\left[\mathrm{i}(\omega t - kx) - z\sqrt{k^2 - \frac{\omega^2}{\alpha^2}} \right], \end{aligned} \tag{2}$$

where α is the compressional velocity of the half space, β is the shear velocity of the half space, ω is the angular frequency, A and B are constants to be determined from the initial conditions, and k is the wavenumber of the plane wave moving with the velocity c ($k = \omega/c$).

It can be shown that $\Phi(\omega)$ and $\Psi(\omega)$ satisfy the wave equation:

$$\nabla^2 \Phi = \frac{1}{\alpha^2}\frac{\partial^2 \Phi}{\partial t^2}, \qquad \nabla^2 \Psi = \frac{1}{\beta^2}\frac{\partial^2 \Psi}{\partial t^2}. \tag{3}$$

Then the vertical and horizontal stresses can be written as

$$p_z = \lambda \nabla \Phi + 2\mu\left(\frac{\partial^2 \Phi}{\partial z^2} + \frac{\partial^2 \Psi}{\partial x \partial z}\right), \qquad p_x = \mu\left(2\frac{\partial^2 \Phi}{\partial x\, \partial z} + \frac{\partial^2 \Psi}{\partial x^2} - \frac{\partial^2 \Psi}{\partial z^2}\right). \tag{4}$$

A plane pressure wave moving with the velocity c can be written as

$$P(\omega) = P_0 \exp\left[\mathrm{i}\omega\left(t - \frac{x}{c}\right)\right], \tag{5}$$

where P_0 is the amplitude of the pressure wave.

The initial conditions require that the horizontal stress vanish at the surface ($z = 0$) and the vertical stress is equal to the stress generated by the load. Therefore, from relations (4) and (5) we have:

$$p_z = \lambda \nabla \Phi + 2\mu\left(\frac{\partial^2 \Phi}{\partial z^2} + \frac{\partial^2 \Psi}{\partial x \partial z}\right) = P(\omega), \qquad p_x = \mu\left(2\frac{\partial^2 \Phi}{\partial x\, \partial z} + \frac{\partial^2 \Psi}{\partial x^2} - \frac{\partial^2 \Psi}{\partial z^2}\right) = 0. \tag{6}$$

The resulting system has two unknown and two equations. Solving the system for A and B we obtain:

$$A = \frac{P_0}{\dfrac{2\mu\omega^2}{c^2} - \dfrac{(\lambda + 2\mu)\omega^2}{\alpha^2} + \dfrac{4\beta^2 \mu \omega^2 \sqrt{\dfrac{1}{c^2} - \dfrac{1}{\alpha^2}}\sqrt{\dfrac{1}{c^2} - \dfrac{1}{\beta^2}}}{c^2 - 2\beta^2}}, \qquad B = \frac{2\mathrm{i}c^3 P_0 \alpha^2 \beta^2 \sqrt{\dfrac{1}{c^2} - \dfrac{1}{\alpha^2}}}{\omega^2\left[\left(c^2 - 2\beta^2\right)\left[-2\alpha^2\mu + c^2(\lambda + 2\mu)\right] - 4c^2\alpha^2\beta^2\mu\sqrt{\dfrac{1}{c^2} - \dfrac{1}{\alpha^2}}\sqrt{\dfrac{1}{c^2} - \dfrac{1}{\beta^2}}\right]}. \tag{7}$$

This result is replaced in eq. (2) and the potential functions can be written as

$$\Phi = \frac{c^2 P_0 \alpha^2 \left(c^2 - 2\beta^2\right) \exp\left[\mathrm{i}(\omega t - kx) - z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\right]}{\omega^2\left[2\alpha^2\left(c^2 - 2\beta^2\right)\mu - c^2\left(c^2 - 2\beta^2\right)(\lambda + 2\mu) + 4c^2\alpha^2\beta^2\mu\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right]},$$

$$\Psi = \frac{2\mathrm{i}c^3 P_0 \alpha^2 \beta^2 \sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}} \exp\left[\mathrm{i}(\omega t - kx) - z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right]}{\omega^2\left[\left(c^2 - 2\beta^2\right)\left[-2\alpha^2\mu + c^2(\lambda + 2\mu)\right] - 4c^2\alpha^2\beta^2\mu\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right]}, \tag{8}$$

then the displacements can be computed from relation (1). The final form of the displacements is in eqs. (9) and (10):

$$u = \frac{\mathrm{i}cP_0\alpha^2\left[\exp\left(z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right)\left(c^2 - 2\beta^2\right) + 2c^2\exp\left(z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\right)\beta^2\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right]}{\omega\left[\left(c^2 - 2\beta^2\right)\left[-2\alpha^2\mu + c^2(\lambda + 2\mu)\right] - 4c^2\alpha^2\beta^2\mu\sqrt{\left(\frac{1}{c^2} - \frac{1}{\alpha^2}\right)\left(\frac{1}{c^2} - \frac{1}{\beta^2}\right)}\right]}$$

$$\times \exp\left[\mathrm{i}(\omega t - kx) - z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right],$$

$$w = -\frac{c^2 P_0 \alpha^2 \sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\left[2\exp\left(z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\right)\beta^2 + \exp\left(z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right)\left(c^2 - 2\beta^2\right)\right]}{\omega\left[2\alpha^2\mu\left(c^2 - 2\beta^2\right) - c^2\left(\lambda + 2\mu\right)\left(c^2 - 2\beta^2\right) + 4c^2\alpha^2\beta^2\mu\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right]} \tag{9}$$

$$\times \exp\left[\mathrm{i}(\omega t - kx) - z\omega\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right].$$

At the surface ($z = 0$) the displacements become:

$$u_0 = \frac{\mathrm{i}cP_0\alpha^2\left(-2\beta^2 + c^2 + 2\beta^2c^2\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right)\exp\left[\mathrm{i}(\omega t - kx)\right]}{\omega\left[\left(c^2 - 2\beta^2\right)\left[-2\alpha^2\mu + c^2\left(\lambda + 2\mu\right)\right] - 4c^2\alpha^2\beta^2\mu\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\right]},$$

$$w_0 = \frac{c^4 P_0 \alpha^2 \sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}} \exp\left[\mathrm{i}(\omega t - kx)\right]}{\omega\left[2\alpha^2\mu\left(c^2 - 2\beta^2\right) - c^2\left(c^2 - 2\beta^2\right)(\lambda + 2\mu) + 4c^2\alpha^2\beta^2\mu\sqrt{\frac{1}{c^2} - \frac{1}{\alpha^2}}\sqrt{\frac{1}{c^2} - \frac{1}{\beta^2}}\right]}. \tag{10}$$

In the original derivation of the formula (Sorrells 1971) velocity c is very small compared to the shear or compressional wave velocity of the medium and the expression was considerably simplified. In our case, the velocity is approximately constant and is equal to the velocity of sound in air (c = 330 m/s), and can even exceed the shear wave velocity of the medium for the soft materials. Therefore, all the terms need to be taken into account in the computation of the displacements. In addition, the observations used to match the theoretical displacements were long period recordings below 1 Hz, whereas the present study uses signals with dominant frequency around 2 Hz.

References

Anglin, F.M., and R.A.W. Haddon (1987), Meteoroid sonic shock-wave-generated seismic signals observed at a seismic array, *Nature* **328**, 607-609, DOI: 10.1038/328607a0.

Bonilla, L.F., J.H. Steidl, G.T. Lindley, A.G. Tumarkin, and R.J. Archuleta (1997), Site amplification in the San Fernando Valley, California: Variability of site-effect estimation using the S-wave, coda and H/V methods, *Bull. Seism. Soc. Am.* **87**, 3, 710-730.

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

Borcherdt, R.D. (1994), Estimates of site-dependent response spectra for design (methodology and justification), *Earthq. Spectra* **10**, 4, 617-653, DOI: 10.1193/1.1585791.

Capon, J. (1969a), High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE* **57**, 8, 1408-1418, DOI: 10.1109/PROC.1969.7278.

Capon, J. (1969b), Investigation of long-period noise at the Large Aperture Seismic Array, *J. Geophys. Res.* **74**, 12, 3182-3193, DOI: 10.1029/JB074i012p03182.

Castro, R.R., M. Mucciarelli, F. Pacor, and C. Petrungaro (1997), S-wave site-response estimates using horizontal-to-vertical spectral ratios, *Bull. Seism. Soc. Am*. **87**, 1, 256-260.

Chen, S.-Z., and G.M. Atkinson (2002), Global comparisons of earthquake source spectra, *Bull. Seism. Soc. Am.* **92**, 3, 885-895, DOI: 10.1785/0120010152.

Chilo, J. (2008), Feature extraction for low-frequency signal classification, Ph.D. Thesis, Department of Physics, Royal Institute of Technology, Stockholm, Sweden.

Chilo, J., A. Jabor, L. Liszka, Å.J. Eide, T. Lindblad, and L. Persson (2006), Infra-sonic and seismic signals from earthquake and explosions in Arequipa, Peru, Western Pacific Geophysics Meeting, 24-27 July 2006, Beijing.

Donn, W.L., I. Dalins, V. McCarty, M. Ewing, and G. Kaschak (1971), Air-coupled seismic waves at long range from Apollo launchings, *Geophys. J. Roy. Astron. Soc.* **26**, 1-4, 161-171, DOI: 10.1111/j.1365-246X.1971.tb03389.x.

Evers, L.G., L. Ceranna, H.W. Haak, A. Le Pichon, and R.W. Whitaker (2007), A seismoacoustic analysis of the gas-pipeline explosion near Ghislenghien in Belgium, *Bull. Seism. Soc. Am.* **97**, 4147-425, DOI: 10.1785/0120060061.

Field, E.H., and K.H. Jacob (1995), A comparison and test of various site-response estimation techniques, including three that are not reference-site dependent, *Bull. Seism. Soc. Am.* **85**, 4, 1127-1143.

Goforth, T.T., and J.A. McDonald (1970), A physical interpretation of seismic waves induced by sonic booms, *J. Geophys. Res.* **75**, 26, 5087-5092, DOI: 10.1029/JB075i026p05087.

Herrin, E.T, J. Bonner, P. Golden, C. Hayward, G. Sorrells, J. Swanson, and I.M. Tibuleac (1998), Reducing false alarms with seismo-acoustic synergy, Proc. 20th Annual Seismic Research Symposium on Monitoring a Comprehensive Nuclear-Test-Ban Treaty, 21-23 September 1998, Santa Fe, NM.

Herrin, E., J.L. Bonner, P. Golden, C. Hayward, G.G. Sorrells, J. Swanson, and I.M. Tibuleac (1999), Reducing false alarms with seismo-acoustic synergy, Proc. 21st Seismic Research Symposium: Technologies for Monitoring the Comprehensive Nuclear-Test-Ban Treaty, 21-24 September 1999, Las Vegas, NV.

Langston, C.A. (2004), Seismic ground motions from a bolide shock wave, *J. Geophys. Res.* **109**, B12309, DOI: 101029/2004JB003167.

Martin, G.R., and R. Dobry (1994), Earthquake site response and seismic code provisions, *NCEER Bull.* **8**, 4, 1-6.

Mayeda, K., S. Koyanagi, and K. Aki (1991), Site amplification from S-wave coda in the Long Valley caldera region, California, *Bull. Seism. Soc. Am.* **81**, 6, 2194-2213.

McDonald, J.A., and T.T. Goforth (1969), Seismic effects of sonic booms: Empirical results, *J. Geophys. Res.* **74**, 10, 2637-2647, DOI: 10.1029/JB074i010p 02637.

McKenna, M.H., and E.T. Herrin (2006), Validation of infrasonic waveform modeling using observations of the STS107 failure upon reentry, *Geophys. Res. Lett.* **33**, L06811, DOI: 10.1029/2005GL024801.

NVAR Certification Manual (1999), Prepared for the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization Provisional Technical Secretariat International Monitoring System Division.

Phillips, W.S., and K. Aki (1986), Site amplification of coda waves from local earthquakes in central California, *Bull. Seism. Soc. Am.* **76**, 3, 627-648.

Riepl, J., P.-Y. Bard, D. Hatzfeld, C. Papaioannou, and S. Nechtschein (1998), Detailed evaluation of site-response estimation methods across and along the

sedimentary valley of Volvi (EURO-SEISTEST), *Bull. Seism. Soc. Am.* **88**, 2, 448-502.

Siddiqqi, J., and G.M. Atkinson (2002), Ground-motion amplification at rock sites across Canada as determined from the horizontal-to-vertical component ratio, *Bull. Seism. Soc. Am.* **92**, 2, 877-884, DOI: 10.1785/0120010155.

Sorrells, G.G. (1971), A preliminary investigation into the relationship between long-period seismic noise and local fluctuations in the atmospheric pressure field, *Geophys. J. Roy. Astron. Soc.* **26**, 1-4, 71-82, DOI: 10.1111/j.1365-246X.1971.tb03383.x.

Sorrells, G.G., J.A. McDonald, Z.A. Der, and E.T. Herrin (1971a), Earth motion caused by local atmospheric pressure changes, *Geophys. J. Roy. Astron. Soc.* **26**, 1-4, 83-98, DOI: 10.1111/j.1365-246X.1971.tb03384.x.

Sorrells, G., J.A. McDonald, and E.T. Herrin (1971b), Ground motions associated with acoustic waves, *Nature Physical Science* **229**, 14-16.

Sorrells, G.G., E.T. Herrin, and J.L. Bonner (1997), Construction of regional ground truth databases using seismic and infrasound data, *Seismol. Res. Lett.* **68**, 743-752.

Sorrells, G., J. Bonner, and E.T. Herrin (2002), Seismic precursors to space shuttle shock fronts, *Pure Appl. Geophys.* **159**, 5, 1153-1181, DOI: 10.1007/s00024-002-8676-0.

Stewart, J.H. (1980), Geology of Nevada. A discussion to accompany the Geologic Map of Nevada, Nevada Bureau of Mines and Geology, Spec. Publ. 4, 136 pp.

Su, F., and K. Aki (1995), Site amplification factors in Central and Southern California determined from coda waves, *Bull. Seism. Soc. Am.* **85**, 2, 452-466.

Theodulidis, N., P.-Y. Bard, R. Archuleta, and M. Bouchon (1996), Horizontal-to-vertical spectral ratio and geological conditions: The case of Garner Valley Downhole Array in southern California, *Bull. Seism. Soc. Am.* **86**, 2, 306-319.

Tibuleac, I.M., and E.T. Herrin (1997), Calibration studies at TXAR, *Seism. Res. Lett*. **68**, 353-365.

Tibuleac, I.M., and E.T. Herrin (2001), Detection and location capability at NVAR for events on the Nevada test site, *Seism. Res. Lett*. **72**, 1, 97-107.

Welch, P.D. (1967), The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms, *IEEE Trans. AU* **15**, 70-73.

Received 12 June 2009
Received in revised form 31 March 2010
Accepted 8 April 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1040-1055
DOI: 10.2478/s11600-010-0023-8

EM and GPR Investigations of Contaminant Spread Around the Hoc Mon Waste Site, Vietnam

Nguyen VAN GIANG[1], Guy MARQUIS[2], and Le Huy MINH[1]

[1]Institute of Geophysics, Vietnamese Academy of Science and Technology, Hanoi, Vietnam, e-mail: giangnv@igp-vast.vn

[2]EOST-IPGS, University of Strasbourg and CNRS, Strasbourg, France
e-mail: gm.eost@gmail.com

Abstract

This paper is presenting the results from near-surface geophysical surveys near the waste site of Hoc Mon in southern Vietnam where leachate contamination has been recognized at the surface. Using EM and GPR surveys, we were able to determine the lateral extent of a contaminated area of high electrical conductivity and have identified channels that concentrate the contaminant flow. The simple relationship between the electrical resistivity and the leachate concentration is suggested and estimated the *in situ* leachate concentration from the inversion of the EM data; values are as high as 40%. Thanks to a permeability barrier leachate flow is confined to the shallow subsurface, making it easier to apply possible site remediation projects.

Key words: electromagnetic methods, ground-penetrating radar, leachate flow, contaminant concentration.

1. INTRODUCTION

Large waste sites are known to have mid- to long-term negative impact on the environment. This is especially true when domestic waste is present because organic matter may be introduced into the subsurface. This organic matter undergoes biochemical transformations that produce toxic wastes

which can be carried over large distances by groundwater flow (e.g., Baedecker and Back 1979, Christensen *et al*. 2000).

Recent geophysical studies around landfills have successfully mapped the extent and/or the depth of subsurface contaminant concentrations. Electromagnetic (e.g., Tezkan 1999, Nobes *et al*. 2000), self-potential (e.g., Naudet *et al.* 2003), borehole (e.g., Jorstad *et al*. 2004) and electrical resistivity (e.g., Bernstone *et al*. 2000, Oh *et al.* 2003, Guérin *et al*. 2004, Abu-Zeid *et al*. 2004) methods have all given reliable results.

The data from a geophysical survey were conducted around the Hoc Mon waste site, 15 km north of Ho Chi Minh City, Vietnam. The site was in operation from June 1991 to December 2002. Today, it consists of a pile of about 30 m (Fig. 1a) of consolidated domestic and construction waste deposited over alluvial sediments from the Saigon River. The subsurface consists of a top clayey sand layer, several metres thick, underlain by sand deposits. The waste pile is in part covered by geo-membrane to prevent landslides that occurred frequently during the rainy season.

The waste site is at a distance of about 250 m from the Sang Canal that reaches the Saigon River about 1.5 km to the east of the study area. Natural groundwater flow is from the waste site toward the Canal. There is therefore concern that contamination from the waste site might reach the Saigon River in the future, which could pose a serious potential health hazard.

To this day, dark water seeps at several locations between the waste site and the Canal (Fig. 1b) suggesting a high degree of underground water contamination by leachate. The rice paddies have now been abandoned there. There is a great uncertainty, however, as to the extent of the contaminated area, as one expects the heavier contamination to be concentrated at depth, and so the contamination observed at the surface may represent only a small fraction of the overall contamination.

As a collaboration programme between the Vietnam Academy of Science and Technology and French National Center for Scientific Research, teams from the Hanoi Institute of Geophysics and Institut de Physique du Globe at Strasbourg have undertaken a geophysical survey of the area between the waste site and the Sang Canal in August 2004, i.e., during the rainy season. We privileged surface electromagnetics for rapid coverage of the subsurface electrical resistivity and ground-penetrating radar to delineate major subsurface structures.

2. GEOPHYSICAL DATA

Controlled-source electromagnetic (EM) and ground-penetrating radar (GPR) were acquired on the site in August 2004, i.e., during the rainy season in southern Vietnam. The difficult field conditions, i.e., working in quick-

Fig. 1: (a) View of the waste hill taken from the field. (b) View of the site investigated taken from the top of the waste pile. The Sang Canal is in the background. The EM survey was made within the box. Leachate is observed at the surface in the dotted area. (c) Location of the profiles from 1 to 10 discussed in the paper.

sand as we got closer to the Canal, prevented us to acquire data all the way to the Sang Canal.

Controlled-source EM

Electromagnetic data were acquired with a Geophex (USA) GEM-2 instrument, along 10 profiles (inset of Fig. 1c). The spatial sampling rate is about 1 m inline and 5 to 10 m cross-line. The GEM2 instrument outputs the ratio of the secondary to primary field (H_s/H_p) in ppm. By acquiring EM data at several frequencies, we expect to be able to determine variations of electrical conductivity with depth (Won *et al.* 1996), although this capability is disputed (McNeill 2000).

Figure 2 shows a plot of the GEM2 response (magnitude of H_s/H_p) for a typical profile at frequencies of 325, 525, 1375, 3525, 9225, 14 875, and 20 025 Hz. As expected from the theory, the amplitude of the response increases with frequency. The overall signal-to-noise ratio (S/N) was calculated by breaking each profile into a set of 10-point windows. We then took each window's mean as the signal and its range of values, i.e., $(H_s/H_p)_{\max} - (H_s/H_p)_{\min}$, as noise, so as to have an S/N value per frequency per window. The values shown in Fig. 3 are obtained by averaging all windows for Profile 4 where the S/N is increased with frequency.

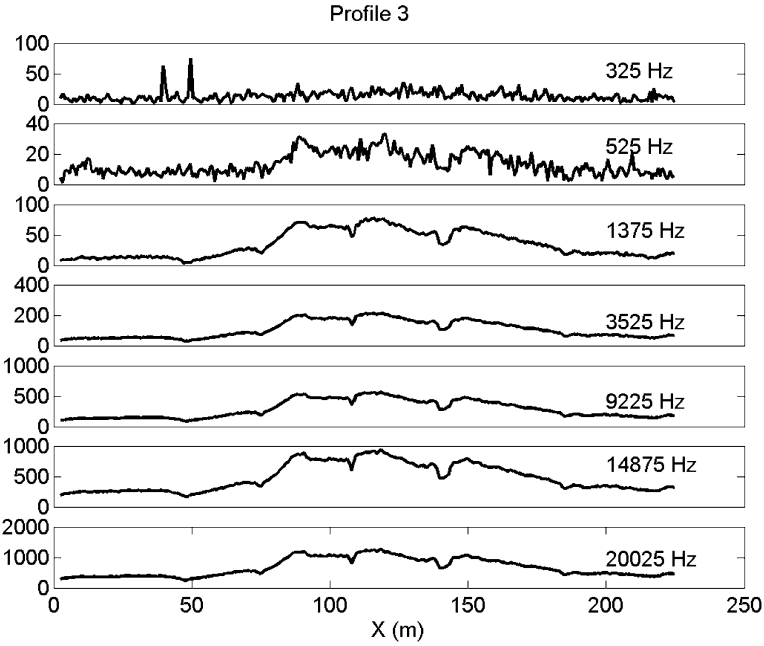


Fig. 2. Magnitude of the H_s/H_p ratio (in ppm) for the seven sampled frequencies along a typical profile. Note the important decrease in S/N at low frequencies.

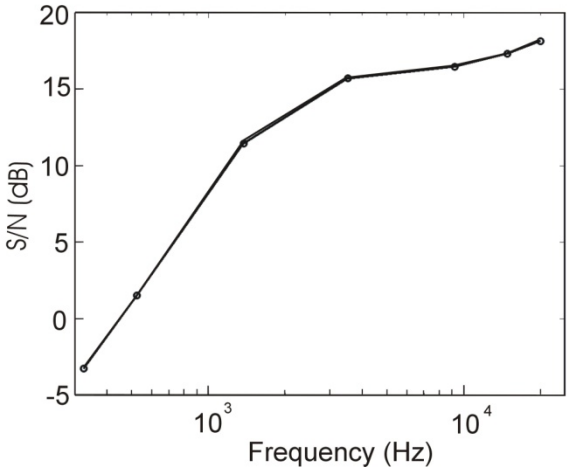


Fig. 3. The GEM2 signal-to-noise ratio in dB for Profile 3. See details in text.

Note that the anomalies observed in the high S/N response curves, i.e., between 1375 and 20 025 Hz in Fig. 2, have similar shapes: anomaly widths are the same and local minima are coincident. This suggests that the area sampled with a single measurement (footprint) does not vary significantly with frequency. It could also just indicate that the conducting target is very shallow.

Figure 4 shows the imaginary response (the part most sensitive to electrical resistivity changes) map for a frequency of 20 025 Hz. We note the presence of a strong EM response – related to a conductive zone – in the bottom of Fig. 4. The observed dark water patches are also within this zone. In addition, two features can be recognized:

- A gradual decrease of the amplitude of the EM imaginary response (i.e., of electrical conductivity) with distance to the waste site. This is "normal" behaviour usually observed where high-conductivity solute migrates from high to low concentration areas, whatever the solute mixing mechanism (i.e., diffusion, dispersion). By analogy with the hand, we will call this the "palm-shaped" anomaly.
- Several narrow, ridge-shaped, high-conductivity anomalies extend in a branch-like fashion toward NW, suggesting the presence of channels of preferential underground water – and contamination – migration. A closer look at data from Profile 5 (Fig. 5) shows clearly these lateral variations with peaks at $X = 90$, 140, and 175 m. This indicates some heterogeneity in the way the alluvial sediments deposited, which is expected in this low-energy depositional environment. We will call these the "finger-shaped" anomalies.

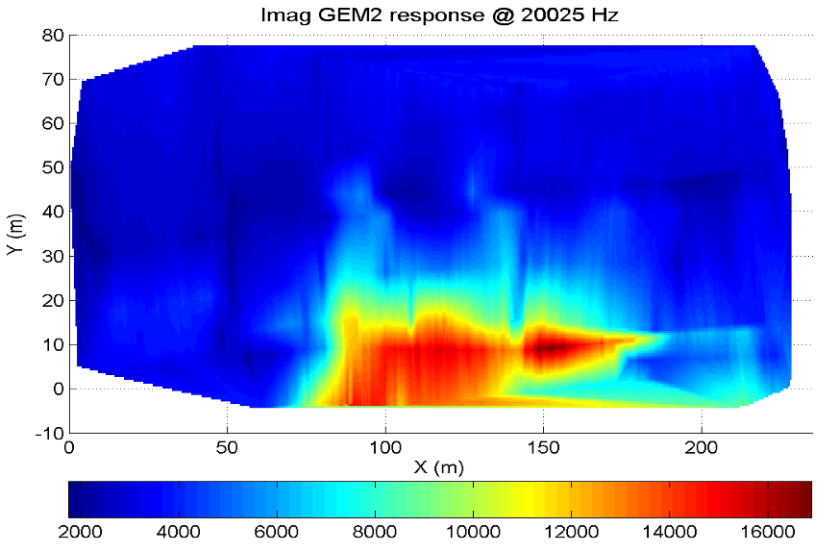


Fig. 4. Map view of the imaginary response (in ppm) at 20 250 Hz. The *X*-axis direction is N70°E. Strong responses (hot colours) are associated to high-conductivity zones. Note the gradual decrease of the response away from the waste site. The arrow-shape pattern indicates the preferential direction of contaminant flow as evidenced by the high-response "fingers". Colour version of this figure is available in electronic edition only.

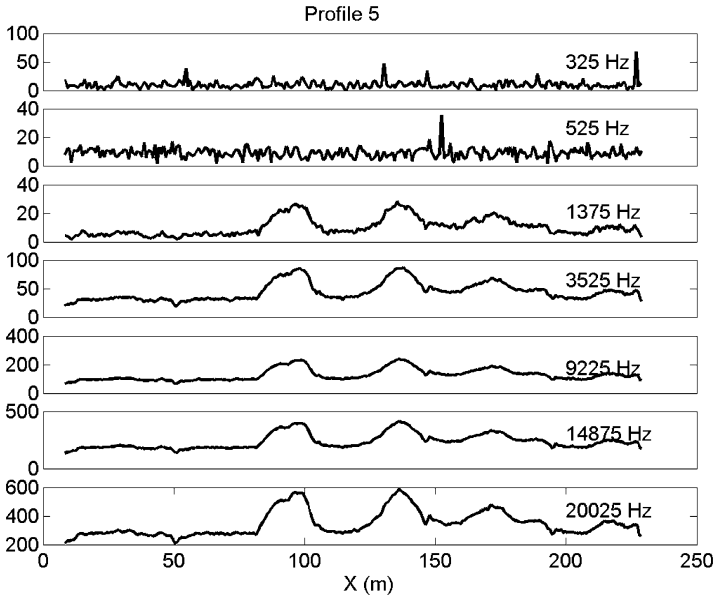


Fig. 5. The GEM2 response for Profile 5 indicating the presence of high-conductivity channels at $X = 90$, 140, and 175 m.

We point out that the "finger" anomalies extend about 20-25 m (i.e., up to 40%) farther away from the waste site than the "palm" anomalies.

Ground-penetrating radar

Bistatic ground-penetrating radar (GPR) data were acquired with a PulseEkko 100 system with 100 MHz unshielded antennae along Profiles 2, 3, and 4. Data acquired in August 2004 did not have a good S/N ratio and were strongly affected by above-surface features that gave diffraction patterns with velocities close to 0.3 m/ns (e.g., Bano *et al*. 2000). We ran, however, another GPR acquisition in May 2005, i.e., during the dry season, and we obtained data of better quality.

The processing was performed with in-house PC based interactive software (Radlab) developed in MatLab® at IPG Strasbourg (Girard 2002). Processing steps included the time-zero correction, running average filter (anti-DC filter), envelope gain, band-pass filtering (10-70 MHz), and constant gain for final display. Figure 6 shows the processed GPR section from Profile 3. A strong reflector can be seen across the profile, undulating around 90 ns two-way-time (twt), and a second, shallower reflector is clearly visible at about 120 ns twt. We believe that these two reflectors contour lens-shaped bodies.

The Common Mid Point (CMP) acquisition was ran in May 2005. We recovered a stacking velocity of 0.138 m/ns for the strongest reflector with a 90 ns twt. These results show that the strong reflector is at about 6.2 m (i.e., 90×0.138/2) depth.

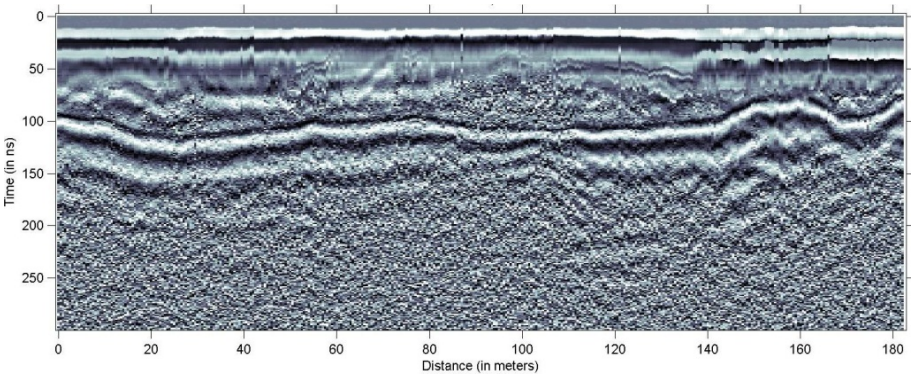


Fig. 6. GPR data (100 MHz) from Profile 3, acquired during the dry season (May 2005), and interpreted reflectors. A strong reflector can be followed throughout the profile at about 90-100 ns twt and an intermediate reflector can be recognised on the left and right sides but is not identified in the middle, suggesting lens-type structures.

We can try to estimate the porosity of this top layer, assuming that the medium follows a simple time-average (i.e., CRIM with an exponent of 0.5) dielectric mixing law and a soil relative dielectric permittivity ε_r ranging between 8 and 12. These values are higher than those usually reported for clayey soils (ε_r = 4-6, e.g., Daniels 1996), but we do not expect this top layer to be dry over its whole 6 metre thickness. We do not have a clear indication on the GPR data (Fig. 6) of where the unsaturated zone ends, but this is expected as the change from dry to saturated medium is gradual (Bano *et al.* 2006). With this approach, we obtain a soil porosity between 16 and 24%. We will choose the mean value of 20% for the remainder of the paper.

3. INVERSION OF EM DATA

As shown earlier, the shape of the EM responses does not vary significantly with frequency, which we interpret as the GE EM2 sampling a similar subsurface area with frequency. The depth range sampled, however, should be different for each frequency as the induction number N depends on the angular frequency ω ($N^2 = \omega\mu\sigma/2$, where μ is the magnetic susceptibility and σ is the electrical conductivity). We have attempted a multi-frequency inversion of the EM data. We have only used the frequencies for which the S/N is always greater than 10 dB, i.e., from 1375 to 20 025 Hz.

Our inversion scheme is a simplified implementation of that proposed by Zhang *et al*. (2000) limited to the case of horizontal coplanar loops. Because the H_s/H_p response varies greatly in magnitude with frequency, we have chosen a simple objective function of the total relative data misfit:

$$\phi = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{d_j^{\mathrm{mod}} - d_j^{\mathrm{obs}}}{d_j^{\mathrm{obs}}}\right)^2, \tag{1}$$

where N is the number of frequencies, d_j^{obs} is the measured imaginary part of the H_s/H_p response at frequency j, and d_j^{mod} is the imaginary part of the H_s/H_p response for a horizontal coplanar (HCP) mode over a layered medium at frequency j (Ward and Hohmann 1988, Won *et al*. 1996):

$$d_j^{\mathrm{mod}} = \mathrm{Im}\left(-r^3\int_0^{\infty}\left(1+R_{TE}\right)\mathrm{e}^{-2\lambda h}\,J_0(\lambda r)\,\mathrm{d}\lambda\right), \tag{2}$$

where R_{TE} is the TE-mode reflection coefficient (Ward and Hohmann 1988), J_0 is the zeroth-order Bessel function, h is the height of the instrument above ground level (1.0 m), and r is the interloop distance (1.66 m). The integration is calculated using the Hankel transform digital filter of Anderson (1979).

Is the use of a 1D inversion scheme actually relevant to study what is clearly a 3D problem (see Fig. 4)? Zhang *et al*. (2000) have tested the validity of this approach by inverting data from synthetic 3D examples: they show that for a horizontal coplanar (HCP) loop configuration, such as the one used in this survey (and also for vertical coplanar and coaxial configurations), the models resulting from 1D inversion are in good agreement with the synthetic models. However, the recovered electrical conductivity values for discrete conductors embedded in a resistive half-space are smaller than expected, though this discrepancy is less significant for shallow targets, which is undoubtedly the case in this study. We are therefore confident that the results from this type of inversion can provide a reliable picture of the near-surface electrical conductivity.

We chose a simple 1D model consisting of a layer of 6 m thickness above a half-space of 100 Ω m resistivity, so only the resistivity of the top layer is allowed to vary. Our choice for the fixed layer thickness was based on the depth of the strong reflector on the GPR data. We assume here that the depth of this reflector, imaged by a single GPR profile, does not vary over the survey area.

The fit of the resulting model to the data is very good: most of the 2859 measurements fall within 5% except for a few (86) measurements that showed misfits as large as 18%. We have checked that there is no spatial

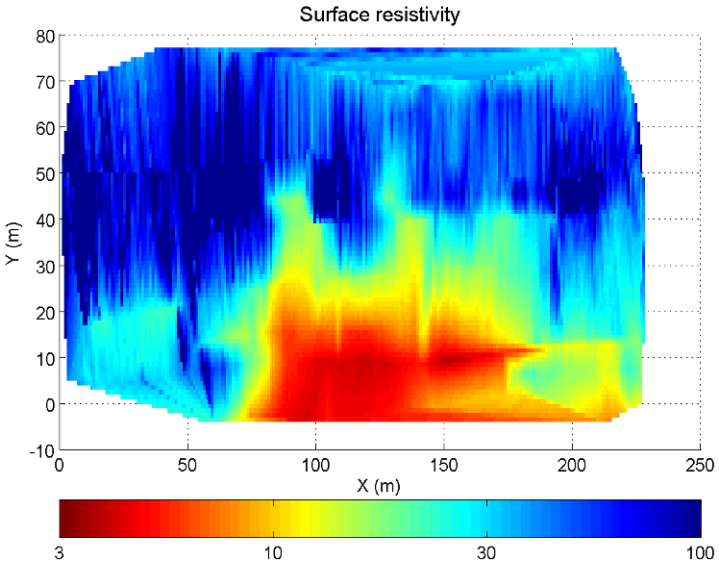


Fig. 7. Electrical resistivity of the top layer (in Ω m) obtained by the inversion of GEM2 data, assuming the top layer is 6 m thick. Colour version of this figure is available in electronic edition only.

correlation of the poor-fitting measurements with the very high or low resistivity values of the top layer.

The 1D inversion was calculated for each data point and the result was a surface electrical resistivity map of the study area (Fig. 7). The surface resistivity values vary from less than 5 Ω m close to the waste site to 40 Ω m away from it.

We also checked for possible artefacts from fixing the top layer thickness and the underlying half-space resistivity by running another 1D inversion with a four-layer model, letting all parameters unconstrained (except for a restriction to positive resistivity and thickness values). The resulting conductance for the top 6 metres reaches values up to 2 S and shows the same anomaly pattern as in Fig. 7.

4. DISCUSSION

Leachate electrical properties

The EM data indicate the presence of a shallow, high electrical conductivity anomaly between the waste site and the Sang Canal. This anomaly extends up to 80 m away from the edge of the waste site (Figs. 4 and 7). The very low resistivity values reported in the literature for soil (e.g., Nobes *et al*. 2000, Bernstone *et al*. 2000, Yoon and Park 2001, Oh *et al*. 2003, Abu-Zeid *et al*. 2004) have usually been related to the presence of leachate infiltration. The dark water observed at the surface in places on the site (Fig. 1a) confirms this interpretation.

Electrical resistivity is a parameter useful for a geophysicist, but if one intends to integrate geophysical data into a hydrological model, say for forecasting purposes, resistivity has to be translated into a relevant hydrological parameter. We already know that porosity and fluid electrical resistivity are among the main factors controlling bulk (i.e., geophysically-determined) electrical resistivity in most shallow environments (e.g., Archie's Law). The latter is almost entirely controlled by solute concentration, if the main conduction mechanism is electrolytic. This suggests that we may use electrical resistivity ρ as a proxy for underground contaminant (here leachate) concentration X in such situations.

As we do not have access to water samples from the study area, we have tried to get a general relationship between ρ and X from published data on the electrical properties of leachates. These data show a great variability that depends not only on leachate concentration, but also on depth, temperature, type of waste, distance to the landfill, etc. Nobes *et al.* (2000) report water conductivity values (at 20°C) between 132 and 302 mS/m (3 to 8 Ω m) in the Burwood landfill in New Zealand, Bernstone *et al.* (2000) have measured 950 to 1370 mS/m (0.8 to 1 Ω m) in the Filborna landfill in Sweden and Jors-

tad *et al*. (2004) report 24 to 165 mS/m (6 to 40 Ω m) at the Astrolabe Park site in Australia. Nobes *et al.* (2000) have obtained a good empirical correlation between the apparent conductivities measured with an EM-31 instrument and the water conductivity. However, this correlation did not enable them to calculate leachate concentrations *in situ*: this extra step requires either a great amount of *in situ* concentration and fluid conductivity measurements or controlled laboratory experiments.

Recent experiments by Oh *et al*. (2003) provide the first – to our knowledge – laboratory results on the effects of leachate concentration on the bulk electrical resistivity of soil samples. They have obtained electrical conductivity-water content relationships for porous sands with aqueous fluids containing 0, 10, 20 and 30% of leachate for water contents between 0 and 30%. These very interesting results are, however, not applicable directly to the situation at Hoc Mon because the soil contains clays that have an important impact on electrical resistivity. We have therefore developed a tentative Rho-*X* relationship for Hoc Mon, following this approach:

- From Fig. 7, we see that away from the waste site, where leachate is not likely to be present, the bulk electrical resistivity is 100 Ω m, which we will assume to be the zero-leachate concentration bulk resistivity value, i.e., the pore fluid already contains other dissolved solids, but no leachate. We will denote its reciprocal σ_0^b, i.e., the bulk electrical conductivity without leachate. We will use a pure leachate resistivity of 0.3 Ω m (3.3 S/m). We can obtain the leachate-only bulk conductivity, denoted σ_{100}^b, using Archie's Law, assuming a 20% porosity (from the GPR velocity, assuming full saturation) and $m = 1.1$, a reasonable value for unconsolidated soils $\sigma_{100}^b = 0.56$ S/m according to:

$$\sigma_{100}^b = \sigma_f \, \phi^m \,. \tag{3}$$

- We also know that fluid electrical conductivity varies linearly with solute concentration, so if Archie's Law is valid for these soils, then so should bulk electrical conductivity. We have therefore used a linear interpolation between the two end-members, $\sigma_0^b = 0.01$ S/m and $\sigma_{100}^b = 0.56$ S/m:

$$\sigma^b = \sigma_0^b + \alpha[X] \tag{4}$$

with $\alpha = 0.55$ S/m per volume concentration. We then take the reciprocal of σ^b to get ρ. We are aware that we neglect surface conduction effects, and remember that clays are present, so this relationship will not be valid at low concentrations. However, as we observed fairly low resistivities throughout, this is likely not a major problem.

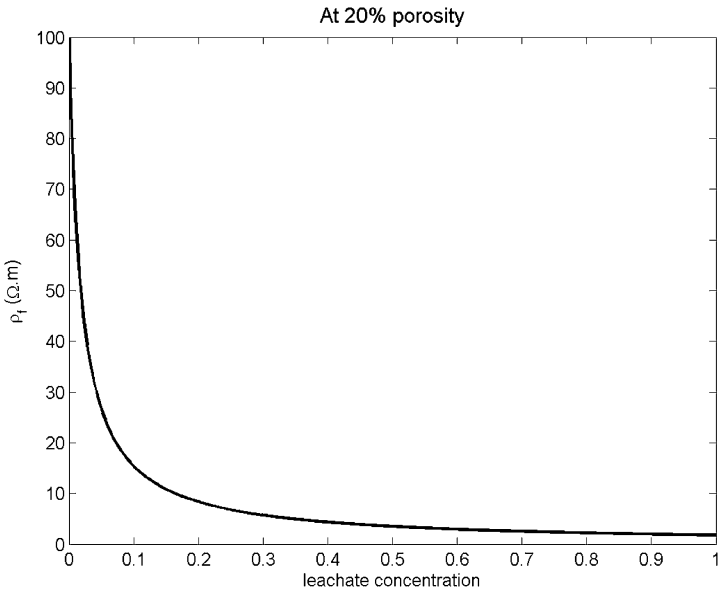


Fig. 8. Bulk resistivity as a function of leachate concentration for Hoc Mon. See text for details of derivation.

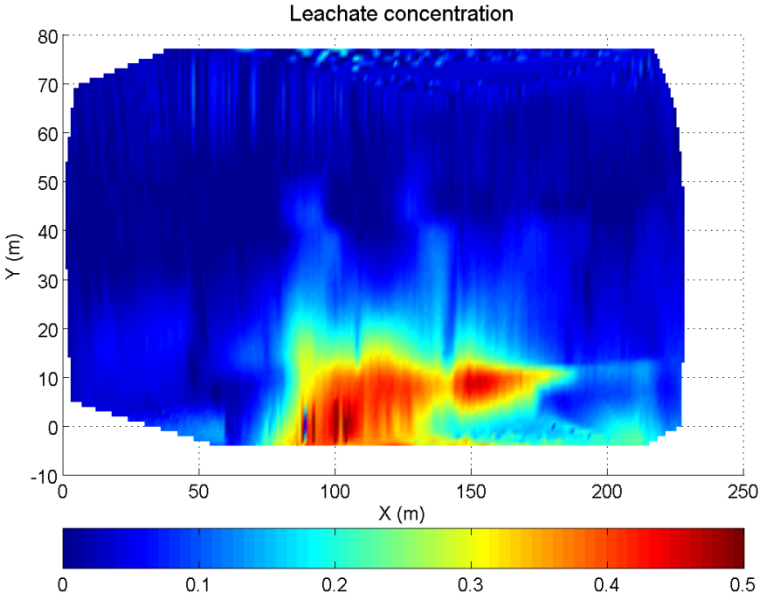


Fig. 9. Shallow leachate concentration obtained by the combination of the electrical resistivities in Fig. 7 and the relationship of eq. (4). Colour version of this figure is available in electronic edition only.

This relationship is illustrated in Fig. 8. Top-layer electrical resistivity values obtained by the inversion of GEM2 data can now be converted into *in situ* leachate concentrations using eq. (2). The results of this calculation are shown in Fig. 9. *In situ* leachate concentrations range between 0 and 30%, with a few extreme values that reach 40%. These leachate concentrations can then be integrated into a hydrological model for history matching and forecasting leachate transport.

Leachate confined to shallow depths

Leachate has a higher density than the ambient groundwater. Therefore, one expects that the leachate will tend to flow not only horizontally along the piezometric head gradient, but also vertically. It is therefore surprising that for the case studied here, leachate is concentrated in the topmost soil layer. Two hypotheses can explain this situation: either the density difference between the leachate-contaminated and the ambient waters is very small and so density-driven flow is negligible, or there is a natural low-permeability barrier that confines the leachate to shallow depths.

We favour this second hypothesis. An impermeable compacted clay layer at depth is expected to give a strong GPR reflector, considering the great contrast in electrical properties (both dielectric permittivity and electrical resistivity) between clay and alluvial sands or gravels (e.g., Ward and Hohmann 1988). There is a strong, continuous reflector at ~6 m depth in the GPR data (Fig. 6) that indeed can be interpreted as such a barrier.

The presence of a natural barrier confining the leachate at shallow depths would protect the deeper aquifers from toxic waste infiltration and also makes it easier to implement the site rehabilitation schemes.

Finger patterns

Figure 4 shows "finger-shaped" anomalies that indicate some form of preferential paths for contaminant transport. Underground flow channels were detected by Nobes *et al*. (2000) by locating differences between high- and low-tide EM responses. These channels could have played an instrumental role in contaminant transport. However, their EM data do not suggest so clear, focused channels as shown here.

As stated earlier, the study area consists of alluvial sediments deposited by the Saigon River, in a low-energy depositional environment. Lateral heterogeneities in sediment composition and depositional geometry are certainly expected. This situation is illustrated in Fig. 10 that compares EM and GPR data for Profile 3. The highest EM responses, i.e., the highest electrical conductivity values, are measured between what we have interpreted as two underground lenses. Accordingly, the channels seen in the EM section are located between these lens-shaped objects.

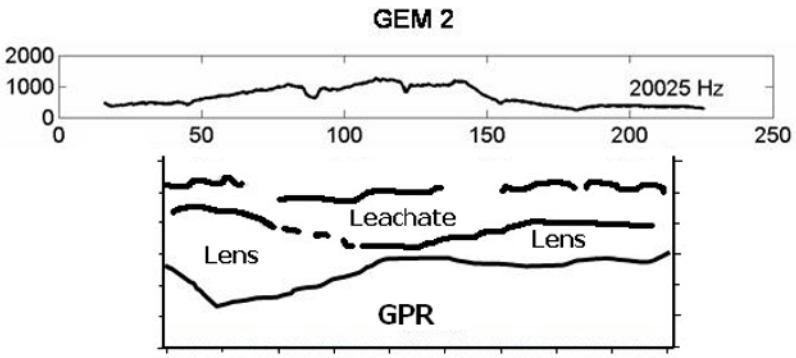


Fig. 10. Comparison of GPR interpretation and EM data for Profile 3. The EM response and hence the electrical conductivity is strong where the lenses are thin. We interpret these lenses as barriers to contaminant flow.

Our results emphasise the relevance of combining surface geophysical methods in contaminated site evaluation studies: a study based only on spot sampling of underground waters could well have missed these narrow channels and thus lead to underestimation of the actual contamination of the site.

5. CONCLUSIONS

EM and GPR studies near a recently-closed waste site in an alluvial plain have enabled delineation of a high-conductivity anomaly that we interpret as leachate-contaminated water, in correlation with surface observations, with leachate concentrations as high as 40%. The shapes of the main anomaly indicate a diffusive pattern for profiles closest to the waste site, but data from far-side profiles reveal a fingering pattern of migration. Heterogeneities in the subsurface sediments, such as clay lenses, can play a major role in the contaminant spread dynamics. According to the geophysical results, leachate flow is confined to the shallow subsurface, thanks to a permeability barrier, facilitating possible site remediation projects.

Acknowledgements. Funding for this study was provided through a cooperation program between French National Center for Scientific Research and Vietnam Academy of Science and Technology. We wish to thank the Director of the Ho Chi Minh City Environmental Company who facilitated the access to its facilities for this case study. We thank Pierre-Daniel Matthey (Ecole et Observatoire des Sciences de la Terre) for their assistance in the field, Maksim Bano for help with GPR data and Mark Everett, Andreas Kemna, and anonymous reviewers for their numerous helpful comments and suggestions.

Paper presented at the Eage 67th Meeting and Exhibition, Madrid.

References

Abu-Zeid, N., G. Bianchini, G. Santarato, and C. Vaccaro (2004), Geochemical characterisation and geophysical mapping of landfill leachates: the Marozzo canal case study (NE Italy), *Environ. Geol*. **45**, 4, 439-447, DOI: 10.1007/s00254-003-0895-x.

Anderson, W.L. (1979), Numerical integration of related Hankel transforms of orders 0 and 1 by adaptive digital filtering, *Geophysics* **44**, 7, 1287-1305, DOI: 10.1190/1.1441007.

Baedecker, M.J., and W. Back (1979), Hydrogeological processes and chemical reactions at a landfill, *Ground Water* **17**, 5, 429-437, DOI: 10.1111/j.1745-6584.1979.tb03338.x.

Bano, M. (2006), Effects of the transition zone above a water table on the reflection of GPR waves, *Geophys. Res. Lett*. **33,** L13309, DOI: 10.1029/2006 GL026158.

Bano, M., G. Marquis, B. Nivière, J.C. Maurin, and M. Cushing (2000), Investigating alluvial and tectonic features with ground-penetrating radar and analyzing diffractions patterns, *J. Appl. Geophys*. **43**, 1, 33-41, DOI: 10.1016/ S0926-9851(99)00031-2.

Bernstone, C., T. Dahlin, T. Ohlsson, and W. Hogland (2000), DC-resistivity mapping of internal landfill structures: two pre-excavation surveys, *Environ. Geol*. **39**, 360-371, DOI: 10.1007/s002540050015.

Christensen, T.H., P.L. Bjerg, S.A. Banwart, R. Jakobsen, G. Heron, and H.J. Albrechtsen (2000), Characterization of redox conditions in groundwater contaminant plumes, *J. Contam. Hydrol*. **45**, 3-4, 165-241, DOI: 10.1016/ S0169-7722(00)00109-1.

Daniels, D.J. (1996), *Surface-Penetrating Radar*, Institution of Electrical Engineers, London, UK, 300 pp.

Girard, J.F. (2002), Imagerie géoradar et modélisation des diffractions multiples, Ph.D. Thesis, University of L. Pasteur, Strasbourg.

Guérin, R., M.L. Munoz, C. Aran, C. Lapperrelle, M. Hidra, E. Drouart, and S. Grellier (2004), Leachate recirculation: moisture content assessment by means of a geophysical technique, *Waste Mgmt*. **24**, 8, 785-794, DOI: 10.1016/j.wasman.2004.03.010.

Jorstad, L.B., J. Jankowski, and R.I. Acworth (2004), Analysis of the distribution of inorganic constituents in a landfill leachate-contaminated aquifer: Astrolabe Park, Sydney, Australia, *Environ. Geol*. **46**, 2, 263-272, DOI: 10.1007/ s00254-004-0978-3.

McNeill, J.D. (2000), Application of dipole-dipole electromagnetic systems for geological depth sounding, Technical Note TN-31, Geonics Limited, Mississauga, Canada.

Naudet, V., A. Revil, J.Y. Bottero, and P. Bégassat (2003), Relationship between self-potential (SP) signals and redox conditions in contaminated groundwater, *Geophys. Res. Lett*. **30**, 21, 2091, DOI: 10.1029/2003GL018096.

Nobes, D.C., M.J. Armstrong, and M.E. Close (2000), Delineation of a landfill leachate plume and flow channels in coastal sands near Christchurch, New Zealand, using a shallow electromagnetic survey method, *Hydrogeol. J*. **8**, 3, 328-336, DOI: 10.1007/s100400000063.

Oh, M.H., J.H. Lee, G.L. Yoon, and J. Park (2003), Pilot-scale field model tests for detecting landfill leachate intrusion into the subsurface using a grid-net electrical conductivity measurement system, *Environ. Geol*. **45**, 2, 181-189, DOI: 10.1007/s00254-003-0880-4.

Tezkan, B. (1999), A review of environmental applications of quasi-stationary electromagnetic techniques, *Surv. Geophys*. **20**, 3-4, 279-308, DOI: 10.1023/A: 1006669218545.

Ward, S.H., and G.W. Hohmann (1988), Electromagnetic theory for geophysical applications. **In:** M.N. Nabighian (ed.), *Electromagnetic Methods in Applied Geophysics, Vol. 1: Theory*, Society of Exploration Geophysicists, Tulsa, 131-311.

Won, I.J., D.A. Keiswetter, G.R.A. Fields, and L.C. Sutton (1996), GEM-2: A new multifrequency electromagnetic sensor, *J. Environ. Eng. Geoph*. **1**, 129-138, DOI: 10.4133/JEEG1.2.129.

Yoon, G.L., and J.B. Park (2001), Sensitivity of leachate and fine contents on electrical resistivity variations of sandy soils, *J. Hazard. Mater*. **84**, 2-3, 147-161, DOI: 10.1016/S0304-3894(01)00197-2.

Zhang, Z., P.S. Routh, D.W. Oldenburg, D.L. Alumbaugh, and G.A. Newman (2000), Reconstruction of 1-D conductivity from dual-loop EM data, *Geophysics* **65**, 492-501, DOI: 10.1190/1.1444743.

Received 3 December 2009
Accepted 1 March 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1056-1071
DOI: 10.2478/s11600-010-0029-2

Fluid Theory with Asymmetric Molecular Stresses: Difference Between Vorticity and Spin Equations

Roman TEISSEYRE

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

A b s t r a c t

We present a new development in fluid theory, incorporating into it the velocity and spin fields; special attention is given to the structure of transport.The theory includes asymmetric molecular stresses and independent rotation velocity, *i.e.*, spin. Our approach is based on our former studies on the asymmetric continuum theory with the balance and constitutive laws for displacement velocity and independent rotation motion, and points out the role of a related characteristic length unit. It is assumed that the vorticity caused by velocities can induce a spin transport counterpart. Thus, under certain conditions, an additional transport term due to rotational velocity fields may be incorporated to the velocity transport, which may lead to the vortex fields included directly into the theory.

The Coriolis effect, important for the vortex processes, is considered and it is demonstrated that the motion equations in our asymmetric theory include this effect automatically. When confinining to 2D case, some compatibilities are found between the relations derived for the rotation motions and the moment formed by the Coriolis forces and applied to such motions. This is an important argument supporting our approach.

The obtained nonlinear vortex equations (solitons) are derived and discussed for a stationary case.

Key words: molecular stresses, vorticity, circular transport, Coriolis effect, asymmetric continuum.

1. INTRODUCTION

We follow the former study on the 10 basic point motions and deformations (three being related to the displacement vector, three to the point rotation vector, one to the axial deformation, and three to the string-string strain vector), as presented by Teisseyre and Górski (2008, 2009), and use the related motion equations, as described by Teisseyre (2009) in the framework of the asymmetric continuum theory. From the physical point of view, the point fields mean those related to the Planck length λ: at this level, rotation can be treated as an independent field that is not related to displacement derivatives. The asymmetric continuum theory (Teisseyre 2009) has some origins in the micomorphic and micropolar theories (see: Eringen and Suhubi 1964, Mindlin 1965, and Eringen 2001), but is more closely related to the Theory of Asymmetric Elasticity (Nowacki 1986) with asymmetric stresses and couple-stresses. The approach adopted is a further simplification of Nowacki's asymmetric theory; the asymmetric stresses and couple stresses are reduced only to the independent symmetric and antisymmetric stresses which are independently related to the displacement and rotation fields (the additional relations introduced by Nowacki, as related to the couple stresses, disappear here; *cf*. Teisseyre 2005).

In our present study we make a similar approach to the mechanics of fluids, applying into it the velocity and independent rotation velocity (spin), as well as molecular symmetric and antisymmetric stresses. We consider the rotation velocity field (spin), treated as the most important factor in vorticity, and we will even consider problems in which transport linear velocities, v_l, remain zero – being replaced by a circular transport.

The classical approach to vorticity and helical motions is based on the Navier-Stokes equation by applying the curl operator to it (see: Saffman 1995); more complicated problems can be treated with the help of microcontinuum theory for fluent media (Eringen 2001). However, it seems that the transport problem might be treated in a more complex way in which the velocity transport can be amended by including a transport related to circular motions. Thus, we present some modification of the Navier–Stokes transport theory and for simplicity we assume that the density and point concentration are constant. In our approach we include the basic balance and constitutive relations for spin (rotation motions independent of velocities).

Many important improvements of the Navier–Stokes transport works include a modification of external forces and studies on the concentration field with advection and diffusion processes, changes in fluid density and point concentration (Kalinowska and Rowiński 2008, and Czernuszenko 1987).

The helical motions and vortex theory have been studied since the XIX century (see, *e.g*., Helmholz 1858, Kelvin 1867, Moore and Saffman 1972, Saffman 1995).

However, in the present study, we propose a quite different approach; we concentrate on a vortex structure considering an internal transport process as based on the spin motion equations. This approach is based on new development in the continuum theory with the asymmetric stresses and phase shift indexes related to the generation processes of displacement and rotation motions in a seismic source when, *e.g*., the breaks of bonds in the microfracture release the rebound rotations (Teisseyre 2008a,b, 2009).

The generalization related to transport motions in fluids leads to a system of non-linear relations for transport velocities; such a transport is caused by the angular moments and spin motions. A comparison with the Coriolis effect is made.

Further, we consider the formation of a micro-vortex dynamical structure, and soliton vortex equations. Thus, the torque moments which could be generated by the antisymmetric molecular stresses will cause the spin motions; such motions, $l\omega$, include a related characteristic length, l, contributing to the transport motion as an additional element to the velocity transport.

2. CONSTITUTIVE RELATIONS

We follow the asymmetric continuum theory developed in our former papers (Teisseyre 2008a,b, 2009); we assume that beside the asymmetric molecular stresses, $\tilde{S}$, there appear also the antisymmetric molecular strain rate, $\tilde{E}$, and spin, ω:

$$\tilde{S}_{kl} = \tilde{S}_{(kl)} + \tilde{S}_{[kl]}, \quad \tilde{E}_{kl}, \quad \omega_{kl}\,, \tag{1}$$

where for any asymmetric tensor, T_{kl}, we define its symmetric and antisymmetric parts as $T_{(kl)}$ and $T_{[kl]}$.

These fields can be related by means of the following constitutive relations:

$$\begin{aligned} &\frac{1}{3}\tilde{S}_{ss} = -p = \frac{1}{3}k\tilde{E}_{ss}\,, \qquad \tilde{S}^{D}_{kl} = \eta\tilde{E}^{D}_{kl}\,, \\ &\tilde{S}^{D}_{kl} = \tilde{S}_{(kl)} - \frac{1}{3}\delta_{kl}\tilde{S}_{ss}\,, \quad \tilde{S}_{[kl]} = \eta\,\omega_{kl} \equiv \eta\,\varepsilon_{kls}\omega_{s}\,, \end{aligned} \tag{2}$$

where $\tilde{S}^{D}_{kl}$ and $\tilde{E}^{D}_{kl}$ mean the deviatoric parts of tensors and η is the viscosity; ε_{kls} means the alternating third-order tensor.

For both constitutive relations in eqs. (2), that is, the molecular symmetric stress-strain and molecular antisymmetric stress-spin relations, we have assumed the same constant η (similar assumption related to rigidity is supported by the similarity between transversal motions and rotations).

In the presented theory, points in the considered fluid continuum have six degrees of freedom: velocity field and spin (rotation velocity).

The molecular strain rate and spin can be related to the derivatives of velocity fields in the following way:

$$\tilde{E}_{kl} = \frac{1}{2}\left(\frac{\partial \upsilon_l}{\partial x_k} + \frac{\partial \upsilon_k}{\partial x_l}\right), \quad \omega_{kl} = \varepsilon_{kls}\omega_s = \frac{1}{2}\left[\frac{\partial \tilde{\upsilon}_l}{\partial x_k} - \frac{\partial \tilde{\upsilon}_k}{\partial x_l}\right], \tag{3}$$

where we distinguish between the velocity field related to molecular strain and to rotation velocity; this follows from the fact that any rotation field can be expressed as a rotation of some velocity field.

We assume an independence between the molecular strain rate and rotation velocity; therefore, the velocity field, $\tilde{\upsilon}_l$, could differ from that considered in the transport relation, υ_l; in some cases the field $\tilde{\upsilon}_l$ may be treated as only a mathematical quantity.

It should also be noted that for $\tilde{\upsilon} = \upsilon$ we obtain the relation between the molecular stresses as a sum of molecular strain rate and spin:

$$\tilde{S}_{kl} = \tilde{S}_{(kl)} + \tilde{S}_{[kl]} = \eta \frac{\partial \upsilon_l}{\partial x_k} = \eta(\tilde{E}_{kl} + \omega_{kl}). \tag{4}$$

3. SYSTEM OF THE NAVIER–STOKES EQUATIONS

Transport processes included in the Navier–Stokes equation at a constant density can be presented symbolically by the following transition from symbol of the partial derivative to substantial derivative:

$$\frac{\partial}{\partial t} \to \frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \upsilon_s \frac{\partial}{\partial x_s} \tag{5a}$$

leading to

$$\rho \frac{\partial \upsilon_i}{\partial t} + \rho \upsilon_s \frac{\partial \upsilon_i}{\partial x_s} = \eta \frac{\partial^2 \upsilon_i}{\partial x_k \partial x_k} - F_i, \tag{6a}$$

where F are the body forces, υ is the velocity field and η is the dynamic viscosity; the pressure gradient, $\partial p / \partial x_i$, is included the force term.

However, using the definitions in eqs. (3), we may present the above relation (5a) in a different form, with the independent fields $\tilde{E}_{ki}$ and ω_{ki}, which will be valid both at $\upsilon' = \upsilon$ and $\upsilon' \neq \upsilon$:

$$\rho \frac{\partial \upsilon_i}{\partial t} + \rho \upsilon_s \frac{\partial \upsilon_i}{\partial x_s} = \eta \frac{\partial}{\partial x_k}\left(\tilde{E}_{ki} + \omega_{ki}\right) - F_i. \tag{6b}$$

This is an equivalent form expressing a possible independence between the molecular strain and rotation fields.

4. VORTICITY AND TRANSPORT PROCESSES

The transport motion is usually related to the velocity field, although in some special cases the independent rotational motions related to angular moment and a characteristic length may contribute to transport phenomena.

Following (5a) and (6a), the transport for the field $\tilde{\upsilon}_k$, where a straight-line transport is replaced by the rotational transport along a curved line, is as follows:

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \tilde{\upsilon}_k \frac{\partial}{\partial x_k} \ . \tag{5b}$$

Let us remind that the classical approach to include rotations was made *via* the definition of vorticity, ς :

$$\varsigma_n = \varepsilon_{npi} \frac{\partial \tilde{\upsilon}_i}{\partial x_p} \ . \tag{7a}$$

Acting with the curl operator on the Navier–Stokes transport equation, we arrive at the relation for vorticity, ς :

$$\rho \frac{\partial \varsigma_k}{\partial t} + \rho \varepsilon_{kni} \frac{\partial}{\partial x_n} \left(\tilde{\upsilon}_s \frac{\partial \tilde{\upsilon}_i}{\partial x_s} \right) = \rho \frac{\partial \varsigma_k}{\partial t} + \rho \upsilon_s \frac{\partial \varsigma_k}{\partial x_s} + \rho \varepsilon_{kni} \frac{\partial \tilde{\upsilon}_s}{\partial x_n} \frac{\partial \tilde{\upsilon}_i}{\partial x_s} = \eta \frac{\partial^2 \varsigma_k}{\partial x_k \partial x_k} - \varepsilon_{kni} \frac{\partial}{\partial x_n} F_i \ ,$$

which may be changed as follows, when assuming the transport motions along the plane, *e.g.*, at $x_3 = \text{constant}$ and $\partial \tilde{\upsilon}_n / \partial x_n = 0$:

$$\varepsilon_{3ni} \frac{\partial \tilde{\upsilon}_s}{\partial x_n} \frac{\partial \tilde{\upsilon}_i}{\partial x_s} = \frac{\partial \tilde{\upsilon}_s}{\partial x_1} \frac{\partial \tilde{\upsilon}_2}{\partial x_s} - \frac{\partial \tilde{\upsilon}_s}{\partial x_2} \frac{\partial \tilde{\upsilon}_1}{\partial x_s} = \frac{\partial \tilde{\upsilon}_1}{\partial x_1} \frac{\partial \tilde{\upsilon}_2}{\partial x_1} - \frac{\partial \tilde{\upsilon}_1}{\partial x_2} \frac{\partial \tilde{\upsilon}_1}{\partial x_1} + \frac{\partial \tilde{\upsilon}_2}{\partial x_1} \frac{\partial \tilde{\upsilon}_2}{\partial x_2} - \frac{\partial \tilde{\upsilon}_2}{\partial x_2} \frac{\partial \tilde{\upsilon}_1}{\partial x_2} \ ,$$

$$\varepsilon_{3ni} \frac{\partial \tilde{\upsilon}_s}{\partial x_n} \frac{\partial \tilde{\upsilon}_i}{\partial x_s} = \left(\frac{\partial \tilde{\upsilon}_2}{\partial x_1} - \frac{\partial \tilde{\upsilon}_1}{\partial x_2} \right) \left(\frac{\partial \tilde{\upsilon}_1}{\partial x_1} + \frac{\partial \tilde{\upsilon}_2}{\partial x_2} \right) = 0 \ , \qquad \text{at} \quad \frac{\partial \tilde{\upsilon}_1}{\partial x_1} + \frac{\partial \tilde{\upsilon}_2}{\partial x_2} = 0$$

leading to

$$\rho \frac{\mathrm{D} \varsigma_k}{\mathrm{D}t} = \rho \frac{\partial \varsigma_k}{\partial t} + \rho \tilde{\upsilon}_s \frac{\partial \varsigma_k}{\partial x_s} = \eta \frac{\partial^2 \varsigma_k}{\partial x_s \partial x_s} - \varepsilon_{kni} \frac{\partial}{\partial x_n} F_i \ , \tag{7b}$$

where we have put $\partial \tilde{\upsilon}_s / \partial x_s = 0$.

Moreover, considering this relation for vorticity we may replace the transport term by the rotation of Coriolis forces related to ς_s and $\tilde{\upsilon}_n$ as a kind of internal damping force:

$$\rho \frac{\mathrm{D} \varsigma_k}{\mathrm{D}t} = \rho \frac{\partial \varsigma_k}{\partial t} - \varepsilon_{kni} \frac{\partial}{\partial x_n} \tilde{F}_i = \eta \frac{\partial^2 \varsigma_k}{\partial x_s \partial x_s} - \varepsilon_{kni} \frac{\partial}{\partial x_n} F_i \ , \tag{8a}$$

where we confine ourselves to the non-compressive case.

We define the Coriolis force and its rotation as follows:

$$\tilde{F}_p = -\rho\varepsilon_{psn}\varsigma_s\tilde{\upsilon}_n = \rho\frac{\partial\tilde{\upsilon}_n}{\partial x_p}\tilde{\upsilon}_n - \rho\frac{\partial\tilde{\upsilon}_p}{\partial x_n}\tilde{\upsilon}_n \ ,$$

$$\varepsilon_{kqp}\ \frac{\partial}{\partial x_q}\tilde{F}_p = \rho\varepsilon_{kqp}\ \frac{\partial}{\partial x_q}\left(\tilde{\upsilon}_n\frac{\partial\tilde{\upsilon}_n}{\partial x_p} - \tilde{\upsilon}_n\frac{\partial\tilde{\upsilon}_p}{\partial x_n}\right),$$

where in this definition the rotation is replaced by vorticity, $\Omega_{ij} \to 1/2\ \varepsilon_{ijs}\varsigma_s$, as shears rotate with the angular velocity $\Omega = 1/2\ \varsigma$ (Saffman 1995).

Hence, with this internal Coriolis rotation, we obtain from (8a) for motions only on the *z* plane:

$$\rho\frac{\partial\varsigma_3}{\partial t} = \eta\frac{\partial^2\varsigma_3}{\partial x_s\partial x_s} - \rho\tilde{\upsilon}_n\frac{\partial\varsigma_3}{\partial x_n} - \varepsilon_{3qp}\ \frac{\partial}{\partial x_q}F_p$$

or

$$\rho\frac{\partial\varsigma_k}{\partial t} = \eta\frac{\partial^2\varsigma_k}{\partial x_s\partial x_s} + \rho\varepsilon_{kpq}\frac{\partial}{\partial x_p}\left(\tilde{\upsilon}_n\frac{\partial\tilde{\upsilon}_n}{\partial x_q}\right) - \rho\varepsilon_{kpq}\frac{\partial\tilde{\upsilon}_n}{\partial x_p}\frac{\partial\tilde{\upsilon}_q}{\partial x_n} - \rho\tilde{\upsilon}_n\frac{\partial\varsigma_k}{\partial x_n} - \varepsilon_{kqp}\ \frac{\partial}{\partial x_q}F_p \ .$$

However, due to symmetry-antisymmetry, the above equation leads to

$$\varepsilon_{kpq}\frac{\partial}{\partial x_p}\left(\tilde{\upsilon}_n\frac{\partial\tilde{\upsilon}_n}{\partial x_q}\right) = \varepsilon_{kpq}\left(\frac{\partial\tilde{\upsilon}_n}{\partial x_p}\frac{\partial\tilde{\upsilon}_n}{\partial x_q} + \tilde{\upsilon}_n\frac{\partial^2\tilde{\upsilon}_n}{\partial x_p\partial x_q}\right) = 0 \ ,$$

and at $\partial\tilde{\upsilon}_n/\partial x_n = 0$ the remaining part disappears; see the derivation of eq. (7b) above

$$\varepsilon_{3pq}\frac{\partial\tilde{\upsilon}_s}{\partial x_p}\frac{\partial\tilde{\upsilon}_q}{\partial x_s} = \left(\frac{\partial\tilde{\upsilon}_2}{\partial x_1} - \frac{\partial\tilde{\upsilon}_1}{\partial x_2}\right)\left(\frac{\partial\tilde{\upsilon}_1}{\partial x_1} + \frac{\partial\tilde{\upsilon}_2}{\partial x_2}\right) = 0 \ , \quad \text{at} \quad \frac{\partial\tilde{\upsilon}_1}{\partial x_1} + \frac{\partial\tilde{\upsilon}_2}{\partial x_2} = 0 \ .$$

We obtain

$$\rho\frac{\partial\varsigma_k}{\partial t} = \eta\frac{\partial^2\varsigma_k}{\partial x_s\partial x_s} - \rho\tilde{\upsilon}_n\frac{\partial\varsigma_k}{\partial x_n} - \varepsilon_{kqp}\ \frac{\partial}{\partial x_q}F_p \ . \tag{8b}$$

Finally, we have obtained the expression in which the total derivative is explained by the internal Coriolis rotation term; in rotational motions on a plane, the Coriolis effect becomes equivalent to the rotational transport. We shall add that when the ratio of absolute values of vorticity to viscosity, $\text{abs}\,\zeta/\eta$, would be greater than some critical value, then one should include

a rotation counterpart to the transport process. Thus, in such a case, in order to describe the vortex motions, we should include into the Navier–Stokes transport equation both the velocity field (this field can be equivalently transformed into the molecular shear oscillations; *cf.* Teisseyre 2008a) and the moment-related velocities caused by the spin motions with the characteristic Cosserat length or even macroscopic arms.

Instead of the Newton law for velocities and the balance relation for the rotation motions, we can rely on the relations for the symmetric molecular stresses and strains and for the antisymmetric molecular stresses and spin.

Thus, in the double transport theory (Teisseyre 2008b) we consider two fields: velocity and the spin-related transport which can lead to more complicated vortex dynamics and nonlinear relations.

In some problems, a macroscopic transport field, υ, can be treated as an imposed external field, while the internal transport rotation will enter into the additional terms introduced to the system of equations. Thus, in motions with advanced vorticity dynamics we assume that a transport can be related both to velocity, υ_k, and to vortex motions, $\tilde{\upsilon}_k$; we follow the approach applied to fragmentation and slip in the fracture processes (Teisseyre 2009). Note that such a flow can occur effectively for the Reynolds numbers above the critical value.

The double transport process with velocities and spin motions generate the micro-vortices; a kind of dynamic vortex structure can be formed, simultaneously undergoing a velocity-related transport process. Including these transport processes, the motion equations become after (6):

$$\rho\frac{\partial \upsilon_i}{\partial t}+\rho\upsilon_s\frac{\partial \upsilon_i}{\partial x_s}+\tilde{\upsilon}_k\frac{\partial \upsilon_i}{\partial x_k}=\eta\frac{\partial^2\upsilon_i}{\partial x_k\partial x_k}-F_i=\eta\frac{\partial}{\partial x_k}\left(\tilde{E}_{ki}+\omega_{ki}\right)-F_i\ . \tag{9}$$

Under some special conditions, an isolated vortex center can be formed; inside it, the velocity transport might be negligible. However, to consider formally the transport related to such extreme conditions we shall return to the basic problem how to incorporate the spin motion into the system of basic relations.

5. MOTION EQUATIONS FOR THE VORTEX ROTATION FIELD

To consider the molecular stress moment balance we shall first define the velocity fields related to spin. The balance equation for the antisymmetric stresses can be obtained for the balance of divergence of the molecular stress moment, similarly to that derived in the asymmetric continuum theory for solids (Teisseyre 2008a, 2009); with points having six degrees of freedom, we consider the balance equation for spin:

$$\tilde{M}_{lk} = l^2 \varepsilon_{lki} \frac{\partial}{\partial x_n} \tilde{S}_{[ni]} = \eta l^2 \varepsilon_{lki} \frac{\partial \omega_{ni}}{\partial x_n} , \qquad \tilde{S}_{[ni]} = \eta \omega_{ni} ,$$
$$\frac{\partial \tilde{M}_{pk}}{\partial x_k} = l^2 \varepsilon_{pki} \frac{\partial^2 \tilde{S}_{[ni]}}{\partial x_k \partial x_n} = l^2 \varepsilon_{pki} \frac{\partial^2 \tilde{S}_{[ki]}}{\partial x_n \partial x_n} = l^2 \rho \varepsilon_{pki} \frac{\partial^2 \omega_{ki}}{\partial t^2} + \varepsilon_{pki} \tilde{K}_{[ki]} , \tag{10a}$$

where the rotation velocity (spin), $\omega_s = 1/2 \ \varepsilon_{ski} \omega_{ki}$, is introduced, and l is the Cosserat length enabling us to formulate the action of molecular stresses and moments on the fluid elements, and $\tilde{K}_p = \varepsilon_{pki} \tilde{K}_{[ki]}$ means the external moment.

Further, we can write

$$\frac{\partial \tilde{M}_{pk}}{\partial x_k} = \varepsilon_{pks} \eta l^2 \frac{\partial^2 \omega_{ks}}{\partial x_n \partial x_n} = \varepsilon_{pks} \rho l^2 \frac{\partial \omega_{ks}}{\partial t} + \varepsilon_{pks} \tilde{K}_{[ks]} ,$$

or

$$\eta l^2 \frac{\partial^2 \omega_p}{\partial x_n \partial x_n} = \rho l^2 \frac{\partial^2 \omega_p}{\partial t^2} + \tilde{K}_{[p]} , \tag{10b}$$

or

$$\eta l^2 \frac{\partial^2 \omega_{ns}}{\partial x_k \partial x_n} = \rho l^2 \frac{\partial^2 \omega_{ks}}{\partial t^2} + \tilde{K}_{[ks]} , \tag{10c}$$

where we may note the important equivalence for any anti-symmetric field fulfilling the non-source field condition $\frac{1}{2} \varepsilon_{pki} \frac{\partial \omega_{ki}}{\partial x_p} = \frac{\partial \omega_p}{\partial x_p} = 0$:

$$\varepsilon_{pki} \frac{\partial^2 \omega_{ni}}{\partial x_k \partial x_n} = \varepsilon_{pki} \frac{\partial^2 \omega_{ki}}{\partial x_n \partial x_n} .$$

We may define now an average rotation over some group of neighbouring rotations considered in an element related to the characteristic length:

$$\eta \frac{\partial^2 \omega_{ns}}{\partial x_k \partial x_n} = \rho \frac{\partial^2 \omega_{ks}}{\partial t^2} \rightarrow \eta \frac{\partial^2 \Omega_{ns}}{\partial x_k \partial x_n} = \rho \frac{\partial^2 \Omega_{ks}}{\partial t^2}, \quad \text{at } \Omega_k = l \frac{\partial \omega_{nk}}{\partial x_n} , \tag{11}$$

where η and ρ are constants.

Then, we obtain

$$\eta l^2 \frac{\partial^2 \Omega_s}{\partial x_n \partial x_n} = \rho l^2 \frac{\partial^2 \Omega_s}{\partial t^2} + K_s , \tag{12a}$$

where $\Omega_s = 1/2 \ \varepsilon_{ski} \Omega_{ki}$. This equation is quite similar to that for spin (10c), so the question may arise why we intend to introduce this quite identical relation, the only difference being a slightly different definition of rotation tensors?

There are two reasons: the one, that the field ω may be reduced to a pure spin (point rotation *sensu stricto)*, and the other is that when we relate the field ω with the Cosserat characteristic length (*e.g.*, characteristic dimension of molecules), then the amplitudes of these spin motions along the rotation path should be averaged to obtain the observed rotations (*cf.* Fig. 1 right part, in which rotations relate to the groups of points). Moreover, while considering the field ω, some difficulties arise how to attribute the point rotation to a vortex structure with dimension r between the characteristic length l and the considered arm of vortex L such a structure as a whole could be attributed to the field Ω, while rotation motions at each of the points, of course, will be described by the field ω. *Vice versa*, with $\Omega_k = l\,\partial\omega_{nk}/\partial x_n = l\,\varepsilon_{nks}\,\partial\omega_s/\partial x_n$, we return to eq. (10c).

We should keep in mind that the obtained relations relate to the fact that each point-element in our approach has six degrees of freedom – compare Fig. 1. However, to the rotation field thus defined, Ω_s , we may attach, instead of the characteristic Cosserat length l, the variable vortex arm L; in that case, instead of (12a), we may have (see further on):

$$\eta L^2 \frac{\partial^2 \Omega_s}{\partial x_n \partial x_n} = \rho L^2 \frac{\partial^2 \Omega_s}{\partial t^2} + K_s \,. \tag{12b}$$

Further, we continue our analysis using the cylindrical system of coordinates.The rotation vector, Ω, related to the spin is given by the following expression:

$$\Omega_s = l\left\{\left(\frac{\partial\omega_\varphi}{\partial z} - \frac{1}{r}\frac{\partial\omega_z}{\partial\varphi}\right), \left(\frac{\partial\omega_z}{\partial r} - \frac{\partial\omega_r}{\partial z}\right), \left(\frac{\partial\omega_r}{r\partial\varphi} - \frac{\partial(r\omega_\varphi)}{r\,\partial r}\right)\right\}, \tag{13}$$

where $s = \{r, \varphi, z\}$.

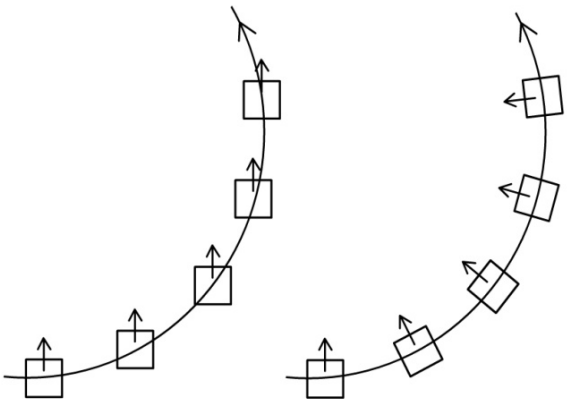


Fig. 1. The transport trajectories with points having three degrees of freedom marked by the arrow (left part) and six degrees (right part) of freedom (additional rotation included).

6. CIRCLE ARC TRANSPORT AND VORTEX MOTION

The transport field related to the vortex motion can be defined basing on the angular molecular moment field related to rotations Ω. Basing on the definition of macroscopic rotation (11) we can define the angular velocity with the help of a vector of the effective arm L_k ; we put

$$\tilde{\upsilon}_p = \varepsilon_{pks}\Omega_k L_s \ . \tag{14}$$

The related circle arc transport, by analogy to (5), can be presented symbolically, with the help of this macroscopic arcuate velocity field, $\tilde{\upsilon}$, by the D/Dt operator given as follows:

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \tilde{\upsilon}_k \frac{\partial}{\partial x_k} \ . \tag{15}$$

Thus, we will focus our analysis on the vortex motions with the vortices oriented along the z-axis; in this case, the arm vector may be time dependent in a horizontal plane. We assume that on the plane $z = \text{const}$ we are dealing with some variable characteristic length, L, related to a possible vorticity; however, along the z-axis assumed as a small one:

$$L(t) >> l \ , \quad L_z = l \ . \tag{16}$$

Accordingly, for the arcuate velocity (14) we obtain:

$$\tilde{\upsilon}_r = l\Omega_\varphi - L\Omega_z \ , \quad \tilde{\upsilon}_\varphi = L\Omega_z - l\Omega_r \ , \quad \tilde{\upsilon}_z = L\Omega_r - L\Omega_\varphi \ ,$$

and with the condition (16):

$$\tilde{\upsilon}_r \approx -L\Omega_z \ , \quad \tilde{\upsilon}_\varphi \approx L\Omega_z \ , \quad \tilde{\upsilon}_z \approx L\left(\Omega_r - \Omega_\varphi\right) \ . \tag{17}$$

When considering only a rotational transport, the Navier–Stokes eq. (9) becomes modified ($\tilde{E}_{ki} \approx 0$) :

$$\rho\frac{\partial\tilde{\upsilon}_i}{\partial t} + \rho\tilde{\upsilon}_k\frac{\partial\tilde{\upsilon}_i}{\partial x_k} = \eta\frac{\partial\omega_{ki}}{\partial x_k} \quad \text{or} \quad \rho\frac{\partial\tilde{\upsilon}_i}{\partial t} + \rho\tilde{\upsilon}_k\frac{\partial\tilde{\upsilon}_i}{\partial x_k} = \eta\frac{1}{l}\Omega_i \ , \tag{18}$$

where in the second relation we have used the definition shown in eq. (11). Using (9) and (14) we will have the following motion equation:

$$\rho L_s\frac{\partial\Omega_{si}}{\partial t} + \rho\Omega_{qk}L_sL_q\frac{\partial\Omega_{si}}{\partial x_k} = \eta\varepsilon_{spi}\frac{1}{2l}\Omega_{sp} \ . \tag{19}$$

Using (17) we obtain directly from (14) and (18) or directly from (19) in the cylindrical coordinates:

$$\rho\frac{\partial \Omega_z}{L\partial t}+\rho\left(-\Omega_z\frac{\partial \Omega_z}{\partial r}+\Omega_z\frac{\partial \Omega_z}{r\partial\varphi}+\left(\Omega_r-\Omega_\varphi\right)\frac{\partial \Omega_z}{\partial z}\right)=-\eta\frac{1}{L^2 l}\Omega_r\,,$$

$$\rho\frac{\partial \Omega_z}{L\partial t}+\rho\left(-\Omega_z\frac{\partial \Omega_z}{\partial r}+\Omega_z\frac{\partial \Omega_z}{r\partial\varphi}+\left(\Omega_r-\Omega_\varphi\right)\frac{\partial \Omega_z}{\partial z}\right)=\eta\frac{1}{L^2 l}\Omega_\varphi\,,$$

hence we shall put $\Omega_r=-\Omega_\varphi$ $(\tilde{\upsilon}_z=2L\Omega_r)$ and for these two relations we obtain:

$$\rho\frac{\partial \Omega_z}{L\partial t}+\rho\left(-\Omega_z\frac{\partial \Omega_z}{\partial r}+\Omega_z\frac{\partial \Omega_z}{r\partial\varphi}+2\Omega_r\frac{\partial \Omega_z}{\partial z}\right)=-\eta\frac{1}{L^2 l}\Omega_r\,,\quad \Omega_r=-\Omega_\varphi \tag{20a}$$

and the last relation becomes

$$\rho\frac{\partial \Omega_r}{L\partial t}+\rho\left(-\Omega_z\frac{\partial \Omega_r}{\partial r}+\Omega_z\frac{\partial \Omega_r}{r\partial\varphi}+2\Omega_r\frac{\partial \Omega_r}{\partial z}\right)=\eta\frac{1}{2L^2 l}\Omega_z\,. \tag{20b}$$

We have obtained the direct relations, eqs. (20), describing the circular transport; the vorticity phenomena become described by the autonomous equations.

We should remember that for the vortices oriented along the z-axis, we have introduced here the macroscopic vortex arm, L; however, for the remaining components we confine ourselves to the microscopic vortex arm, l. We might try to solve this system, but it seems more reasonable to adopt further assumptions related to the macroscopic vortex motion.

However, let us first compare the relations just derived with the Coriolis effect applied to the rotation motion.

7. COMPARISON BETWEEN THE CORIOLIS EFFECT AND EXACT MOTION EQUATIONS

We may introduce to the motion equations for fluids, (19) or (20), an influence of inner Coriolis forces as an external moment.

However, we shall first simplify the obtained motion equations by considering the motion in horizontal direction only. Thus, for the rotation field, Ω_z, we introduce the equal rotation arms, L, in the horizontal plane:

$$\Omega_q\rightarrow\{0,\,0,\,\Omega_z\} \tag{21}$$

and from (20a) and (21) we get at z = const:

$$\rho\frac{\partial \Omega_z}{\partial t}=\rho L\left(\Omega_z\frac{\partial \Omega_z}{\partial r}-\Omega_z\frac{\partial \Omega_z}{r\partial\varphi}\right)-M_z\,, \tag{22a}$$

where we have included moment, M_z, around the z-axis.

Assuming for the plane problem that the angular changes shall be small in comparison with that along an arm, $\frac{\partial \Omega_z}{\partial r} >> \frac{\partial \Omega_z}{r\,\partial \varphi}$, and with $\Omega_r \approx 0$, we write

$$\rho \frac{\partial \Omega_z}{\partial t} = \rho L \Omega_z \frac{\partial \Omega_z}{\partial r} - M_z \,. \tag{22b}$$

Now we consider a compatibility between the rotation motions after our theory (eq. 22b) with those obtained taking into account the influence of Coriolis effects that may act on the rotation motion. However, we must be aware of the different definitions of rotations, as introduced in our theory, and in the Coriolis approach. In the Coriolis forces there appear some rotation field $\bar{\Omega}_z$ and the independent rotational transport along the plane $z = 0$: $\tilde{\upsilon} = \tilde{L} \times \tilde{\Omega}_z$; further, this transport field will be considered with some stationary rotation, $\tilde{\Omega}_z(r)$, an arm $\tilde{L}$, and at a constant density. For the Coriolis forces we write at $\partial \Omega_z / \partial t = 0$:

$$\frac{1}{2\rho} F_p = -\varepsilon_{psn} \bar{\Omega}_s \tilde{\upsilon}_n \,, \quad \frac{F_1}{2\rho} = \tilde{\upsilon}_2 \bar{\Omega}_z \,, \quad \frac{F_2}{2\rho} = -\tilde{\upsilon}_1 \bar{\Omega}_3 \,, \tag{23}$$

and we define the moment of Coriolis forces acting inside the fluid:

$$M_z^C = \frac{1}{2} \rho \in_{3ks} L_k F_s = -\rho \tilde{L} \tilde{\Omega}_z \bar{\Omega}_z \,, \quad \text{at} \quad \tilde{\upsilon}_1 \approx \tilde{\upsilon}_2 \approx \tilde{L} \tilde{\Omega}_3 \,. \tag{24}$$

The required compatibility will appear in the form of proportionality between this Coriolis moment and that following from our theory:

$$M_3^C \propto M_z \,. \tag{25}$$

We consider a stationary case on the plane $z = 0$, under the assumptions:

$$\Omega_z \propto \frac{1}{r} \;\leftrightarrow\; \Omega_z \frac{\partial \Omega_z}{\partial r} \propto \frac{1}{r^3} \,. \tag{26}$$

With this solution and the required compatibility (25) we shall demand that, when comparing eqs. (22b) and (24), the following relation holds:

$$-\tilde{L} \tilde{\Omega}_z \bar{\Omega}_z \propto L \Omega_z \frac{\partial \Omega_z}{\partial r} \propto -\frac{L}{r^3} \,. \tag{27}$$

This proportionality relation can be fulfilled when assuming that there exists a constant rotation acting on all fields on this plane:

$$\bar{\Omega}_z \propto \text{constant} \tag{28}$$

and for the fields $\tilde{L}$, $\tilde{\Omega}_3$ we may assume the following expressions which will fulfill the conditions (26), (27) and (28) given above:

$$\tilde{L}\tilde{\Omega}_z \propto \frac{L}{r^3} \quad \text{with} \quad \tilde{\Omega}_z \propto \frac{L}{r^2} \quad \text{and} \quad \tilde{L} \propto \frac{1}{r}. \tag{29}$$

This stationarity condition assures compatibility between the derived equations in our vortex theory and the Coriolis effects acting on rotational motions; here, the rotation arcs are formed by the constant interactive velocity at a distance r from that rotation center

$$V_r \propto r\tilde{L}\bar{\Omega}_z \propto \bar{\Omega}_z \ . \tag{30}$$

This means that we will arrive at the situation similar to a straight stream flow (Fig. 2). However, at the same time the rotation velocity field, $\tilde{\Omega}_z$, shall decrease more rapidly with growing distance than the field Ω_z (*cf*. eq. 26).

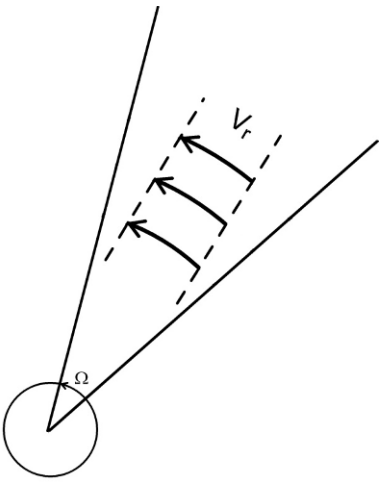


Fig. 2. Visualision of the constant interactive velocity V_r at a distance r, related to the condition joining the Coriolis effect applied to rotation motions and the results from the vortex motion theory on a plane.

8. VORTEX STRUCTURES: LINEAR STRUCTURE AND STATIONARY SOLUTION

We return to the derived system of equations (20a) and (20b); for the micro- and macroscopic vortex structures thus formed, we can easily find some stationary conditions.

First, we assume according to (16), that there may appear a variable arm, $L = L(t)$, related to some structure formed, which may be much greater that the constant Cosserat length l ($l \ll L$). Further, we consider one linear statio-

nary vortex structure around the z axis, Ω_z, (Fig. 3) with a constant torque moment around z, and, moreover, we assume that the rotation motion components, Ω_r and $\Omega\varphi$, are much smaller.

The constancy of rotation moment, M_z, may be supported by some external conditions. For a constant moment, M_z, and a stationary relation for Ω_z, we come, after (22b), to the expression for the vortex arm (with M^0 being a constance):

$$M_z \approx \frac{\tilde{L}}{r^3} \;\rightarrow\; \Omega_z \approx \frac{M^0}{\sqrt{r^2+z^2}}. \tag{31}$$

For a stationary case we obtain for the arm length $\tilde{L}$:

$$\tilde{L} \approx \frac{M_z}{\rho}\left(\Omega_z \frac{\partial \Omega_z}{\partial r}\right)^{-1}, \qquad \tilde{L} \approx \frac{1}{M_z \rho}\frac{\left(r^2+z^2\right)^2}{r}. \tag{32}$$

The arm, $\tilde{L}$, for a constant z increases like r^3, while for a constant r an increase with z becomes even more sharp, by one order of magnitude: z^4.

In Fig. 4 we present functions $\tilde{L}_n = \tilde{L}(z; r_n)$ obtained according to (32) for the different boundary conditions $\tilde{L} = r_n$, at plane z_0, these plots are to be compared with our visualization in Fig. 2.

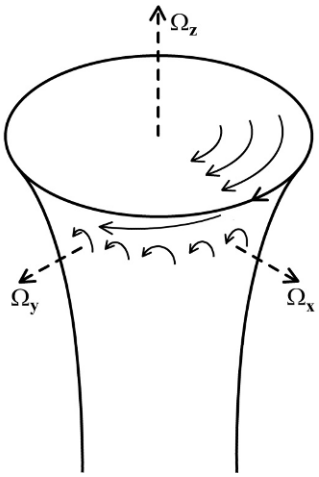


Fig. 3. Vortex structure and macroscopic rotation with a variable arm: $L(r, \varphi, z, t)$.

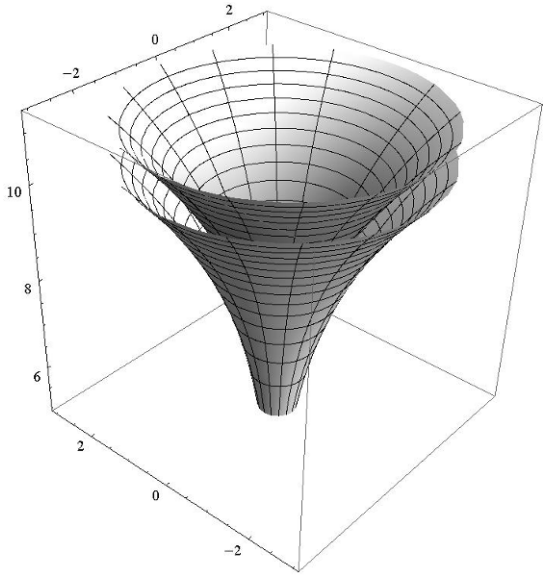


Fig. 4. The examples of the dependence of the vortex arm, L, on z.

9. CONCLUSION

We have applied some results of the asymmetric continuum theory (Teisseyre 2008a, 2009) to the processes in fluids with the transport phenomena produced by the velocity field related to the molecular rotation moments.

We have obtained the nonlinear relations describing the complicated vortex structures and related interaction in transport motion; the vorticity phenomena become described by autonomous equations.

We have proved that there exists some compatibility between the derived equations of rotation motions and the moment formed by the Coriolis forces, as applied to such motions.

We have presented the solution for a stationary vortex structure.

Finally, we have discussed a very simple case of the macroscopic stationary vortex structure; this would be a first step to study a dynamic case.

Acknowledgement. Special thanks are due to Dr. Zbigniew Czechowski and to the reviewers for their important and critical remarks, while for the help in preparation of the manuscript and figures the author sincerely thanks Ms Anna Dziembowska and Dr. Marek Górski.

References

Cosserat, E., and F. Cosserat (1909), *Theorie des Corps Déformables*, Librairie Scientifique A. Hermann et Fils, Paris.

Czernuszenko, W. (1987), Dispersion of pollutants in rivers, *Hydrol. Sci. J.* **32**, 1, 59-67, DOI: 10.1080/02626668709491162.

Eringen, A.C. (2001), *Microcontinuum Field Theories. II: Fluent Media*, Springer, Berlin, 340 pp.

Eringen, A.C, and E.S. Suhubi (1964), Nonlinear theory of simple micro-elastic solids I, *Int. J. Eng. Sci*. **2**, 2, 189-203, DOI: 10.1016/0020-7225(64)90004-7.

Helmholtz, H. (1858), Über Integrale der hydrodynamischen Gleichungen, welche den Wirbelbewegungen entsprechen, *Crelle's J.* **1858**, 55, 25-55, DOI: 10.1515/crll.1858.55.25.

Kalinowska, M.B., and P.M. Rowiński (2008), Numerical solutions of two-dimensional mass transport equation in flowing surface waters, *Publs. Inst. Geophys. Pol. Acad. Sc*. **E-8**, 404, 200 pp.

Kelvin, L. (1867), The translatory velocity of a circular vortex ring, *Phil. Mag*. **33**, 511-512.

Mindlin, R.D. (1965), On the equations of elastic materials with microstructure, *Int. J. Solid Struct*. **1**, 1, 73.

Moore, D.W., and P.G. Saffman (1972), The motion of a vortex filament with axial flow, *Phil. Trans. Roy. Soc. Lond.* **A 272**, 403-429.

Nowacki, W. (1986), *Theory of Asymmetric Elasticity*, PWN – Warszawa, Pergamon Press – Oxford, 383 pp.

Saffman, P.G. (1995), *Vortex Dynamics*, Cambridge University Press, 311 pp.

Teisseyre, R. (2005), Asymmetric continuum mechanics: Deviations from elasticity and symmetry, *Acta Geophys. Pol.* **53**, 115-126.

Teisseyre, R. (2008a), Introduction to asymmetric continuum: Dislocations in solids and extreme phenomena in fluids, *Acta Geophys.* **56**, 2, 259-269, DOI: 10.2478/s11600-008-0010-5.

Teisseyre, R. (2008b), Asymmetric continuum: Standard theory. **In:** R. Teisseyre, H. Nagahama, and E. Majewski (eds.), *Physics of Asymmetric Continua: Extreme and Fracture Processes*, Springer, 95-109.

Teisseyre, R. (2009), Tutorial on new developments in the physics of rotational motions, *Bull. Seism. Soc. Am.* **99**, 2B, 1028-1039, DOI: 10.1785/ 0120080089.

Teisseyre, R., and M. Górski (2008), Introduction to asymmetric continuum: Fundamental point deformations. **In:** R. Teisseyre, H. Nagahama, and E. Majewski (eds.), *Physics of Asymmetric Continua: Extreme and Fracture Processes*, Springer, 3-16.

Teisseyre, R., and M. Górski (2009), Fundamental deformations in asymmetric continuum: Motions and fracturing, *Bull. Seism. Soc. Am.* **99**, 2B, 1132-1136, DOI: 10.1785/0120080091.

Received 23 October 2009
Received in revised form 28 April 2010
Accepted 7 May 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1072-1093
DOI: 10.2478/s11600-010-0020-y

Statistics and Characteristic Scales for Bed Load in a Channel Flow with Sidewall Effects

Alessio RADICE[1], Francesco BALLIO[1], and Vladimir NIKORA[2]

[1]Politecnico di Milano, Dipartimento I.I.A.R., Milan, Italy
e-mails: alessio.radice@polimi.it, francesco.ballio@polimi.it

[2]University of Aberdeen, School of Engineering, Aberdeen, Scotland, UK
e-mail: v.nikora@abdn.ac.uk

Abstract

Results are presented for a long-duration sediment transport experiment with a plane bed. High-frequency time series of sediment concentration, sediment velocity and solid discharge per unit width were measured using image analysis of video records. A range of transport intensities from a single experiment has been analysed considering transport at different distances from the sidewalls as an independent variable. First and second order temporal statistics of the spatially-averaged values of the transport parameters are presented, which are in agreement with previous studies. A refined statistical analysis of the sediment concentration dynamics, reflecting bed load process, is also given. The characteristic scales of the sediment concentration dynamics are evaluated and analyzed in conjunction with those of the near-bed and bulk flow fields.

Key words: sediment transport, particle concentration, particle velocity, sediment transport rate, double-average, spatial-average, characteristic scales for bed load.

1. INTRODUCTION

Research on sediment transport in river hydraulics has been ongoing for many years, aiming to provide reliable tools for quantitative prediction of sediment fluxes in natural streams and constructed waterways for given flow conditions. Pioneering studies on this topic were made, for example, by

Shields (1936), Einstein (1937, 1950) and Meyer-Peter and Müller (1948). Several follow-up studies have been conducted in the following decades. Reviews of the presently available bed-load formulae can be found in many fluvial hydraulics books, such as, e.g., Chanson (1999). Despite the great deal of work done, the prediction of sediment transport rates in rivers is still associated with large uncertainty, as documented by Gomez and Church (1989) or Martin (2003).

During the recent decades, research on sediment transport has moved towards studies of the small-scale processes of particle entrainment and motion (e.g., Drake *et al.* 1988, Nelson *et al.* 1995, Niño and Garcìa 1998a, b, Nikora *et al.* 2002, Ancey *et al.* 2008). Although such an approach typically does not immediately lead to ready-to-use upgrades for sediment transport formulae, it increases understanding of the physical mechanisms governing solid transport and creates a rigorous foundation for development of practical relationships. Recent developments in measurement instrumentation and techniques open fresh perspectives for advancing this field. For example, the pioneering studies of Fernandez Luque and van Beek (1976) and of Drake *et al.* (1988) were largely based on manually-done measurements, with a huge amount of time-consuming work. In the subsequent years, significant effort has been made in developing automatic tools for measuring sediment motion, in many cases by means of image processing techniques (e.g., Lee and Hsu 1994, Pilotti *et al.* 1997, Papanicolaou *et al.* 1999, Böhm *et al.* 2006, Radice *et al.* 2006).

In this paper we present results obtained from a long-duration experiment with high-frequency measurements of bed-load particle motion. The paper continues a research line of Radice and Ballio (2008) and Radice *et al.* (2009), partly reviewing and corroborating previous results as well as presenting new findings. Radice and Ballio (2008) analyzed the double-average features of sediment motion, quantifying the relationships between sediment transport rate, sediment concentration and sediment velocity, as well as links between the sediment transport rate and the mean and fluctuating values of sediment concentration and velocity. Yet, the reliability of reported results was not high, for lowest transport rates, because of relatively short duration of the available time series. In the present paper, we analyse much more reliable data on sediment concentration, based on long-term measurements of sediment motion. Also, the analysis made by Radice *et al.* (2009) was limited to a single sediment transport intensity, while in this paper we explore a range of transport conditions to quantify variability in the sediment concentration dynamics. Furthermore, we make an extensive analysis of the characteristic time scales of the concentration fluctuations, obtained by means of several statistical tools. Finally, we compare the temporal scales of

sediment motion with those estimated for the near-bed and bulk flow field. Radice *et al.* (2009) focused on the statistical properties of sediment concentration, since earlier results (Radice and Ballio 2008) and some literature indications (e.g., Ancey *et al.* 2008) suggest that the statistics of sediment concentration may largely represent those of the sediment transport rate. In the study reported here we follow the same assumption.

2. THE BED-LOAD EXPERIMENT

The experiment for this study was conducted at the Hydraulics Laboratory of the Politecnico di Milano, Italy, and was described in Radice *et al.* (2009). The same experimental installation as in Radice and Ballio (2008) was used. Here we describe the experimental setup only briefly, as additional details can be found in the above-mentioned papers.

A straight 5.8 m long rectangular duct was used, with width and depth equal to 0.40 and 0.16 m, respectively. In a 2 m long working section in the middle of the duct, bed sediments were movable whereas in the remaining parts of the duct bed they were fixed (glued) to the bottom. The sediments were simulated with uniform PVC circular cylinders. The latter were used instead of spheres because this is the shape in which PVC particles are typically produced as semi-manufactured materials for several industrial applications. The particle shape should influence only weakly the results documented in the following. A characteristic size of the grains was however determined as the diameter of a sphere having the same volume as a PVC particle; this size equals to 3.6 mm and in the following will be referred to as d. The specific gravity of the PVC circular cylinders was $\Delta = (\rho_g - \rho)/\rho = 0.43$, with ρ_g and ρ being the densities of particles and water, respectively.

The channel was completely covered by a transparent lid creating a duct with a pressurized flow that guaranteed high quality videoing of particle motion on the bed. Lau and Krishnappan (1985) and Smith and Ettema (1997) found that sediment transport in covered flows can be successfully modelled using the typical tools for free-surface flows and therefore our results should be equally applicable to free-surface and covered (duct-type) flows.

The threshold water discharge, corresponding to the beginning of particle motion on the bed, was estimated by Radice and Ballio (2008) to be equal to 18.95 l/s. The experiment reported here was conducted at flow rate of 21 l/s, i.e., just above the threshold value. The sediment motion was filmed from the top using a CCD capable of 25 fps, and the duration of the film was 20 min. The focus area on the bed was 20×28 cm^2 (see Fig. 1, where also the Cartesian system of reference used is depicted).

The sediment motion was measured using the technique proposed by Radice *et al.* (2006). Their methodology allows measuring the concentration

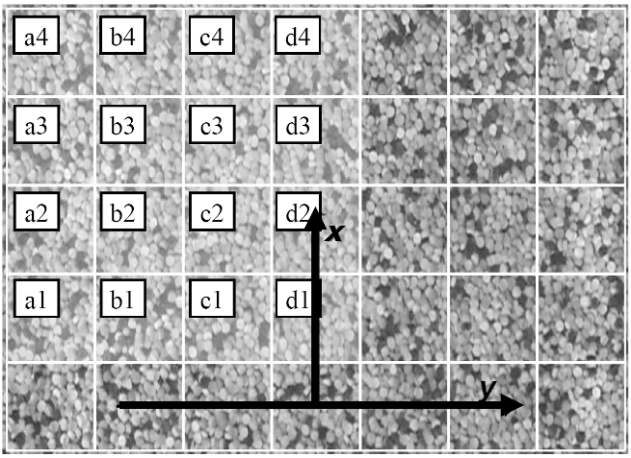


Fig. 1. Sketch of the bed area under investigation. Water flow is from bottom to top. The image also presents the reference system and spatial averaging areas of 4×4 cm in size. The brightened portion of the image is analyzed in the following figures; the corresponding spatial averaging areas are labelled for future reference. Cells a_i, b_i, c_i, and d_i correspond to $y = -12, -8, -4$ cm, and 0, respectively.

and the velocity of the particles in motion at a certain instant, with both measurements being referred to an ***interrogation area*** (a term borrowed from the literature on Particle Image Velocimetry, PIV). In the following, we will consider spatially-averaged quantities and therefore we will use an alternative term of ***spatial-averaging area***. The bed sediment concentration is defined as $\langle C \rangle = N V_g / (A d)$, where N is the number of moving grains and V_g is the volume of a single particle, A is the spatial-averaging area and the angular brackets indicate spatial averaging. The moving particles between two consecutive frames are identified through image difference and filtering. The sediment velocity is measured using a cross-correlation PIV technique and therefore an assumption of the measured velocity values being spatial averages is not fully rigorous. Indeed, the PIV measures the most probable displacement of a group of particles instead of the mean velocity of individual particles involved in motion. The bias, induced by PIV, generally decreases with increase in sampling frequency (e.g., Forliti *et al.* 2000). Radice *et al.* (2006) showed that sampling rates similar to those used here showed that sampling rates similar to those used here resolve a local range of sediment motion as defined by Nikora *et al.* (2002) and therefore PIV limitations will not influence significantly the results presented below.

The time series of the measured sediment concentration are shown in Fig. 2 for the locations brightened and labelled in Fig. 1; similar plots are obtained for the particle velocity. A significant sidewall effect is evident from

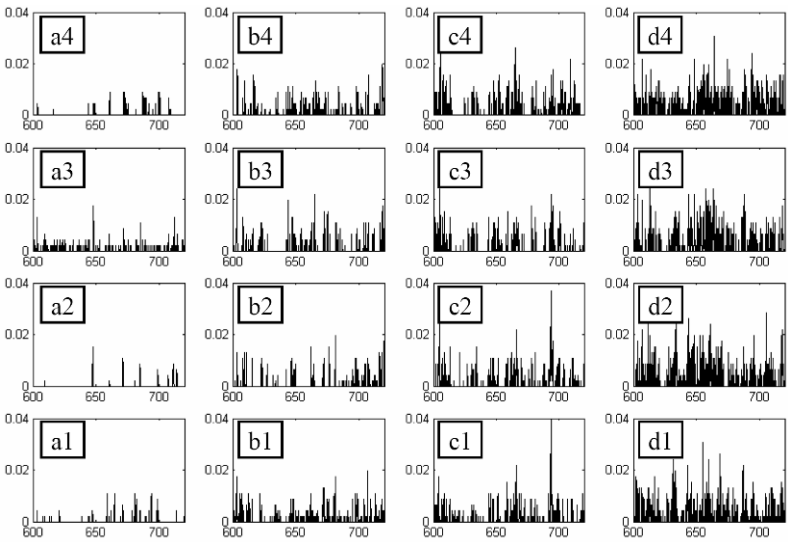


Fig. 2. Time series of sediment concentration over a 2 min period. On the horizontal axes: time (in s); on the vertical axes: concentration $\langle C \rangle$. The plots correspond to the locations brightened in Fig. 1 and are labelled in the same way.

the plots, from which it is clear that, on average, values of the concentration are larger for positions closer to the duct centreline. Moreover, as the duct wall is approached, the concentration time series include a larger number of zero values (identifying a no-motion state at that position for a given instant). Radice and Ballio (2008) defined an intermittency coefficient $\overline{\langle I \rangle}$ as a temporal average of a delta function that assumes 1 when no-motion state is detected and 0 otherwise, and showed that the intermittency coefficient $\overline{\langle I \rangle}$ increases when the transport intensity lowers. Clearly, in Fig. 2 the coefficient $\overline{\langle I \rangle}$ increases from the centreline of the channel towards the sidewalls. A certain similarity can be qualitatively observed between the series for the same *y* and different *x* positions. This will be analyzed in Sub-section 4.2.

The spatially-averaged, vector solid discharge per unit width is defined, similar to other studies (e.g., Niño and Garcìa 1998b, Parker *et al.* 2003), as $\langle \boldsymbol{q}_s \rangle = \langle C \rangle \langle \boldsymbol{v}_s \rangle d$, where $\langle \boldsymbol{v}_s \rangle$ is the particle velocity (vector quantities are indicated with bold symbols). Both particle velocity and solid discharge may be made non-dimensional using the sediment size and density: the dimensionless longitudinal sediment velocity is defined as $\langle F \rangle = \langle u_s \rangle / (g \Delta d)^{0.5}$, where $\langle u_s \rangle$ is the streamwise component of sediment velocity and *g* is grav-

ity. The dimensionless longitudinal sediment transport rate per unit width is defined as $\langle \Phi \rangle = \langle q_s \rangle / \sqrt{g \Delta d^3}$ (where $\langle q_s \rangle$ is the streamwise component of solid discharge), which results in $\langle \Phi \rangle = \langle C \rangle \langle F \rangle$. The data collection described above was supplemented with velocity measurements at the duct centreline using an Ultrasonic Doppler Profiler. Full details are given in Radice *et al.* (2009).

3. THE DOUBLE-AVERAGED PARAMETERS OF SEDIMENT MOTION

In this section, the double-averaged values of the measured quantities are considered. Since the instantaneous values measured are spatially-averaged, the double-averaged statistics are obtained by time-averaging the instantaneous spatially-averaged quantities, similar to Nikora *et al.* (2007).

Figure 3 depicts the first and second order statistics of particle motion parameters. In general, the quantities $\overline{\langle C \rangle}$, $\overline{\langle F \rangle}$ and $\overline{\langle \Phi \rangle}$ reveal similar behaviour, with highest transport activity at the duct centreline and vanishing transport near the sidewalls, mainly occurring due to the sidewall effects on flow dynamics (i.e., increase of the mean and shear velocities from the sidewalls towards the duct centreline). The intermittency factor $\overline{\langle I \rangle}$ decreases from the sidewalls to the duct centreline, as expected, with increasing sediment transport activity. There is some asymmetry in the transverse distributions that is probably due to the non-perfectly symmetrical inlet conditions in the duct. This effect, however, does not change the validity of the results presented in the following. As for longitudinal distributions, a high level of homogeneity is observed in the streamflow direction (Fig. 3).

The observed transverse changes in transport statistics due to the sidewall effect in our experiment make it possible to study a range of transport conditions (changing across the flow), even using a data set from a single run. In other words, we assume that the different transport conditions across the flow can be viewed as a set of separate experiments and this assumption is strongly supported by Fig. 4. In that figure, we compare the sediment motion statistics at different locations across the flow with those of Ballio and Radice (2007) and Radice and Ballio (2008), who used the data from a range of experiments with different transport intensities (water discharges ranged from 19.8 to 27.5 l/s). Thus, the results of the earlier works closely match the data from much longer time series of sediment parameters. However, some differences between the previous and the new data can be detected in Fig. 4a, b for the mean and standard deviation of sediment concentration at the lowest transport rates. It is believed, in the light of considerations that are made in the following paragraph, that the new-date trend is more reliable.

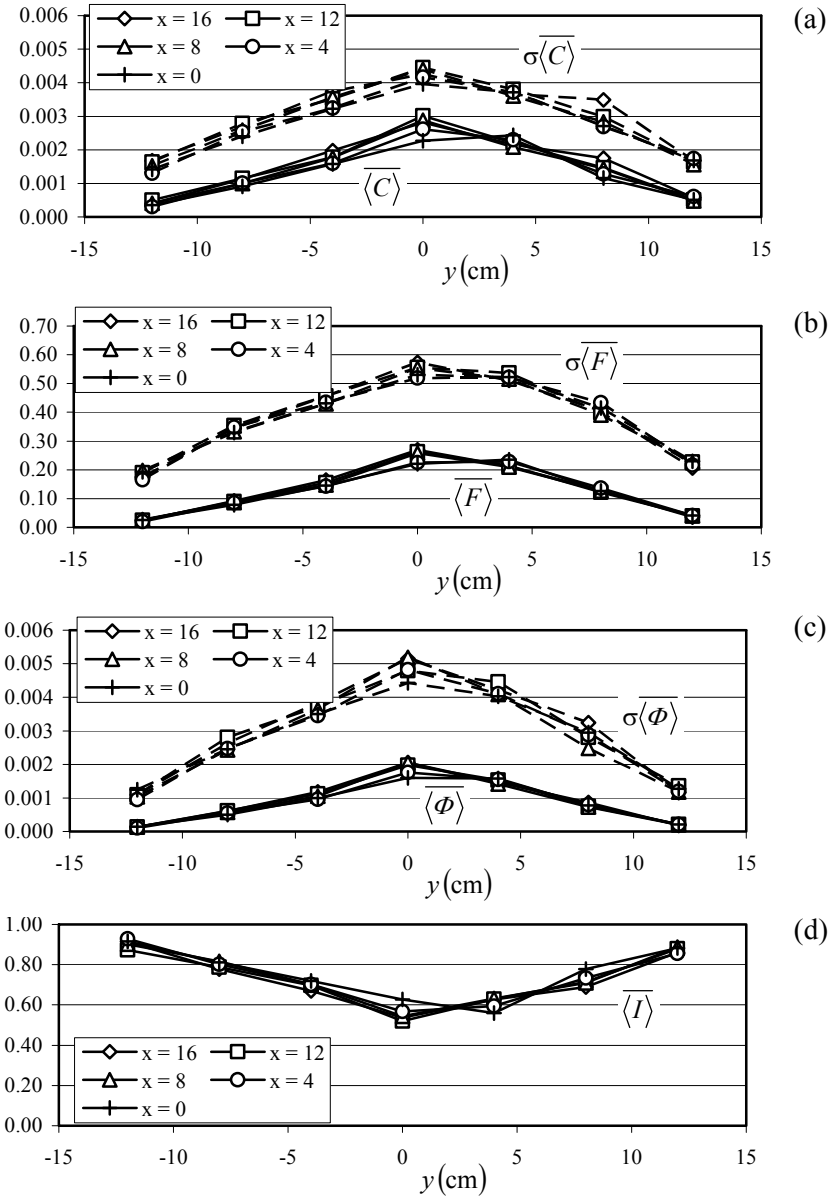


Fig. 3. Temporal statistics of spatially-averaged sediment motion: mean (solid line) and standard deviation (dashed line). (a) Sediment concentration, (b) dimensionless sediment velocity, (c) dimensionless solid discharge per unit width, (d) intermittency coefficient. The positions (x) of spatial-averaging areas along the flow are given in cm and defined in Fig. 1.

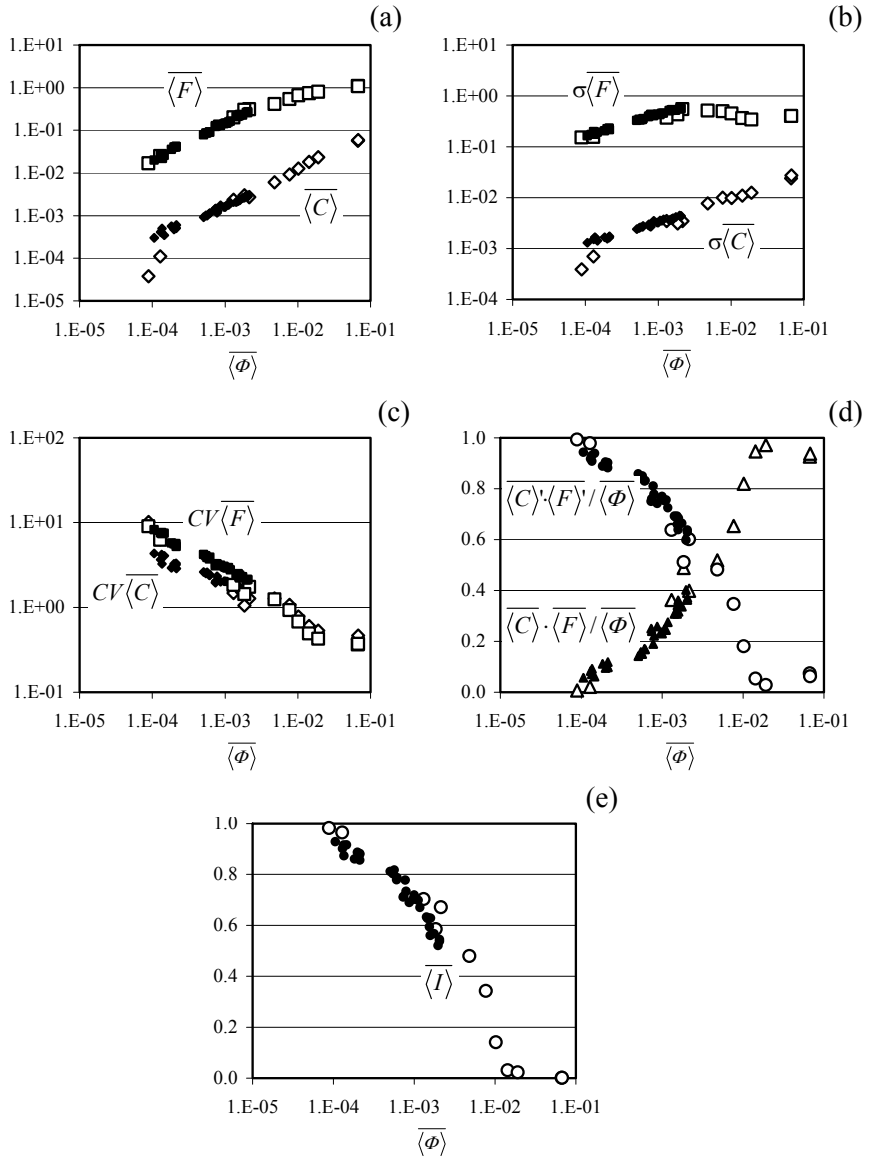


Fig. 4. First and second order statistics of spatially-average sediment concentration (diamonds) and velocity (squares) as a function of the double-average solid discharge: (a) mean; (b) standard deviation; (c) coefficient of variation; (d) mean and fluctuating contributions of concentration and velocity to the overall transport as a function of the double-average solid discharge; and (e) intermittency coefficient as a function of the double-average solid discharge. In all plots, the present experimental data (filled markers) are compared with those (open markers) of Ballio and Radice (2007) or Radice and Ballio (2008).

Considering that the instantaneous, spatially-averaged sediment transport rate per unit width is defined as $\langle \Phi \rangle = \langle C \rangle \langle F \rangle$, its double-averaged value can be represented, using a Reynolds decomposition (Radice and Ballio 2008), as $\overline{\langle \Phi \rangle} = \overline{\langle C \rangle}\,\overline{\langle F \rangle} + \overline{\langle C \rangle' \langle F \rangle'}$ (where the prime indicates the deviation from the mean value). Again, the results of the present work are consistent with those of Radice and Ballio (2008): for the lowest values of the solid discharge the relative contribution of the mean term is negligible, whereas the opposite effect takes place for the largest sediment transport rates (Fig. 4d). The relative contributions of the mean and fluctuating terms have a complementary behaviour, i.e., the sum of two contributions is identically equal to unity. Figure 4e also shows that the intermittency variation is similar for both previously published results and the new long-term experiment.

It is important to stress here that the results of the present work significantly strengthen the results of Ballio and Radice (2007) and Radice and Ballio (2008), where the duration of each experiment was only 20 s with a total sampling area 4 cm long and 8 cm wide. In the study reported here, much longer time series are available for a much larger bed area videoed, significantly increasing the statistical significance of the results. Furthermore, the good agreement between the present data and the earlier results corroborates the previous trends also for the highest transport conditions.

4. HIGH-ORDER STATISTICS OF THE SPATIALLY-AVERAGED SEDIMENT CONCENTRATION

As stated above, some studies, recently reported in the literature, showed that dynamics of the sediment concentration can be assumed to fully represent that of the bed-load process. Therefore, in this section we present analyses of the high-order statistics of sediment concentration time series keeping in mind their significance for the physics of the sediment transport phenomena.

4.1 Cumulative frequency distribution of sediment concentration

Figure 5a presents the Cumulative Frequency Distributions (CFDs) for the concentration data expressed as the ratio of the deviation from the mean to the mean. The asymmetrical nature of concentration fluctuation is evident, reflecting the fact that this quantity is bounded by zero as its minimum value (so that the displayed ratio is obviously bounded by –1) and virtually unbounded above. Furthermore, the discrete nature of the concentration fluctuation can be detected from the CFD curves which are stair-shaped. Towards the duct centreline, the average concentration level becomes progressively larger and the frequency of values identical to the lower bound

decreases. The frequency of –1 values in the distributions is indeed the $\overline{\langle I \rangle}$ coefficient, which was already observed to increase with decreasing transport intensity (Fig. 4). On the other hand, the presence of the no-motion values is obviously off-set by several values which considerably exceed the mean one: the value corresponding to the 95% frequency increases from about 3 at $y = 0$ to more than 8 at $y = -12$ cm (the corresponding coefficient of concentration variation changes from 1.5 to 3.9). We have already mentioned the agreement between the present results and those of Radice and Ballio (2008) in relation to the increasing intermittency for decreasing sediment transport intensity. In addition, our findings also agree with those of Ballio and Radice (2007) and Ancey *et al.* (2008) for the shape of the distribution that becomes more regular with increase in sediment transport (Ballio and Radice 2007 analyzed CFDs of solid discharge and Ancey *et al.* 2008 analyzed CFDs of concentration; consistency of the results is a further indicator that concentration dynamics directly reflects sediment transport rate dynamics).

Ballio and Radice (2007), Radice and Ballio (2008) and Radice (2009) performed statistical analyses of sediment motion considering data sets in which the no-motion samples were excluded ('non-zero' data sets). Following those studies, we present here the CFDs obtained for the non-zero data sets of concentration (Fig. 5b). The curves in Fig. 5b are consistent with the conclusion that was drawn by Ballio and Radice (2007) for the solid discharge, i.e., with zero values removed the concentration distributions for different transport intensities become similar. In this case, the distribution value

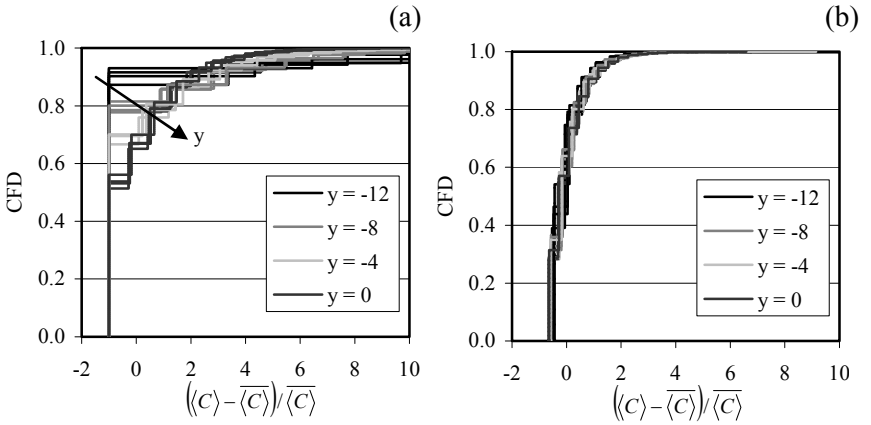


Fig. 5. Cumulative Frequency Distributions of spatially-averaged sediment concentration: (a) (concentration-mean)/mean, all values; (b) (concentration-mean)/mean, non-zero values.

corresponding to the 95% frequency is around 1.3 (coefficient of variation of about 0.7). The collapse among the distributions is also consistent with the results of Radice (2009), who analyzed temporal samples of solid discharge and found that the fluctuations of the values were largely dependent on the process intermittency and exhibited a significant degree of self-similarity as the no-motion events were excluded from the samples. The documented findings indicate that the flow influences the bed load mainly by controlling the frequency of the motion events. In addition, the same relative concentration fluctuations occur during motion events, regardless of transport intensity.

4.2 Analyses in the time and frequency domains

In this sub-section we evaluate the characteristic time scales of the concentration fluctuations using different approaches. A range of estimates obtained for these scales shall be compared in Section 5.

Figure 6 presents the result of the auto-correlation analysis of the concentration data. The curves for different transverse positions (i.e., for different sediment transport intensities) are similar, except for a difference in concavity for the smallest lags (upwards concavity in the curve for $y = -12$ cm and downwards for all others). An autocorrelation (time) length may be computed by fitting parabolas to the first three points of the curves (estimated length will be indicated with "Autocorr-parabola" in the following), obtaining values around 0.16-0.18 s. Such procedure may be however subject to imprecision and should be accounted for with care; in addition, for the obtained lags the correlation coefficient is still quite high, being equal to almost 0.5. A correlation length of 0.4 s (hereafter identified with "Autocorr-0.2") can be estimated for all curves as that corresponding to a correlation coefficient of about 0.2.

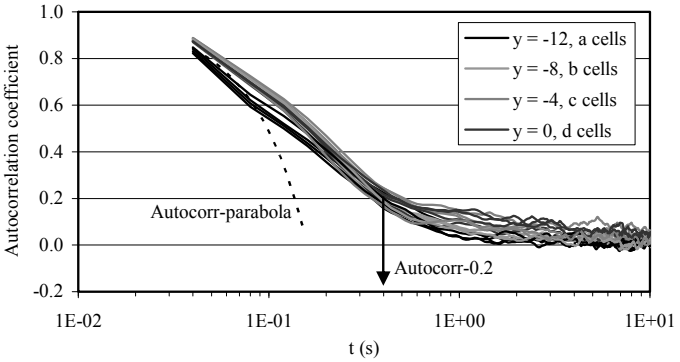


Fig. 6. Autocorrelation functions for the concentration time series. For transverse positions and cell labels refer to Fig. 1.

The temporal dynamics of the sediment concentration was also analyzed by means of the high-order structure functions. The pth order Generalized Structure Function (GSF) of an X quantity is defined as $D_{Gp}(\tau) = <|\Delta X(\tau)^p|>$, where τ defines the time lag between any two values and the angular brackets indicate the average. The GSF values are generally increasing as the lag increases from the sampling interval to larger values, since the average difference between X pairs increases as correlation is lost. On the other hand, GSFs typically attain an asymptotic level at large lags, since the data sets have finite variances. If a τ range exists where a relationship $D_{Gp}(\tau) \propto \tau^K$ holds, it is said that the X variable exhibits a scaling behaviour in this range. Furthermore, if a relationship of the type $K(p) = pH$ holds with H being a constant, the X quantity is said to exhibit simple scaling behaviour. In this case, the H parameter is called the Hurst exponent. The latter assumes a 1/3 value in turbulence inertial subrange in fluid dynamics. A potential benefit of application of structure functions to sediment transport data was demonstrated by Nikora and Goring (2001). Radice *et al.* (2009) have applied this tool to a subset of the experimental data used here (i.e., for $y = 0$ only). In this paper we consider the structure functions for the whole range of y positions available from this experimental dataset. The GSFs for the 4 locations with the same y and $4 \le x \le 16$ were averaged to obtain a single GSF representative of a transport condition. Figure 7 depicts the GSFs obtained. An evident scaling range can be detected for all the transport conditions except that for the weakest one ($y = –12$ cm). We estimated the upper limit of the scaling range from the 3rd order structure function. An estimation of a time scale can be made using an 'ideal' time lag at which the linear trend and the horizontal line of the saturation region intersect ("StrFunct" in the following).

Figure 8 summarises the information on the scaling exponents for a range of structure function orders. All the curves can be approximated as linear functions supporting the simple scaling in time series fluctuations. Furthermore, the line slopes generally grow with increase in the sediment transport capacity. The line slope is around 1/3 for $y = –4$ cm and $y = 0$. The curve obtained for $y = –12$ cm (near the sidewall of the duct, not shown in the plot) could not be approximated by a linear trend because the apparent scaling range for the highest order structure functions was limited to the first two points in the structure functions (Fig. 7). Therefore, the scaling range would correspond to only twice the sampling interval and it was considered insignificant.

The spectra of the concentration time series were computed for each location and are depicted in Fig. 9. The spectra for $y > –12$ cm contain a scaling range where they can be approximated with a power function with –5/3

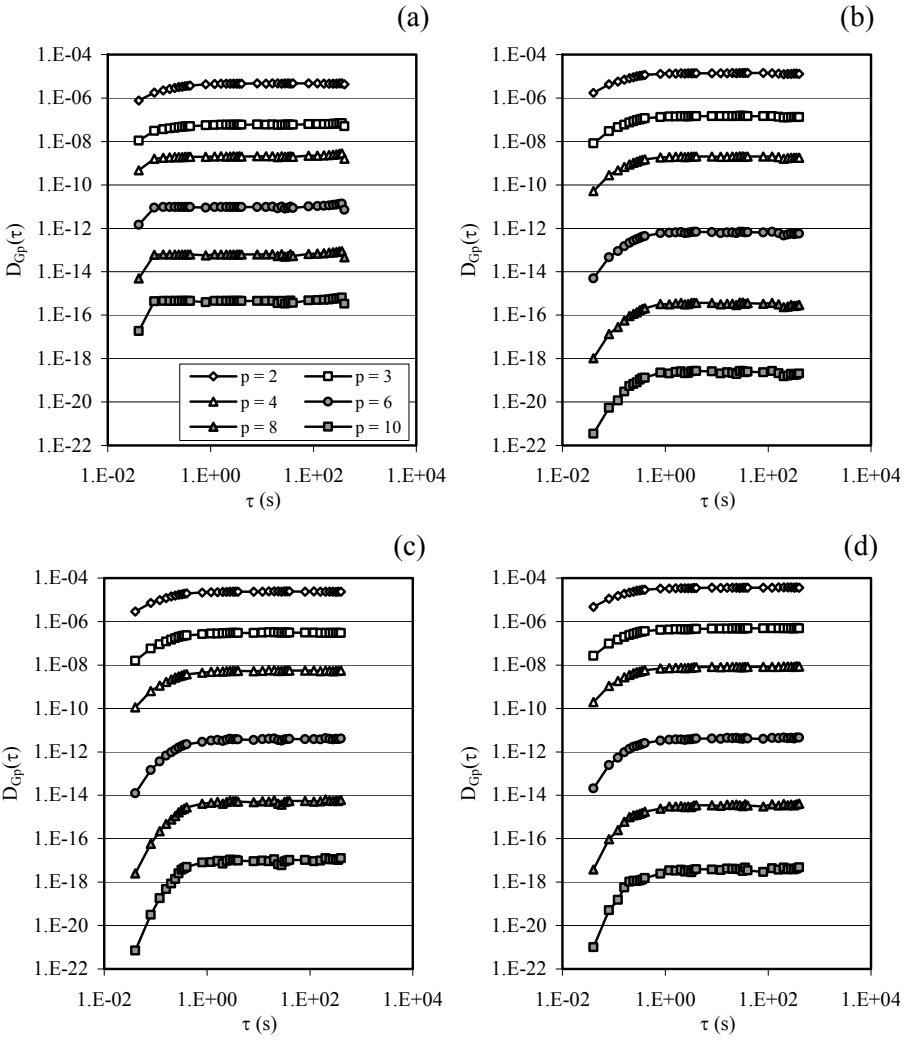


Fig. 7. Structure functions for $y = -12$ cm (a), $y = -8$ cm (b), $y = -4$ cm (c), and $y = 0$ (d). Legend in the first plot holds also for the others.

exponent, while the spectrum for $y = -12$ cm shows a lower spectral slope. The limits of the scaling range can be identified as the frequencies where the spectrum deviates from the straight line in log-log coordinates; the corresponding time scales can be calculated as the inverse of the frequency (in the following, "Spect-min" and "Spect-max"; note that minimum time corresponds to the high bound of frequency range and *vice versa*).

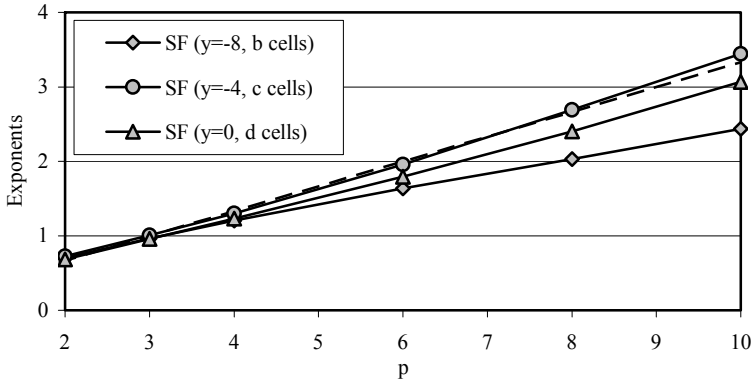


Fig. 8. Exponents for the scaling range of the SFs. The dashed line indicates a linear function with 1/3 coefficient.

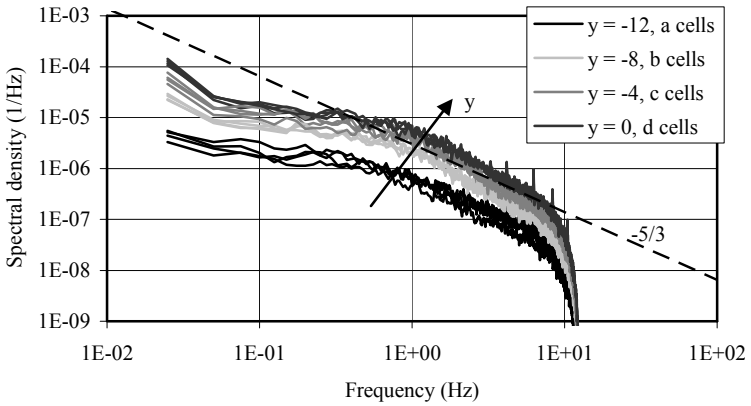


Fig. 9. Spectra of concentration fluctuations. Note that the spectra for the cells at the same y nearly coincide.

The 1/3 value for the scaling exponent and the –5/3 value for the spectrum one are, as mentioned, typical of turbulence inertial subrange in fluid mechanics. Our results are therefore consistent with a picture of the sediment concentration dynamics reflecting a hierarchy of turbulent eddies and thus being a dynamic footprint of the turbulent flow structure.

As highlighted above (Section 2), the time signals of sediment concentration for different longitudinal positions and the same transverse position present a certain degree of similarity that can be quantified using a cross-correlation analysis, separately for each y position. Figure 10 depicts the cross-correlation functions for any possible couple of locations for $y = -4$ cm (we have three couples for 4 cm separation distance, two couples for 8 cm separation distance, and one couple for 12 cm separation distance). Although

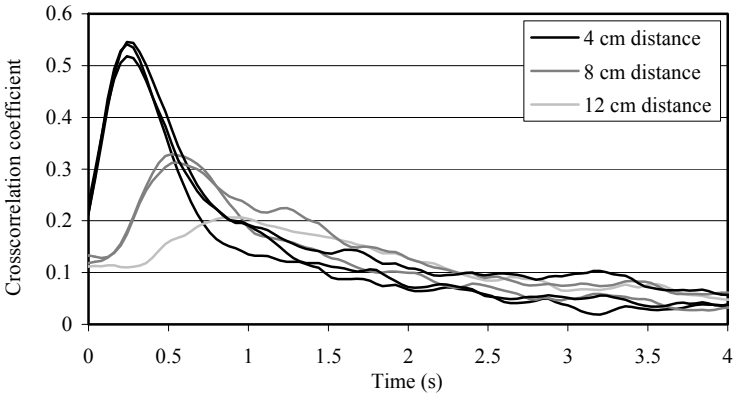


Fig. 10. Cross-correlations for all the possible couples of locations with $y = -4$ cm (c cells).

the cross-correlation peak is less pronounced for larger distance, the results are in agreement, indicating a constant velocity of migration of the concentration fluctuations. We can estimate an average celerity of concentration fluctuations (c in the following) as the ratio of the separation distances to the time lags of the cross-correlation peaks.

5. ANALYSIS OF CHARACTERISTIC SCALES

In this section we compare the characteristic time scales for the dynamics of bed-load sediment concentration, obtained above, and for the turbulent bursting cycle (indicated with "Burst") that can be evaluated as $T = k\delta/U_{max}$, with k from 2 to 5 (Nezu and Nakagawa 1993, Nikora and Goring 2000). We obtain T values ranging from 0.4 s to 1.1 s for k from 2 to 5 (having $\delta = 8$ cm and $U_{max} = 37$ cm/s). This scale is representative of the centreline, where we measured the velocity profile. In addition, we will consider the following parameters:

- The section-averaged water velocity U, which is equal to 32.8 cm/s.
- The "near bed velocity scale" u_{nb}, which is the longitudinal water velocity at an elevation of 4 mm above the mean sediment bed. The vertical velocity profile was measured at the centreline using an Ultrasonic Doppler Profiler (see Radice *et al.* 2009 for details). This velocity scale, equal to 11.0 cm/s, has to be viewed as representative of the centreline only.
- The double-averaged particle velocity $\overline{\langle u_s \rangle}$. It should be borne in mind that such velocity is the average particle velocity within the observation period and, therefore, accounts for the no-motion period. We will consider also the 'non-zero' particle velocity $\overline{\langle u_{sNZ} \rangle}$ (that is, the average velocity of particles during motion only).

❑ The celerity c of propagation of concentration fluctuations estimated using cross-correlation functions.

All the scales are depicted in Fig. 11. The scaling range of the spectra (Fig. 11a) comprehends all the other time scales. It is reasonable to assume that this range reflects a hierarchy of turbulent eddies spanning several scales. A separation can be observed between the scales of the correlation of the particle concentration and those of the turbulent bursting (the latter are represented as a grey sector that accounts for the variability of the k coefficient in the previously mentioned expression for T). The correlation pattern of concentration seems to be related to smaller structures compared with those characterizing the turbulent bursts. The velocity scales (Fig. 11b) show a similarity between u_{nb} and $\overline{\langle u_{sNZ} \rangle}$, whilst $\overline{\langle u_s \rangle}$ is much lower than the former and U is much larger. Thus, the physical velocity of particles during motion seems dominated by the time-averaged near-bed water velocity (and in turn by the time-averaged bed shear stress) while, on the contrary, the double-averaged particle velocity is affected by the intermittent features of the process. The variability of $\overline{\langle u_{sNZ} \rangle}$ is much smaller than that of $\overline{\langle u_s \rangle}$, in agreement with previous findings of Radice and Ballio (2008) and with the results, outlined above, showing that significant similarities emerge as intermittency is removed from the analysis (Ballio and Radice 2007, Radice 2009). The celerity of propagation of concentration fluctuations, c, is slightly larger than the near-bed flow velocity and physical velocities of particles (Fig. 11b).

The double-averaged sediment motion (represented here by $\overline{\langle u_s \rangle}$ in Fig. 11b and $\overline{\langle C \rangle}$ and $\overline{\langle I \rangle}$ in Fig. 11c) is, as also previously seen, highly dependent on the transverse position. On the contrary, the physical scales of particle motion ($\overline{\langle u_{sNZ} \rangle}$, previously discussed, c, and all the time scales of concentration dynamics) do not vary with the position and, in turn, with the double-averaged sediment transport intensity. To the best of our knowledge, this is the first time that such a behaviour emerges so clearly from experimental observation. The low variability of the characteristic time scales across the flow is striking considering that the intermittency factor, which is indeed a key item in the temporal dynamics of particle concentration, is highly dependent on the transport intensity. This finding deserves further attention in the follow-up studies.

Radice *et al*. (2009) described a conceptual picture according to which the dynamics of the bed-load sediment concentration is triggered by near-bed turbulent events which migrate along the flow, with the correlation of the concentration dynamics reasonably reflecting that of the flow field dy-

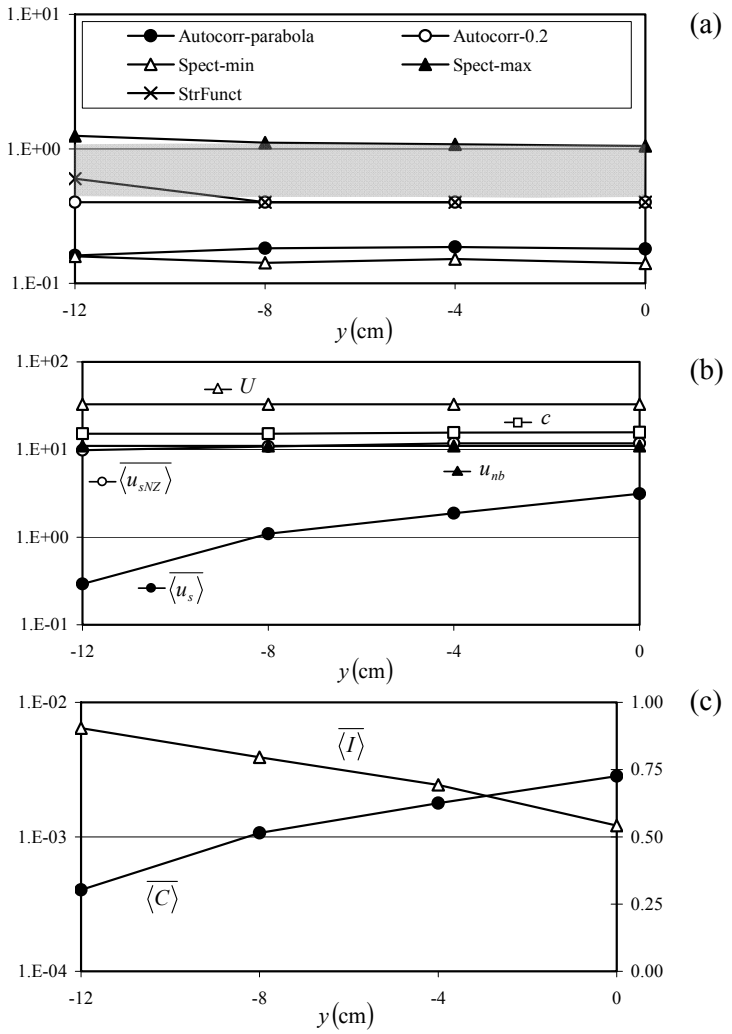


Fig. 11: (a) Temporal scales (s) of particle concentration fluctuations estimated with different methods; the grey area depicts the range of variability of the time scale indicated as "Burst" in the text. (b) Velocity scales (cm/s). (c) Transverse profile of double-averaged concentration and intermittency factor (averaged curves from Fig. 3a, d).

namics. The picture was limited to the sediment transport dynamics at the channel centreline. The results of this study are consistent with this concept. The present findings suggest that as the duct wall is approached, the strength of the turbulent events becomes lower and induces a less intense sediment transport, although the temporal scales for the propagation of the triggering events do not change significantly.

6. VISUALIZATIONS

The analyses outlined above demonstrated the spatial and temporal correlations in the sediment concentration dynamics, suggesting that the sediment transport process may be a succession of correlated events. In this section, a sample visualization of the sediment motion is presented to qualitatively depict the interpretation above (i.e., in Section 5). The sediment motion correlation has to be viewed in a statistical sense, as a result of contributions of many transport events. One of such transport events is described in this section.

Figure 12 presents a sample visualization of particle motion in the centreline region ('d' cells of Fig. 1). The width of the focus area is 4 cm and the length is 16 cm. The white blobs correspond to particles in motion recognized by image filtering. The images present accumulations of blobs moving at different instants, to show the positions progressively taken by the particles in motion. The visualization is compound of five frames with increasing accumulation time; the latter equals 0.2 s in the first left plot, 0.4 s in the second, and so on, so that the visualization period in the last picture on the right is 1 s. In the all pictures, the sediments accumulated in the last 0.2 s interval are white, while those accumulated in previous 0.2 s periods are grey. In this example, we detect one event in the first 0.2 s (highlighted with

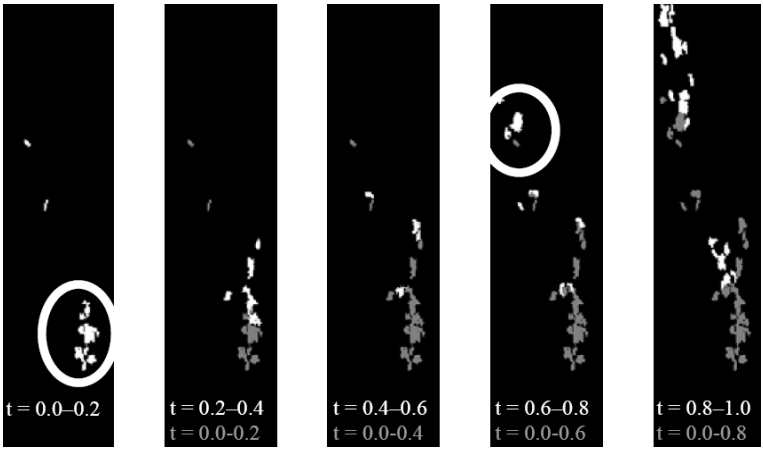


Fig. 12. Sample visualization of particle motion. Pictures for $y = 0$ and for a 4×16 cm^2 region (i.e., an observation area corresponding to the 'd' cells in Fig. 1). In the first picture (left), the images of the moving particles are accumulated over a 0.2 s period. In the successive pictures (from left to right) the accumulation period is progressively increased by 0.2 s increments, so that in the last picture (right) the accumulation period equals 1 s. In all the pictures the sediments accumulated in the last 0.2 s are white, while those accumulated previously are grey. All times t in the pictures are indicated in s.

the white oval). As time progresses, the sediments set in motion during this event come to rest (between 0.4 and 0.6 s, third picture) while we detect another event about 8 cm downstream of the previous one. Hypothesizing that the two events are induced by the same migrating turbulent structure, we would estimate a celerity of propagation of 13 cm/s, which is consistent with previous evaluations based on cross-correlation analysis.

7. CONCLUSIONS

This paper reports an analysis of the sediment motion dynamics in one-dimensional bed load. We have presented experimental results obtained from a long-duration experiment in which the sediment kinematics has been measured by means of image processing techniques. The focus area investigated was wide enough to observe a significant sidewall effect in the experimental duct. As a result, the single experiment made has enabled a variety of transport intensities to be observed and analyzed.

We have analyzed the relationships between the double-averaged values of the measured quantities. The contribution of the mean and temporally fluctuating values of sediment concentration and particle velocity to the resulting sediment transport rate has been also shown. The present results represent a corroboration of earlier ones, which were obtained from data sets whose statistical significance was much lower compared to that for this study.

The temporal dynamics of the spatially-averaged concentration has been analyzed as representative of that of the sediment transport rate. All the analyses made have highlighted a significant degree of correlation in the sediment concentration pattern. A number of characteristic time scales have been estimated with different methods, including auto- and cross-correlation analysis, high order structure functions and spectra. In addition, some velocity scales for the sediment motion and for the flow field have been considered. The analysis of the large data set allowed to clearly demonstrate some key issues: (i) the characteristic time scales of sediment concentration dynamics are almost constant for a range of the double-averaged sediment concentration spanning one order of magnitude; and (ii) the physical velocity of sediment particles during motion is similar to the near-bed longitudinal water velocity and slightly lower than the celerity of propagation of the sediment concentration patterns. The findings lead to a picture where the bed-load sediment concentration is triggered by near-bed turbulent events which migrate along the flow, with the celerity of propagation of turbulent events being similar for the different transport intensities.

Even though the earlier findings reported in the literature suggest that the dynamics of sediment concentration could effectively represent the sediment

transport rate, analyses similar to those reported in this paper should be also conducted for the sediment velocity fluctuations.

Acknowledgments. The research was funded by the Italian Ministry of University and Research (M.I.U.R.) under the "MOMICS" PRIN Project. Two anonymous reviewers are gratefully acknowledged for the comments given on an earlier version of the manuscript.

References

Ancey, C., A.C. Davison, T. Böhm, M. Jodeau, and P. Frey (2008), Entrainment and motion of coarse particles in a shallow water stream down a steep slope, *J. Fluid Mech*. **595**, 83-114, DOI: 10.1017/S0022112007008774.

Ballio, F., and A. Radice (2007), Grain kinematics in weak linear transport, *Arch. Hydro-Eng. Environ. Mech.* **54**, 223-242.

Böhm, T., P. Frey, C. Ducottet, C. Ancey, M. Jodeau, and J.L. Reboud (2006), Two-dimensional motion of a set of particles in a free surface flow with image processing, *Exp. Fluids* **41**, 1, 1-11 DOI: 10.1007/s00348-006-0134-9.

Chanson, H. (1999), *The Hydraulics of Open Channel Flow: An Introduction*, Elsevier Butterworth-Heinemann, Amsterdam.

Drake, T.G., R.L. Shreve, W.E. Dietrich, P.J. Whiting, and L.B. Leopold (1988), Bedload transport of fine gravel observed by motion-picture photography, *J. Fluid Mech*. **192**, 193-217, DOI: 10.1017/S0022112088001831.

Einstein, H.A. (1937), *Der Geschiebebetrieb als Wahrscheinlichkeitsproblem*, Verlag Rascher, Zurich.

Einstein, H.A. (1950), The bed-load function for sediment transportation in open channel flows, *Tech. Bull. U.S. Dept. of Agric*., 1026.

Fernandez Luque, R., and R. van Beek (1976), Erosion and transport of bed-load sediment, *J. Hydraul. Res*. **14**, 127-144.

Forliti, D.J., P.J. Strykowski, and K. Debatin (2000), Bias and precision errors of digital particle image velocimetry, *Exp. Fluids* **28**, 5, 436-447, DOI: 10.1007/s003480050403.

Gomez, B., and M. Church (1989), An assessment of bed load sediment transport formulae for gravel bed rivers, *Water Resour. Res*. **25**, 6, 1161-1186, DOI: 10.1029/WR025i006p01161.

Lau, Y.L., and B.G. Krishnappan (1985), Sediment transport under ice cover, *J. Hydr. Engrg*. **111**, 6, 934-950, DOI: 10.1061/(ASCE)0733-9429(1985) 111:6(934).

Lee, H.Y., and I.S. Hsu (1994), Investigation of saltating particle motions, *J. Hydr. Engrg*. **120**, 7, 831-845, DOI: 10.1061/(ASCE)0733-9429(1994)120:7(831).

Martin, Y. (2003), Evaluation of bed load transport formulae using field evidence from the Vedder River, British Columbia, *Geomorphology* **53**, 1-2, 75-95, DOI: 10.1016/S0169-555X(02)00348-3.

Meyer-Peter, E., and R. Müller (1948), Formulas for bed-load transport, Proc. of the II Meeting of IAHR, Stockholm, Sweden.

Nelson, J.M., R.L. Shreve, S.R. McLean, and T.G. Drake (1995), Role of near-bed turbulence structure in bed load transport and bed form mechanics, *Water Resour. Res*. **31**, 8, 2071-2086, DOI: 10.1029/95WR00976.

Nezu, I., and H. Nakagawa (1993), *Turbulence in Open-Channel Flows*, Balkema, Rotterdam.

Nikora, V., and D. Goring (2000), Flow turbulence over fixed and weakly mobile gravel beds, *J. Hydr. Engrg*. **126**, 9, 679-690, DOI: 10.1061/(ASCE)0733-9429(2000)126:9(679).

Nikora, V.I., and D.G. Goring (2001), Extended self-similarity in geophysical and geological applications, *Math. Geol*. **33**, 3, 251-271, DOI: 10.1023/A:1007630021716.

Nikora, V., H. Habersack, T. Huber, and I. McEwan (2002), On bed particle diffusion in gravel bed flows under weak bed load transport, *Water Resour. Res*. **38**, 6, 1081, DOI: 10.1029/2001WR000513.

Nikora, V., I. McEwan, S. McLean, S. Coleman, D. Pokrajac, and R. Walters (2007), Double-averaging concept for rough-bed open-channel and overland flows: Theoretical background, *J. Hydr. Engrg*. **133**, 8, 873-883, DOI: 10.1061/(ASCE)0733-9429(2007)133:8(873).

Niño, Y., and M. Garcìa (1998a), Experiments on saltation of sand in water, *J. Hydr. Engrg*. **124**, 10, 1014-1025, DOI: 10.1061/(ASCE)0733-9429(1998)124:10(1014).

Niño, Y., and M. Garcìa (1998b), Using Lagrangian particle saltation observations for bedload sediment transport modelling, *Hydrol. Process*. **12**, 8, 1197-1218, DOI: 10.1002/(SICI)1099-1085(19980630)12:8<1197::AID-HYP612>3.0.CO;2-U.

Papanicolaou, A.N., P. Diplas, M. Balakrishnan, and C.L. Dancey (1999), Computer vision technique for tracking bed load movement, *J. Comp. Civ. Engrg*. **13**, 2, 71-79, DOI: 10.1061/(ASCE)0887-3801(1999)13:2(71).

Parker, G., G. Seminara, and L. Solari (2003), Bed load at low Shields stress on arbitrarily sloping beds: Alternative entrainment formulation, *Water Resour. Res*. **39**, 7, 1183, DOI: 10.1029/2001WR001253.

Pilotti, M., G. Menduni, and E. Castelli (1997), Monitoring the inception of sediment transport by image processing techniques, *Exp. Fluids* **23**, 3, 202-208, DOI: 10.1007/s003480050103.

Radice, A. (2009), Use of the Lorenz curve to quantify statistical nonuniformity of sediment transport rate, *J. Hydr. Engrg*. **135**, 4, 320-326, DOI: 10.1061/(ASCE)0733-9429(2009)135:4(320).

Radice, A., and F. Ballio (2008), Double-average characteristics of sediment motion in one-dimensional bed load, *Acta Geophys*. **56**, 3, 654-668, DOI: 10.2478/s11600-008-0015-0.

Radice, A., S. Malavasi, and F. Ballio (2006), Solid transport measurements through image processing, *Exp. Fluids* **41**, 5, 721-734, DOI: 10.1007/s00348-006-0195-9.

Radice, A., F. Ballio, and V. Nikora (2009), On statistical properties of bed load sediment concentration, *Water Resour. Res*. **45**, W06501, DOI: 10.1029/2008WR007192.

Shields, A. (1936), Anwendung der Aehnlichkeitsmechanik und der Turbulenz Forschung auf die Geschiebebewegung, *Mitt. der Preussische Versuchanstalt für Wasserbau und Schiffbau*, Tech. Hochsch., Berlin, Germany, n. 26.

Smith, B.T., and R. Ettema (1997), Flow resistance in ice-covered alluvial channels, *J. Hydr. Engrg*. **123**, 7, 592-599, DOI: 10.1061/(ASCE)0733-9429(1997)123:7(592).

Received 17 November 2009
Accepted 2 February 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1094-1114
DOI: 10.2478/s11600-009-0063-0

Analysis of Solute Transport in Rivers with Transient Storage and Lateral Inflow: An Analytical Study

Akhilesh KUMAR and Durga C. DALAL

Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati, India
e-mails: akhileshk@iitg.ernet.in (corresponding author), durga@iitg.ernet.in

A b s t r a c t

Solute transport in rivers with transient storage and lateral inflow has been studied by Transient Storage Model (TSM). Analytical solution of the TSM is obtained by means of Laplace transform. In order to illustrate the use of present analytical solution, physical transport parameters are estimated for the observational data of Uvas Creek tracer experiment. Analytical concentration-time breakthrough curves are found to be in good agreement with the observed concentration-time breakthrough curves. A sensitivity analysis has been performed in order to identify the most critical parameter for predicting concentration. It has been found that lateral inflow rate is the most sensitive and the ratio of cross-sectional areas is the least sensitive parameter. A hypothetical situation has been considered to study the effects of background concentrations, lateral inflow concentration and its rate. The analytical solutions show that the solute concentration gets diluted or concentrated due to lateral inflow. Physical mechanisms of the problems are well reproduced by the present analytical solutions and these results can be used for analysis of tracer experiments.

Key words: solute transport in river, transient storage model, lateral inflow, physical phenomena in river.

1. INTRODUCTION

Analysis of solute transport in rivers with the effects of transient storage has been an active research area since the early 1980s. Waste materials discharged from chemical plants or industries into rivers, undergo stages of mixing while being transported downstream. Stagnant water zones (or storage zones) are normally found in rivers. Waste materials are usually temporarily trapped in these stagnant water zones and slowly released back into the main stream (which can be seen from Fig. 1). Several models have been proposed to describe these physical phenomena in rivers (e.g., Hays *et al.* 1966, Bencala and Walters 1983, Runkel and Chapra 1994, Harvey *et al.* 1996, Wagner and Harvey 1997, Czernuszenko and Rowiński 1998, Wörman 1998). A commonly referenced model known as the Transient Storage Model (TSM) or the Dead Zone Model (DZM) is based on the classical advection-dispersion phenomenon with the exchange of pollutants between the main stream of river and the stagnant water zones such as pools, gravel beds, side arms of the river or adjacent wetland areas (Bencala and Walters 1983, Runkel and Chapra 1993, Wagner and Harvey 1997). Generally, rivers are divided into two distinct zones in the TSM. The first zone represents the main flow region that includes the process of advection, dispersion and lateral inflow. The second one represents the storage zone area, which is stagnant relative to the main channel flow of a river. Both zones are linked by the mass exchange process between them (which can be seen from Fig. 2).

Both semi-analytical and analytical solutions are available for the model. Hays *et al.* (1966) described the transient storage model in the form of two partial differential equations, one for the main stream of river and the other one for the storage zone. They showed that the approximate solutions of these coupled partial differential equations could describe the natural behavior

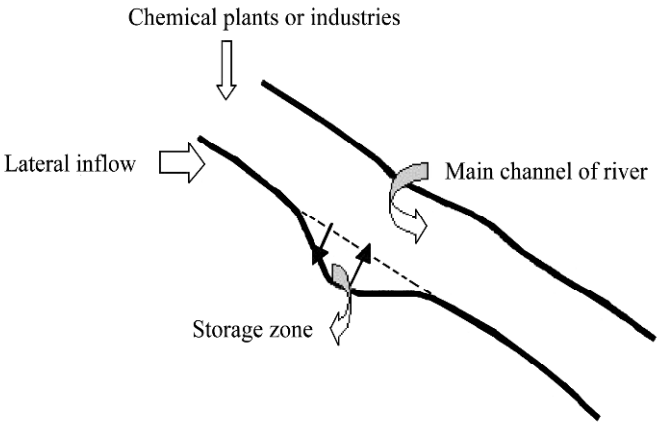


Fig. 1. Schematic physical system for stream and storage zone hydrologic process.

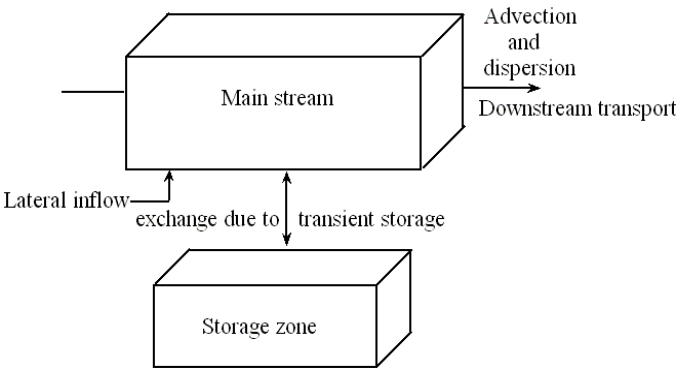


Fig. 2. Conceptual model describing the stream and storage zone hydrologic process for conservative solute.

of longitudinal concentration distribution more accurately than the solution of classical advection-dispersion equation. They also stated that the exchange coefficient and the dead zone ratio (i.e., two additional parameters due to the transient storage) were difficult to predict from bulk hydraulic or geometrical parameters. Thackston and Schnelle (1970) evaluated the dead zone parameters using similar approximate solution given by Hays *et al.* (1966) with the help of field data. In their analysis, they assumed the exchange coefficient to be constant and the dead zone volume to be a function of the friction factor. Hart (1995) developed a semi-analytical solution for conservative solutes, whereas Schmid (1997) developed it for decaying solutes. Both of them considered the physical mechanisms of transient storage as stochastic processes and they did not consider the effects of lateral inflow in their studies. Earlier, Bencala and Walters (1983) included the term due to the lateral inflow in the dead zone model proposed by Hays *et al.* (1966). Parameters for the TSM were evaluated for the best fit of the observed solute concentration curve by using the numerical solution of the TSM. Bencala and Walters (1983) found that the effect of lateral inflow was negligible. Subsequently, several experimental studies have been made to characterize the physical transport properties and hydrological processes in rivers (e.g., Harvey *et al.* 1996, Scott *et al.* 2003). The effects of transient storage on solute transport in streams and rivers with lateral inflow were among the main issues discussed in these experimental studies.

Runkel and Chapra (1993) presented a numerical solution of the TSM by using the Crank–Nicolson numerical scheme. This is the basis of the One-dimensional Transport with Inflow and Storage (OTIS). It also includes the numerical solution for reactive solute transient storage model. OTIS was extended by Runkel (1998) with the nonlinear least squares method to optimize the parameters. This extended OTIS is known as OTIS-P. It has been exten-

sively used to estimate the transient storage behavior of rivers through tracer experiments (Choi *et al.* 2000, Fernald *et al.* 2001).

Davis *et al.* (2000) and De Smedt *et al.* (2005) found analytical solutions of the TSM for instantaneous solute release. In 2006, De Smedt (2006) obtained analytical solutions of the TSM for decaying solutes. In his study, De Smedt did not consider the effects of lateral inflow and also the effects of background concentration in the main stream as well as in the storage zone.

The purpose of the present work is to find exact analytical solutions of the conservative solute transport in rivers with transient storage by including the effects due to background concentrations, lateral inflow solute concentration and its rate have been derived. Analytical solution for the case without storage can easily be obtained from the present analytical solution. In order to illustrate the use of present analytical solutions, parameters are estimated by using large-scale Newton reflexive method (Coleman and Li 1994, 1996) for the tracer experiment conducted in the Uvas Creek. A sensitive analysis has been performed to identify the most critical parameter for predicting concentration. A hypothetical situation has been considered to study the effects of background concentrations, lateral inflow concentration and its rate on the breakthrough curves. Step concentration-time profile is considered as upstream boundary condition for the analytical solutions.

2. THEORY

2.1 Governing equations

The governing equations for the TSM are as follows (Bencala and Walters 1983, Runkel and Chapra 1994):

$$\frac{\partial c_m^*}{\partial t^*} + u_m^* \frac{\partial c_m^*}{\partial x} = D_m^* \frac{\partial^2 c_m^*}{\partial x^{*2}} + \alpha^* \left(c_s^* - c_m^* \right) + \frac{q_L^*}{A_m^*} \left(c_L^* - c_m^* \right), \tag{1}$$

$$\frac{\partial c_s^*}{\partial t^*} = \frac{\alpha^* A_m^*}{A_s^*} \left(c_m^* - c_s^* \right), \tag{2}$$

where c_m^* is the solute concentration in the main flow direction of river, c_s^* is the solute concentration in the storage zone, c_L^* is the constant solute concentration in lateral inflow, $Q_{x^*}^*$ is the in-stream volumetric flow rate, $u_m^* = Q_{x^*}^* \big/ A_m^*$ is the cross-sectional average velocity in the main channel, x^* is the longitudinal distance in the main stream channel from the upstream location, t^* is a time, D_m^* is the longitudinal dispersion coefficient in the main flow, A_m^* is the cross-sectional area of the main stream of river, A_s^* is

the cross-sectional area of the storage zone, α^* is the stream storage exchange coefficient, and q_L^* is the lateral inflow rate into the stream per unit stream length.

In tracer experiments, the river reaches are usually divided into pieces which may be assumed to be homogeneous, and therefore the transport parameters are assumed to be constant, except the volumetric flow rate $Q_{x^*}^*$ which varies linearly between the upstream and downstream sampling location.

Equation (1) is obtained by linking the one-dimensional classical advection-dispersion equation with a first order mass exchange term between the main channel and the storage zone (the second term on the right-hand side) and a term for lateral inflow that enters directly into the main channel bypassing the storage zone (the last term on the right-hand side). Equation (2) represents the rate of change of solute concentration in the storage zone assuming it is proportional to the difference in the solute concentration between the main channel and the storage zone.

2.2 Dimensionless forms of governing equations and initial-boundary conditions

The dimensionless forms of the governing eqs. (1) and (2) are given as

$$\frac{\partial c_m}{\partial t} + \frac{\partial c_m}{\partial x} = \frac{1}{Pe_m}\frac{\partial^2 c_m}{\partial x^2} + Da_{em}\left(c_s - c_m\right) + q_L\left(c_L - c_m\right), \tag{3}$$

$$\frac{\partial c_s}{\partial t} = \frac{Da_{em}}{k_r}\left(c_m - c_s\right), \tag{4}$$

where $x = x^*/l^*$ is the distance in the longitudinal direction, $l^* \approx \sqrt{2.5\,A_m^*}$ is the characteristic length, $t = u_m^* t^*/l^*$ is the dimensionless time, u_m^* is a constant quantity, $c_m = c_m^*/c_0^*$ is the concentration in the main stream of the river, c_0^* is the plateau concentration at the head of the tracer experiment, $c_s = c_s^*/c_0^*$ is the concentration in the storage zone, $c_L = c_L^*/c_0^*$ is the concentration in the lateral inflow, $Pe_m = u_m^* l^*/D_m^*$ is the Péclet number in the main stream of river, $Da_{em} = \alpha^* l^*/u_m^*$ is the exchange Damköhler number in the main stream, $k_r = A_s^*/A_m^*$ is the ratio of cross-sectional area of storage zone and the cross-sectional area of the main stream, and $q_L = q_L^* l^*/u_m^* A_m^*$ is the lateral inflow rate.

The initial conditions for the main stream and the transient storage are

$$c_m\left(x,t=0\right)=c_{m\text{init}}\,, \tag{5}$$

$$c_s\left(x,t=0\right)=c_{s\text{init}}\,, \tag{6}$$

where $c_{m\text{init}}$ and $c_{s\text{init}}$ are the background concentrations in the main stream and in the storage zone, respectively.

Generally, in tracer experiments, a known amount of solute is injected at a constant rate at the injection location, over a certain interval of time, and then the solute concentrations are measured at different sampling locations (Bencala and Walters 1983, Schmid 2003). So, the upstream boundary condition is considered to be a step concentration profile and is expressed as

$$c_m\left(x_0,t\right)=c_u\left(t\right), \tag{7}$$

where x_0 is the upstream boundary location, and $c_u(t)$ is the dimensionless step concentration profile at x_0. It is also assumed that Laplace transform of $c_u(t)$ exists and is bounded.

The other boundary condition, at the far downstream location, is taken as

$$\frac{\partial c_m}{\partial t}=0 \quad \text{as} \quad x\to+\infty\,. \tag{8}$$

2.3 Calculation of the cross-sectional average velocity u_m^*

Both equations (1) and (2) are based on the theory of conservation of mass. The conservation of mass of water can be written, following Ge and Boufadel (2006), as $\mathrm{d}Q_{x^*}^*/\mathrm{d}x^*=q_L^*-q_{L\text{out}}^*$, where $q_{L\text{out}}^*$ is the lateral outflow per unit stream length. In our study, we have assumed that the water does not leave the main stream throughout the study reach, which means $q_{L\text{out}}^*$ is zero. So, the volumetric flow rate at the downstream river reach Q_1^* can be obtained, integrating the resulting equation over $\left[x_0^*,x_1^*\right]$, as $Q_1^*=Q_0^*+q_L^*\left(x_1^*-x_0^*\right)$, where subscripts 0 and 1 represent the upstream location and downstream location, respectively. We consider the averaged value of the volumetric flow rate Q^*, which is a constant quantity throughout the study reach, as $Q^*=\left(Q_0^*+Q_1^*\right)/2$, to calculate the cross-sectional average velocity $u_m^*=Q^*/A_m^*$, which is also a constant quantity throughout the study reach.

2.4 Solute residence time in the storage zone

The solute residence time in the storage zone (Thackston and Schnelle 1970, Harvey *et al.* 1996) in the dimensionless form is computed as $t_s = k_r / Da_{em}$, where t_s is the storage zone residence time.

2.5 Analytical solution

The solution of initial-boundary value problem (3)-(8) is obtained by means of Laplace transforms. The Laplace transform of a function $f(t)$ is defined as

$$\overline{f}(s) = L\{f(t); t \to s\} = \int_0^\infty f(t)\exp(-st)\,\mathrm{d}t\,, \tag{9}$$

where s is the Laplace variable.

By using the initial condition (6) in the Laplace transform of eq. (4), we get

$$\overline{c}_s = \frac{c_{s\text{init}} + (Da_{em}/k_r)\overline{c}_m}{s + (Da_{em}/k_r)}\,, \tag{10}$$

where $\overline{c}_m$ and $\overline{c}_s$ are the Laplace transforms of c_m and c_s, respectively. Applying Laplace transform to eq. (3) and using eq. (10), we obtain a linear nonhomogeneous differential equation as

$$\frac{1}{Pe_m}\frac{\mathrm{d}^2\overline{c}_m}{\mathrm{d}x^2} - \frac{\mathrm{d}\overline{c}_m}{\mathrm{d}x} - R\overline{c}_m = -c_{m\text{init}} - \frac{Da_{em}c_{s\text{init}}}{s + (Da_{em}/k_r)} - \frac{q_L c_L}{s}\,, \tag{11}$$

where

$$R = s + q_L + Da_{em} - \frac{Da_{em}^2}{k_r\left[s + (Da_{em}/k_r)\right]}\,.$$

Using the boundary conditions (7) and (8) in the solution of eq. (11), we obtain

$$\overline{c}_m(x,s) = \overline{c}_u(s)\exp(m_1 x) + \left(\frac{c_{m\text{init}}}{R} + \frac{Da_{em}c_{s\text{init}}}{R\,k_r\left[s + (Da_{em}/k_r)\right]} + \frac{q_L c_L}{R\,s}\right)\left[1 - \exp(m_1 x)\right], \tag{12}$$

where $\overline{c}_u(s)$ is the Laplace transform of $c_u(t)$ and

$$m_1 = (Pe_m/2) - \sqrt{Pe_m}\sqrt{(Pe_m/4) + R}\,.$$

With the help of convolution theorem (given in Carslaw and Jaeger 1959, Sneddon 1972, p. 228), the inverse Laplace transform of eq. (12) gives

$$c_m(x,t) = \int_0^t \left\{ c_u(\eta) f(x,t-\eta) + h(x,\eta)\left[g(a,b) + c_L q_L J(a,b) \right] \right\} \mathrm{d}\eta + c_{m\mathrm{init}} \exp(-Da_{em} t)\, h(x,t) \,, \tag{13}$$

where:

$$a = Da_{em}\eta, \quad b = \frac{Da_{em}}{k_r}(t-\eta),$$

$$f(x,t-\eta) = c_0(x,t-\eta)\exp\left[-Da_{em}(t-\eta)\right] + \left(Da_{em}/k_r\right) \times \int_0^{t-\eta} c_0(x,\gamma)\exp(-a_1-b_1)\sqrt{a_1/b_1}\, I_1\left(2\sqrt{a_1 b_1}\right)\mathrm{d}\gamma \,, \tag{14}$$

$$a_1 = Da_{em}\gamma, \quad b_1 = \left(Da_{em}/k_r\right)(t-\gamma-\eta),$$

I_1 is the modified Bessel function of the first kind and of order one,

$$h(x,t) = \exp(-q_L t)\left[1 - c_1(x,t)\right] \tag{15}$$

$$g(a,b) = \exp(-a-b)\left[c_{m\mathrm{init}}\left(Da_{em}/k_r\right)\sqrt{a/b}\, I_1\left(2\sqrt{ab}\right) + c_{s\mathrm{init}}\, Da_{em}\, I_0\left(2\sqrt{ab}\right)\right] \,, \tag{16}$$

I_0 is the modified Bessel function of the first kind and of order zero,

$$c_0(x,t) = \frac{x\sqrt{Pe_m}}{2\sqrt{\pi t^3}} \exp\left(-\frac{Pe_m (x-t)^2}{4t} - q_L t \right) \,, \tag{17}$$

$$c_1(x,t) = \frac{1}{2}\left[\mathrm{erfc}\left(\frac{\sqrt{Pe_m}(x-t)}{2\sqrt{t}} \right) + \exp(Pe_m x)\,\mathrm{erfc}\left(\frac{\sqrt{Pe_m}(x+t)}{2\sqrt{t}} \right)\right] \,, \tag{18}$$

erfc(.) is the complementary error function and is expressed as

$$\mathrm{erfc}(y) = 1 - \mathrm{erf}(y) = 1 - \frac{2}{\sqrt{\pi}} \int_0^y \exp\left(-z^2\right)\mathrm{d}z \,,$$

and $J(a,b)$ is the Goldstein-function (Goldstein 1953) which is defined as

$$J(a,b) = 1 - \exp(-b)\int_0^a \exp(-\gamma)\, I_0\left(2\sqrt{b\gamma}\right)\mathrm{d}\gamma \,. \tag{19}$$

The analytical solution for storage zone is obtained by taking the inverse Laplace transform of eq. (10) as

$$c_s(x,t) = c_{sinit} \exp\left[-\left(Da_{em}/k_r\right)t\right] + \frac{Da_{em}}{k_r}\int_0^t c_m(x,\eta)\exp\left[-\left(Da_{em}/k_r\right)(t-\eta)\right]\mathrm{d}\eta \,. \quad (20)$$

The expressions given in eqs. (13) and (20) are the analytical solutions of the TSM, i.e., eqs. (3) and (4), with lateral inflow for conservative solute.

The analytical solution for classical advection-dispersion equation with the lateral inflow in the absence of storage zone, i.e., $Da_{em} = 0$, can be obtained from eq. (13) as

$$c_m(x,t) = \int_0^t \left[c_u(\eta)c_0(x,t-\eta) + h(x,\eta)\right]\mathrm{d}\eta + c_{minit}\, h(x,t) \,. \quad (21)$$

3. RESULTS AND DISCUSSION

In order to illustrate the use of present analytical solution, the physical transport parameters are estimated for Uvas Creek tracer experiment described in the work of Avanzino *et al.* (1984) and Bencala and Walters (1983). Results obtained from the present analytical solution are compared with the observed data, and a sensitivity analysis is reported that illustrates to which parameter the concentrations are most and least sensitive. Finally, a hypothetical situation is considered to discuss the effects of background concentrations in the main channel, the storage zone and the lateral inflow, as well as the lateral inflow rate. Integrations appearing in the different analytical expressions are evaluated by Simpson's rule. The Bessel functions I_0 and I_1 are approximated by the polynomials given in Carslaw and Jaeger (1959). The polynomial approximation of Goldstein J function as given in De Smedt and Wierenga (1979) is used.

3.1 Application to tracer experiment

In order to illustrate the use of analytical solutions, we used experimental data from Uvas creek tracer experiment (Avanzino *et al.* 1984). In this experiment, a concentrated solution of dissolved sodium chloride was added to the stream at 50 ml/min for 3 hours. The conservative solute (chloride) concentration was 11.7 mg/l at the injection location. The injection duration was three hours. The volumetric flow rate was 0.0125 m^3/s at the injection location and the background concentration in the main channel, background concentration in the storage zone and the lateral inflow concentration were 3.7 mg/l. The concentration-time breakthrough curves were measured at different downstream locations from the injection location (i.e., $x^* = 38$, 105, 281, 433 and 619 m). The solute concentration profile at upstream boundary

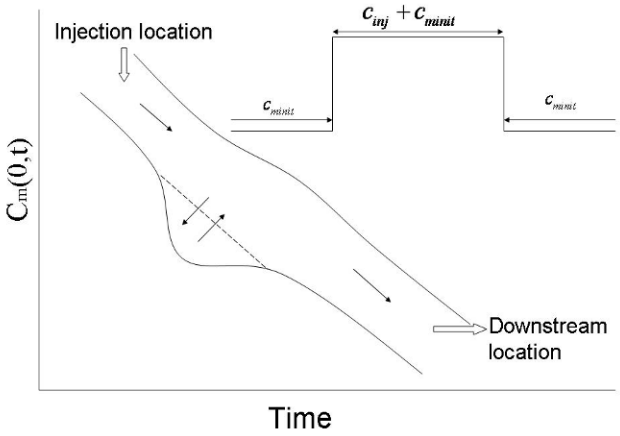


Fig. 3. Sketch of the solution domain, step-concentration profile as an upstream boundary condition.

location ($x_0 = 0$) is chosen to be a step-concentration profile (as shown in Fig. 3), to simulate the observed concentration profiles at each downstream location by the analytical solution given in eqs. (13) and (21). Parameters $D_m^*, A_m^*, A_s^*, q_L^*, \alpha^*$ are estimated by minimizing the sum of the squared differences between the observed and theoretical result. The large scale Newton reflexive method (Coleman and Li 1994, 1996) is used for the parameter estimation. Estimated values of the dimensional parameters are given in Table 1.

The estimated value of lateral inflow q_L^* is increased in the downstream sampling locations as given in Table 1. Consequently, there is an increase in the value of volumetric flow rate, which acts to dilute solute concentrations. The lateral inflow is in good agreement with the calculated value of lateral

Table 1

Estimated values of parameters for observed chloride concentrations for the Uvas Creek tracer experiment for TSM

x^* [m]	A_m^* [m^2]	Q^* [m^3/s]	D_m^* [m^2/s]	$q_L^*(\times 10^{-6})$ [m^3/s m]	A_s^* [m^2]	$\alpha^*(\times 10^{-6})$ [1/s]
38	0.3496	0.0125	0.015785	1.64	0	0
105	0.3880	0.0126	0.08295	1.5	0	0
281	0.3771	0.0136	0.1217	8.3	0.0689	7.5
433	0.4201	0.0141	0.181	7.556	0.0999	8.1667
619	0.4897	0.0154	0.1798	9.34	0.3822	0.14

inflow from the mass balance equation (as described in Section 2.3). Thus, the lateral inflow is accurately estimated. The Péclet number in the last sampling location (619 m) is lower than the Péclet number in the first sampling location (38 m), which implies the spread of solute increases in the longitudinal direction of stream. A similar observation can be noticed from the observed data (which can also be seen from Fig. 4) and there is also a good agreement with those presented in the work of Bencala and Walters (1983). So, the Péclet number is accurately estimated. The stream storage exchange coefficient, which represents interaction between the main channel and the storage zone, increases from sampling location 281 m to sampling location 433 m and decreases from sampling location 433 m to sampling location 619 m. The values of cross-sectional areas are increased in the downstream location as compared to the upstream location, whereas all the parameters are found to be nearly constant or in the same order of magnitude along the downstream direction of the river, which confirms the validity of the assumption made for the mathematical models. The values of parameters also follow a similar trend to that presented in the work of Bencala and Walters (1983). Using the estimated values of parameters, results are calculated by eqs. (13) and (21) in the dimensionless form and presented in dimensional form in Fig. 4. It shows that the analytical concentration–time curves agree well with the observed concentration–time curves of Uvas Creek tracer experiment.

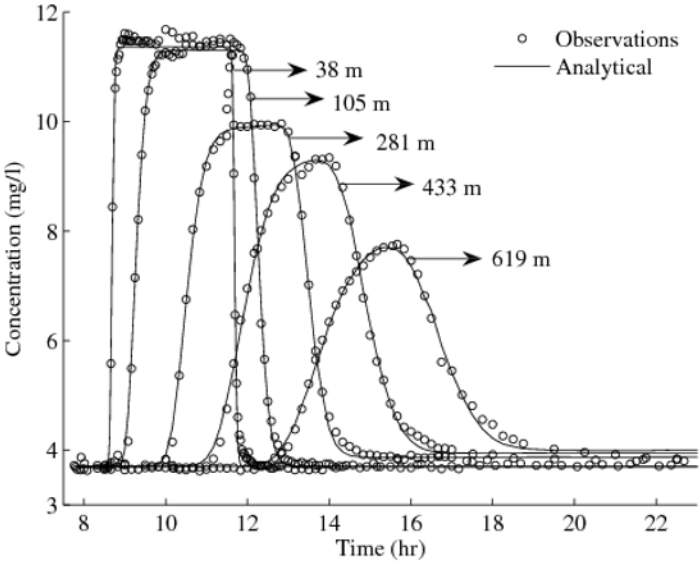


Fig. 4. Analytical and observed chloride concentration–time breakthrough curves at five sampling locations for the Uvas Creek tracer experiment.

Sensitivity analysis

A sensitivity analysis has been performed in order to identify the most critical parameter for predicting concentrations for the present situation. Sensitivity is the partial derivative of the simulated stream tracer concentration with respect to a change in the value of a parameter (Wagner and Harvey 1997)

$$s_{ij} = \frac{\partial c_{m_i}}{\partial p_j}, \tag{22}$$

where s_{ij} is the sensitivity of simulated stream tracer concentration c_{m_i} to the parameter p_j. Sensitivity of Péclet number, exchange Damköhler number in the main stream, ratio of cross-sectional areas lateral inflow rate, and stream storage exchange coefficient are calculated at each sampling location. We have calculated the minimum and maximum values of sensitivity of solute concentration in the main channel to each parameter. Results are presented in Table 2 for each sampling location, but the plots are shown only at 433 m for the most Da_{em} and the least k_r sensitive parameter (Fig. 5b,a). It was found that, among these parameters, concentrations were most sensitive to the lateral inflow rate q_L and were least sensitive to the ratio of the cross-sectional areas k_r (which can be seen from Table 2). The storage zone parameters (Da_{em}, α^*, k_r) are the most sensitive at 433 m. This represents that the change in solute concentration due to the presence of the storage zone is prominent at 433 m compared to that at other locations.

Table 2

Difference between the minimum and maximum values of sensitivity of solute concentration in the main channel to the parameters p_j

x \ p_j	Pe_m	q_L	Da_{em}	k_r	α^*
38	0.1033	37.85	–	–	–
105	0.7013	103.63	–	–	–
281	0.7492	225.77	260.78	0.0828	1.9988
433	1.18	310.43	379.30	0.31	3.1320
619	1.2165	403.29	174.45	4.7×10^{-6}	1.5419

3.2 Hypothetical experiment

In order to study the effects of background concentrations in the main channel and the storage zone, the lateral inflow rate and its concentration, a hypothetical situation was considered based on the values of the parameters given

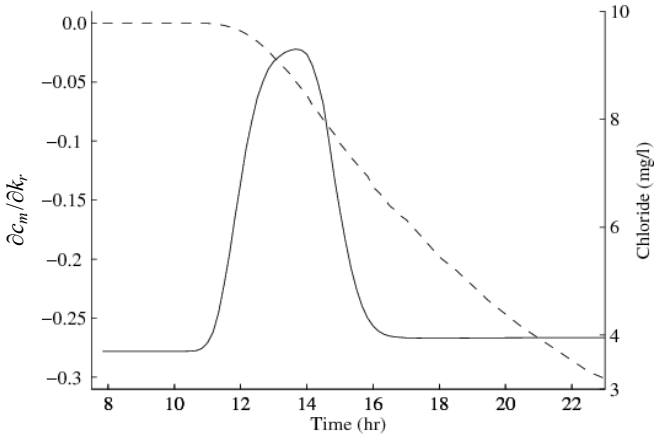


Fig. 5a. Sensitivity of analytical chloride concentration–time breakthrough curve (solid line) to ratio of cross-sectional areas (dashed line) at 433 m.

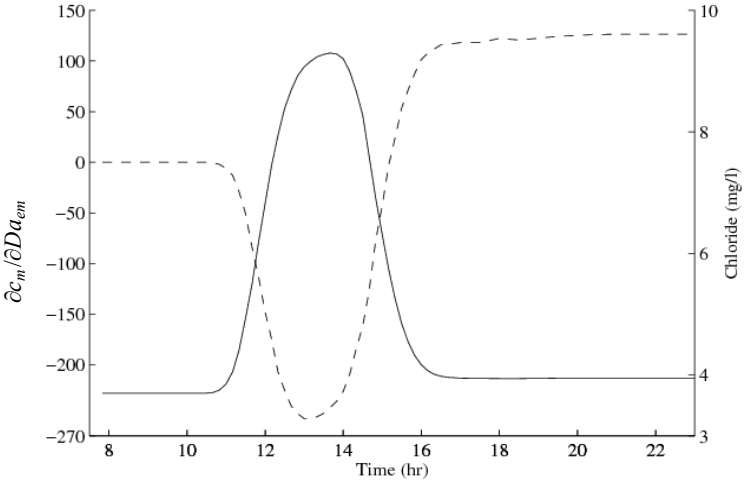


Fig. 5b. Sensitivity of analytical chloride concentration–time breakthrough curve (solid line) to exchange Damköhler number in the main stream (dashed line) at 433 m.

in Table 3. It was assumed that 25 units of conservative solute were injected at the injection location $x_0 = 0$ with a constant rate over time interval of 252 units.

Results are obtained at 1000 units downstream from the injection location. The background concentrations and lateral inflow solute concentration are considered to be constant and can be chosen independently in the present analytical solution (i.e., they need not be equal).

Table 3

Dimensionless parameter values for numerical experiments

c_{inj}	x	Q_0^*	Pe_m	Da_{em}	k_r	q_L
25	1000	10	0.21	9.5238×10^{-4}	0.2	4.7619×10^{-5}

Equal background and lateral inflow concentrations

In order to discuss the effects of lateral inflow rate and exchange Damköhler number, the background concentrations and lateral inflow concentration are considered to be equal to 5 units. The breakthrough curves are presented in Fig. 6. If lateral inflow solute enters into the main stream of river due to lateral inflow rate, then the solute mass is reduced at a location (c_2, c_4, c_6) compared to the case of no lateral inflow (c_1, c_3, c_5). When a solute is injected into the main stream of river, a portion of it initially enters into the storage zone and slowly is released back in the main stream of the river. This exchange of mass is the main cause of reduction in the solute mass and also forms a long tail in the downstream edge of the solute concentration profiles in the main stream ($c_3 - c_4$) at a downstream location compared to the case in which the storage zone is absent ($c_1 - c_2$). Figure 6 also shows that the effects of lateral inflow are not influenced by the presence of dead zone ($c_5 - c_6$).

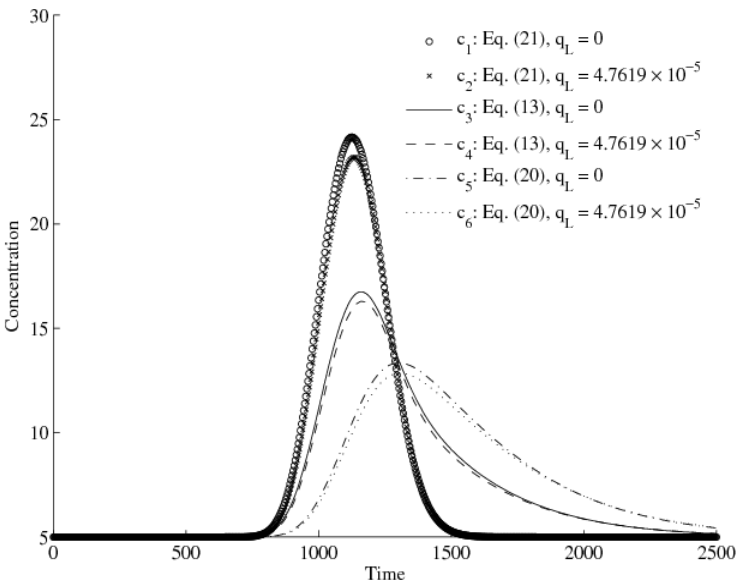


Fig. 6. Effects of lateral inflow rate and exchange rate on solute concentration at 1000 units downstream from the injection location.

Unequal background and lateral inflow concentrations

In this subsection, the effects of lateral inflow solute concentration and background concentrations are discussed for the cases where these concentrations are constant and unequal. In order to discuss the effects of lateral inflow concentration, main stream solute concentrations are calculated for different values of lateral inflow concentration considering the background concentrations in the main channel and in the storage zone as 5 units; the results are presented in Fig. 7.

The increase in the lateral inflow concentration will increase the background concentration in the main channel which, in turn, raises the concentration profiles ($c_2 - c_3$, $c_5 - c_6$), which can be seen from Fig. 7. By comparing Figs. 6 and 7, it is clear that the main stream concentration gets diluted or densified, respectively, depending on whether the lateral inflow concentration value is less than or greater than a particular fixed value. This fixed value is the lateral inflow concentration that does not change the peak concentration value of the breakthrough curve obtained in the case of no lateral inflow. The fixed value depends on the set of input data. In the present situation, the fixed value is found to be 11 unit. Due to the presence of dead zone, the fixed concentration value is reduced from 11 to 10.4 unit. It has been observed that there is a sharp decrease in breakthrough curve (Fig. 7: $c_1 - c_3$) after reaching its peak value when the dead zone is absent whereas in the presence of dead zone this decrease is much slower. The slow decrease in

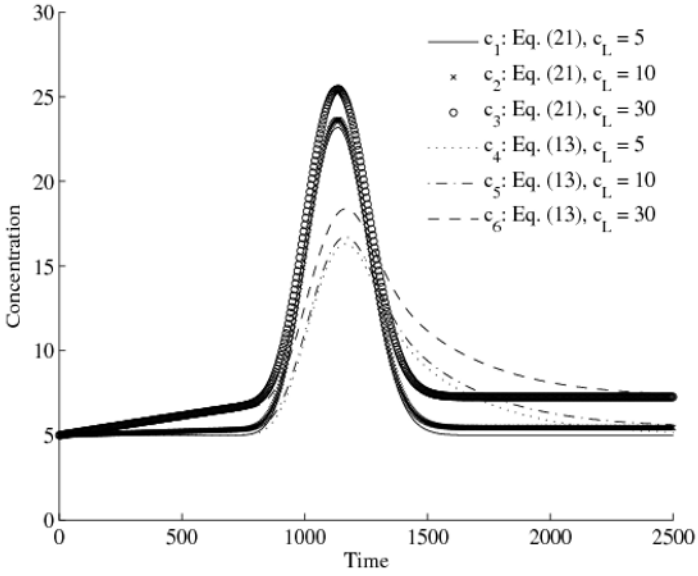


Fig. 7. Effects of lateral inflow rate and exchange rate on solute concentration at 1000 units downstream from the injection location.

the breakthrough curve (Fig. 7: $c_4 - c_6$) is due to the exchange Damköhler number, which does not allow the solute to cross the sampling location very fast with time.

It has been noticed that the effects of the lateral inflow on the storage zone solute concentration profiles are similar to the profiles shown in Fig. 7 (no figure is presented for this case).

The effects of main stream background concentration are shown in Fig. 8 by considering that the lateral inflow concentration and the background concentration in the storage zone are equal to 5 units. The solute concentration profiles for larger values of background concentrations ($c_2 - c_3$, $c_5 - c_6$) are raised up the graph due to the increasing background concentration in the main channel. The effects of exchange Damköhler number on breakthrough curves ($c_4 - c_6$) are similar to the curves ($c_3 - c_4$) shown in Fig. 6.

Effects of c_{minit} on c_s are shown in Fig. 9. Due to increase in the main stream background concentrations, the initial solute concentration in the storage zone increases and reaches up to the main stream background concentration. As time elapses, the solute concentrations get delayed and the tails in the storage zone concentration profiles (Fig. 9: $c_2 - c_3$) become more pronounced compared to the main stream concentration profiles (Fig. 8: $c_5 - c_6$).

In order to describe the effect of storage zone background concentration, breakthrough curves are drawn for different values of background concentra-

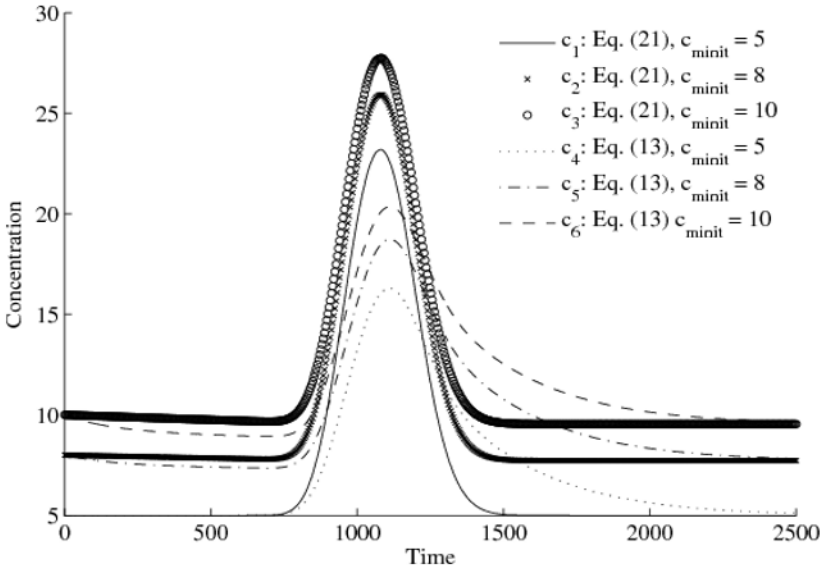


Fig. 8. Effects of main stream background concentration at 1000 units downstream from the injection location in the main stream of the river.

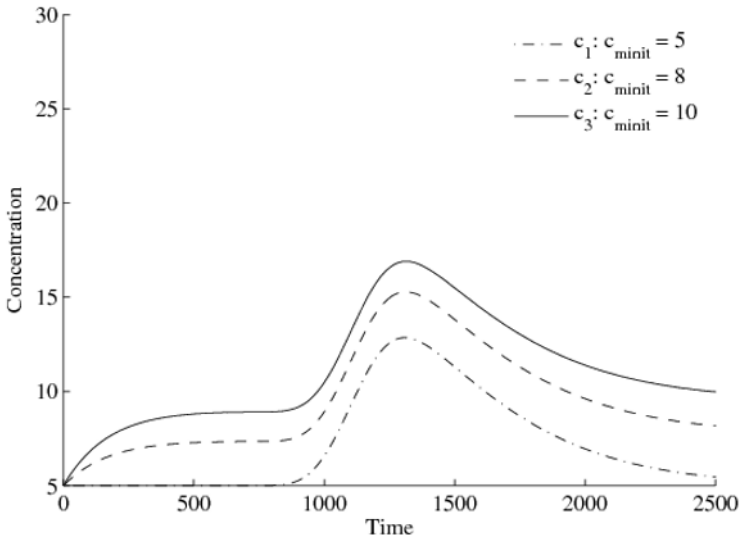


Fig. 9. Effects of main stream background concentration at 1000 units by analytical solution (20) in the storage zone.

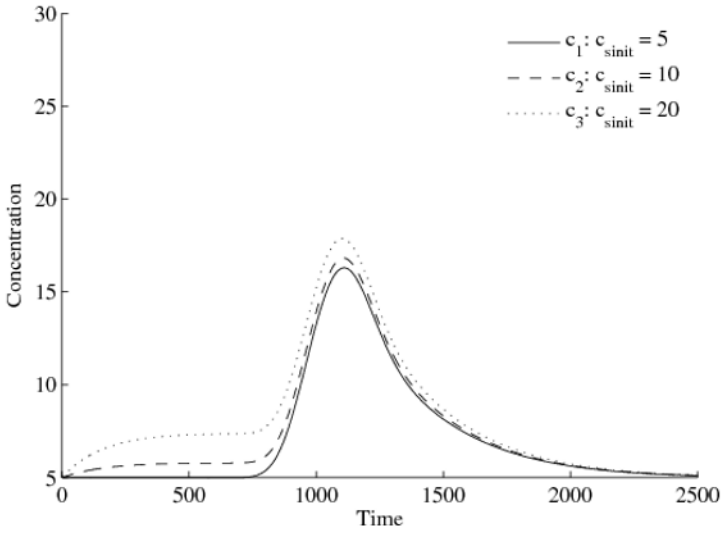


Fig. 10. Effects of the storage zone background concentration at 1000 units downstream from the injection location by analytical solution (13) in the main stream of the river.

tion and presented in Figs. 10 and 11. In these cases, the background concentration in the main stream and lateral inflow concentration are considered to be 5 unit. It is noticed that the solute concentration in the main stream in-

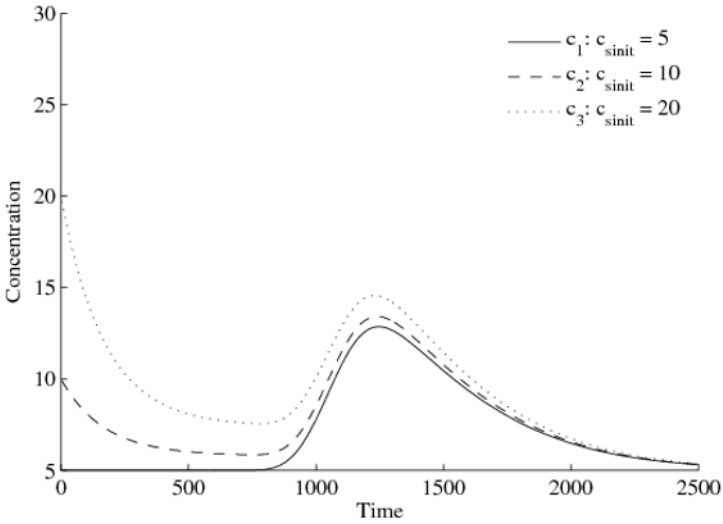


Fig. 11. Effects of the storage zone background concentration at 1000 units downstream from the injection location by analytical solution (20) in the storage zone.

creases with increasing storage zone background concentration (see Fig. 10: $c_2 - c_3$). The response of the storage zone concentration is a little more complex (see Fig. 11: $c_2 - c_3$). Since the initial value is relatively high, concentrations decrease as solute diffuses into the main channel until the arrival of the patch of solute being transported down the hypothetical stream. This reduction is steeper, but does not reach so low a minimum value, as the storage zone background concentration increases. Figure 10 ($c_2 - c_3$) and Fig. 11 ($c_2 - c_3$) show that the concentration-time distribution curves merge at a longer time with the curve c_1 of Figs. 10 and 11. It shows that the patch of solute passed the particular location as time goes on.

4. CONCLUSIONS

A new analytical solution for conservative solute concentrations in rivers with lateral inflow and transient storage is presented. The existing solution for the case without transient storage can easily be obtained from the new one. The solution for concentration in the main channel of the river is applied to the analysis of the Uvas Creek tracer experiment. A sensitivity analysis for this case is also carried out in order to study the sensitivity of predicted concentrations to controlling parameters. Finally, a hypothetical case is considered to investigate the effects of background concentrations in the main channel, the storage zone and the lateral inflow.

The analytical breakthrough curves are found to be in good match with observed breakthrough concentrations of the Uvas Creek tracer experiment.

The sensitivity analysis shows that concentrations are most sensitive to the lateral inflow rate q_L and least sensitive to the ratio of cross-sectional areas k_r. Sensitivity analysis also shows that the storage zone parameters are most sensitive at 433 m.

It has been noticed that lateral inflow rate dilutes the solute concentration at a particular location. Due to the exchange Damköhler number, the solute mass gets reduced and long tail appears in the downstream edge of the concentration profile due to the slow release of solute mass from the storage zone into the main stream. Effects of lateral inflow rate in the case of unequal background and lateral inflow concentrations are similar to the case of equal background and lateral inflow concentrations, however the effects of exchange coefficients in the case of unequal background and lateral inflow concentrations are not similar to the case of equal background and lateral inflow concentrations. Present analytical solutions clearly reveal the effects of lateral inflow rate and exchange Damköhler number.

The solution allows one to determine the fixed concentration of lateral inflow. On the basis of this fixed concentration, it can be verified that the main stream solute concentration is diluted or concentrated if the concentration value is less or greater than the fixed value, respectively. Due to an increase in the main stream background concentrations, the long tail in the breakthrough curves of the storage zone concentration profiles is more pronounced compared to that of main stream concentration profiles. This indicates that solute concentrations can remain in the storage zone long after the patch of solute has cleared the main stream. The effect of storage zone background concentration on the breakthrough curves is prominent over a longer period of time but at longer time effect of storage zone background concentration on the breakthrough curves is negligible. This means that the patch of solute has cleared the main stream as well as the storage zone. The present analytical solutions clearly reveal the effects of background concentrations in the main channel, the storage zone, and the lateral inflow. It can be concluded that our analytical solutions can be applied for equal as well as for unequal solute concentrations, in the main stream background, in the storage zone background, and the lateral inflow. It can also be concluded that solutions can be used for the analysis of solute transport in tracer experiments.

References

Avanzino, R.J., G.W. Zellweger, V.C. Kennedy, S.M. Zand, and K.E. Bencala (1984), Results of a solute transport experiment at Uvas Creek, September 1972, *US Geol. Surv. Water Resour. Invest*. Rep. 84-236, US Geol. Surv., Menlo Park, CA.

Bencala, K.E., and R.A. Walters (1983), Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model, *Water Resour. Res*. **19**, 3, 718-724, DOI: 10.1029/WR019i003p00718.

Carslaw, H.S., and J.C. Jaeger (1959), *Conduction of Heat in Solids*, 2nd ed., Clarendon Press, Oxford.

Choi, J., J.W. Harvey, and M.H. Conklin (2000), Characterizing multiple timescales of stream and storage zone interaction that affect solute fate and transport in streams, *Water Resour. Res*. **36**, 6, 1511-1518, DOI: 10.1029/2000WR 900051.

Coleman, T.F., and Y. Li (1994), On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds, *Math. Program.* **67**, 1-3, 189-224, DOI: 10.1007/BF01582221.

Coleman, T.F., and Y. Li (1996), An interior trust region approach for nonlinear minimization subject to bounds, *SIAM J. Optim.* **6**, 418-445, DOI: 10.1137/ 0806023.

Czernuszenko, W., and P.M. Rowiński (1997), Properties of the dead-zone model of longitudinal dispersion in rivers, *J. Hydraul. Res.* **35**, 4, 491-504.

Davis, P.M., T.C. Atkinson, and T.M.L. Wigley (2000), Longitudinal dispersion in natural channels: 2. The roles of shear flow dispersion and dead zones in the River Severn, UK, *Hydrol. Earth Syst. Sci*. **4**, 3, 355-371.

De Smedt, F. (2006), Analytical solutions for transport of decaying solutes in rivers with transient storage, *J. Hydrol.* **330**, 3-4, 672-680, DOI: 10.1016/ j.jhydrol.2006.04.042.

De Smedt, F., and P.J. Wierenga (1979), A generalized solution for solute flow in soils with mobile and immobile water, *Water Resour. Res.* **15**, 5, 1137-1141, DOI: 10.1029/WR015i005p01137.

De Smedt, F., W. Brevis, and P. Debels (2005), Analytical solution for solute transport resulting from instantaneous injection in streams with transient storage, *J. Hydrol.* **315**, 1-4, 25-39, DOI: 10.1016/j.jhydrol.2005.04.002.

Fernald, A.G., P.J. Wigington Jr., and D.H. Landers (2001), Transient storage and hyporheic flow along the Willamette River, Oregon: Field measurements and model estimates, *Water Resour. Res*. **37**, 6, 1681-1694, DOI: 10.1029/ 2000WR900338.

Ge, Y., and M.C. Boufadel (2006), Solute transport in multiple-reach experiments: Evaluation of parameters and reliability of prediction, *J. Hydrol*. **323**, 1-4, 106-119, DOI: 10.1016/j.jhydrol.2005.08.021.

Goldstein, S. (1953), On the mathematics of exchange processes in fixed columns. I. Mathematical solutions and asymptotic expansions, *Proc. R. Soc. London A* **219**, 1137, 151-171.

Hart, D.R. (1995), Parameter estimation and stochastic interpretation of the transient storage model for solute transport in streams, *Water Resour. Res*. **31**, 2, 323-328, DOI: 10.1029/94WR02739.

Harvey, J.W., B.J. Wagner, and K.E. Bencala (1996), Evaluating the reliability of the stream traces approach to characterize stream-subsurface water exchange, *Water Resour. Res*. **32**, 8, 2441-2451, DOI: 10.1029/96WR01268.

Hays, J.R., P.A. Krenkel, and K.B. Schnelle (1966), Mass transport mechanisms in open-channel flow, *Sanitary and Water Resources Engineering*, Dept. Civil Engng., Techn. Rep. 8, Vanderbilt University, Nashville, TE.

Runkel, R.L. (1998), One-dimensional transport with inflow and storage (OTIS): A solute transport model of streams and rivers, US Geol. Surv. Water Resour. Invest. Rep. 98-4018.

Runkel, R.L., and S.C. Chapra (1993), An efficient numerical solution of the transient storage equations for solute transport in small streams, *Water Resour. Res*. **29**, 1, 211-215, DOI: 10.1029/92WR02217.

Runkel, R.L., and S.C. Chapra (1994), Reply, *Water Resour. Res*. **30**, 10, 2863-2865, DOI: 10.1029/94WR00986.

Schmid, B.H. (1997), Analytical solution of the transient storage equations accounting for solute decay, Proc. 27th IAHR Congress, Water Resour. Engng. Div., August 10-15, San Francisco, CA, **B** 15-20.

Schmid, B.H. (2003), Temporal moments routing in streams and rivers with transient storage, *Advances in Water Resour.* **26**, 9, 1021-1027, DOI: 10.1016/S0309-1708(03)00086-1.

Scott, D.T., M.N. Gooseff, K.E. Bencala, and R.L. Runkel (2003), Automated calibration of a stream solute transport model: Implications for interpretation of biogeochemical parameters, *J. N. Am. Benthol. Soc.* **22**, 4, 492-510, DOI: 10.2307/1468348.

Sneddon, I.N. (1972), *The Use of Integral Transforms*, McGraw-Hill Publ., New Delhi, 539 pp.

Thackston, E.L., and K.B. Schnelle (1970), Predicting effects of dead zones on stream mixing, *J. Sanit. Eng. Div., Am. Soc. Civ. Eng.* **96**, SA2, 319-331.

Wagner, B.J., and J.W. Harvey (1997), Experimental design for estimating parameters of rate-limited mass transfer: analysis of stream tracer studies, *Water Resour. Res*. **33**, 7, 1731-1741, DOI: 10.1029/97WR01067.

Wörman, A. (1998), Analytical solution and timescale for transport of reacting solutes in rivers and streams, *Water Resour. Res.* **34**, 10, 2703-2716, DOI: 10.1029/98WR01338.

Received 29 April 2009
Received in revised form 12 November 2009
Accepted 27 November 2009

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1115-1133
DOI: 10.2478/s11600-010-0026-5

Seasonal Temperature Extremes in Potsdam

Zbigniew W. KUNDZEWICZ[1,2] and Shaochun HUANG[1]

[1]Potsdam Institute for Climate Impact Research, Potsdam, Germany
e-mails: zbyszek@pik-potsdam.de, huang@pik-potsdam.de

[2]Institute for Agricultural and Forest Environment, Polish Academy of Sciences, Poznań, Poland, e-mail: zkundze@man.poznan.pl

Abstract

The awareness of global warming is well established and results from the observations made on thousands of stations. This paper complements the large-scale results by examining a long time-series of high-quality temperature data from the Secular Meteorological Station in Potsdam, where observation records over the last 117 years, *i.e.*, from January 1893 are available. Tendencies of change in seasonal temperature-related climate extremes are demonstrated. "Cold" extremes have become less frequent and less severe than in the past, while "warm" extremes have become more frequent and more severe. Moreover, the interval of the occurrence of frost has been decreasing, while the interval of the occurrence of hot days has been increasing. However, many changes are not statistically significant, since the variability of temperature indices at the Potsdam station has been very strong.

Key words: temperature extremes, seasonality, climate variability, climate change.

1. INTRODUCTION

The time series of global mean air temperature, compiled by the Climatic Research Unit of the University of East Anglia, jointly with the UK Met Office Hadley Centre (*cf.* http://www.cru.uea.ac.uk/cru/data/temperature/, Brohan *et al.* 2006) or in NASA in the USA (http://data.giss.nasa.gov/gistemp/), convincingly illustrate global warming.

As noted in IPCC (2007), warming of the global climate system is unequivocal. This is now evident from observations of increases in air temperature, which show clear growing trends at a range of scales, from local, *via* regional, to continental, hemispheric, and global. The updated 100-year linear trend (1906 to 2005) reflects a 0.74°C (0.56 to 0.92°C) global mean temperature increase, while global warming rates over the last 50 years and over the last 25 years were much stronger (0.128°C/decade and 0.177°C/decade, respectively). That is, the global warming rate over the last 25 years is over 2.4 times faster than it was over the last 100 years.

As reported by Trenberth *et al*. (2007), there has been a widespread increase in the number of warm nights between 1951 and 2003, and a decrease in the number of cold nights. Trends in the number of cold and warm days are also consistent with warming, but are less marked than at night. This is a general tendency, yet there are regional differences. Over the last half-century, nearly two-thirds of the global land area have experienced a significant decrease in the annual occurrence of cold nights, while a significant increase in the annual occurrence of warm nights also took place at nearly two-thirds of the global land area. The distributions of minimum and maximum temperatures have not only shifted to higher values, consistent with overall warming, but the cold extremes have warmed more than the warm extremes. More warm extremes imply an increased frequency of heat waves. Associated with the warming there has been a global trend towards fewer frost days.

Although global warming is unabated, one cannot claim that the values of the annual global mean temperature since 1850 are known with good accuracy. Indeed, the uncertainty range has been considerable (but not overshadowing global warming) and has varied with time, being highest in the 1850s and lowest in the 1980s. Since the 1980s, the ground observation networks have been shrinking in many areas, so that recently the uncertainty does not decrease with time.

The paper complements the large-scale aggregate results by demonstrating tendencies for a long time series of good-quality instrumental observation records. It examines the details of seasonal warming *via* analysis of temperature-related climate extremes in the unique long-term gap-free record from Potsdam, from January 1893 to February 2009.

2. DATA

Since the accuracy and homogeneity of a long time series of records of temperature observations is often problematic, it is essential to look for data from highest-quality stations, where the time series of records are long, reliable and gap-free, and where changes of location, surrounding environment, instruments, observation principles and methods, are limited. Such condi-

tions are not easy to find, but they are fulfilled at the Secular Meteorological Station in Potsdam (Germany).

The Potsdam Station (co-ordinates: 52°23′N, 13°04′E, elevation 81 m AMSL) is located to the south-west of town in Potsdam, approximately 600 m away from the built-up area, so that the urban heat island effect is not present there. It is a notable station, probably the only meteorological observatory, world-wide, with uninterrupted observations of many variables carried out every day since 1 January 1893. The Secular Meteorological Station in Potsdam was established with the purpose of serving for a long time (the word *saeculum* means longevity in Latin). The manned observations were continued even during nearly all (except for only three) the days of World War II. Since measurements from self-recording instruments were available and incorporated into the long time series of records, there have been no gaps in the data, even in 1945.

Efforts have been made to keep the observation conditions homogeneous, by maintaining the station location (on an empty plot in Telegrafenberg in Potsdam, in a considerable distance from buildings and trees), the character of the environment (*e.g.*, removing any tree seedlings from station's environment), and methods and principles of instrumental observation (*cf.* Lehmann and Kalb 1993).

Besides air temperature (mean, minimum, maximum), many other meteorological variables are being measured at the station. These include soil temperature, air pressure, relative air moisture contents, water vapour pressure, wind, precipitation, cloudiness, snow cover, frost depth, and sunshine hours. Great efforts have been made to keep the observation conditions homogeneous, by maintaining the station location, conditions of the environment, methods and principles of instrumental observation. Measurements at the station have been carried out three times a day (7:08, 14:08, and 21:08 CET, *i.e.*, UTC + 1 hour). The daily mean temperature is calculated as $(T_7 + T_{14} + 2T_{21})/4$, where T_N represents air temperature at hour N.

In the present internet era, open access to observation records on the web is very important for scientists, decision makers and the broader public alike, and contributes in a substantial way to the awareness on climate variability and change. Time series of daily meteorological records from Potsdam, extending since 1 January 1893, are freely available on the web portal: http://www.klima-potsdam.de/, together with comprehensive information about the station. The station has international reputation and its website has been frequently visited in Germany and abroad (*cf.* Kundzewicz *et al.* 2007, Kundzewicz and Józefczyk 2008).

The diagram of the mean annual temperature observed at Potsdam shows a clear increasing trend (Fig. 1), and the rate of increase grows with time. The slope of the regression line for the last 25 years (1984-2008) was

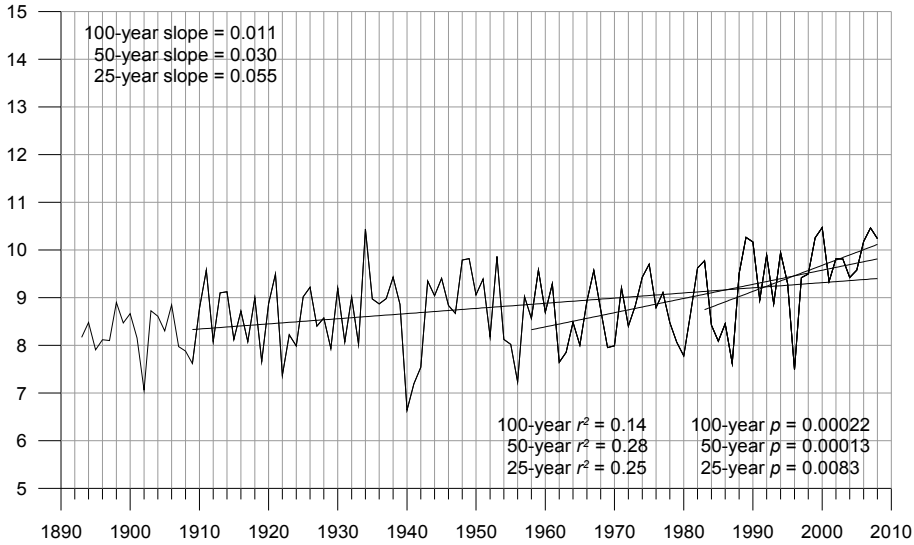


Fig. 1. Changes in mean annual temperature in Potsdam. Notation: r^2 is the correlation coefficient, p is the significance level.

0.55°C/decade, that is nearly twice stronger than during the last 50 years (1959-2008) (0.3°C/decade), and five times stronger than for the last 100 years (1909-2008) (0.11°C/decade). All these changes are statistically significant at the 0.01 level (or better). It should be noted that by shifting the time horizons of regressions in Fig. 2, even slightly, one could get different results. Selection of the last 25, 50, and 100 years for comparison follows the approach taken in IPCC (2007). The present results show that the recent acceleration of warming in Potsdam is much stronger than the global average. There have been eight calendar years on record with a mean annual temperature in excess of 10°C, five of which were in the recent 10 years (Table 1).

Table 1

Mean annual temperature of warmest years in Potsdam

Rank	Year	Mean annual temperature [°C]
1	2000	10.47
2	2007	10.46
3	1934	10.44
4-5	1989, 1999	10.26
6	2008	10.24
7-8	1990, 2006	10.17

The mean annual temperature, which is presented in Table 1, is determined from 365 or 366 values of daily mean temperature (each of which, in turn, is calculated from three values measured every day with 0.1°C accuracy). Presenting of mean annual temperature with 0.01°C resolution allows, for instance, ordering of the years 2000 and 2007. Rounding up to 0.1°C resolution (matching the observation accuracy) would not make it possible to distinguish between the mean annual temperature in these years.

The (upwards) departure from the long-term mean annual temperature in 1934, when the pre-2000 record was settled (the only excursion above 10°C until 1989) was a rare case. But annual mean temperature in excess of 10°C has become much more frequent in the last 20 years (Table 1).

The records of mean annual temperature, presented in Fig. 1, show strong and rapid oscillations. Sudden changes can be noted between two adjacent years, *e.g.*, 1933 (8.02°C) and 1934 (10.44°C), *i.e.*, the difference of 2.42°C, or between 1939 (8.86°C) and 1940 (6.64°C), *i.e.*, the absolute difference of 2.22°C, or – more recently – 1995 (9.29°C), 1996 (7.48°C) or 1997 (9.42°C), *i.e.*, 1.79°C or 1.94°C, respectively. Within the seven-year interval, from 1934 to 1940, the mean annual temperature differed by 3.80°C (from 10.44°C in 1934 down to 6.64°C in 1940).

Departures from the long-term trend, as illustrated by the regression lines, can be strong in individual years. Upwards excursions are a little more frequent, but less pronounced than downward excursions. Deviations of annual mean temperature from the regression line in individual years may even reach 2°C. For instance, the downward excursions from the linear regression in 1940 and 1996 were 2.01°C and 1.78°C, respectively, while the strongest upward excursion in 1934 was 1.85°C.

When looking not only at the annual temperature (in the sense of a calendar year, from 1 January to 31 December), but also at the mean air temperature of any consecutive 12-month period that commences on the 1st of any month, one can find several recent records.

The pre-2007 record of a mean 12-month temperature, 10.70°C (July 1999 – June 2000) has been exceeded six times in 2007, reaching a very high level of 12.09°C in the period from July 2006 to June 2007 (Kundzewicz *et al*. 2007). In the latter record-breaking interval, there were four months with the highest monthly mean temperature ever observed in Potsdam. In July 2006, the mean temperature was 23.69°C (compared to the long-term mean 17.97°C for 1961-1990), in December 2006 it was 5.17°C (long-term mean: 0.69°C), in January 2007, 4.98°C was observed (long-term mean: –0.80°C), and in April 2007, the temperature was 12.02°C (long-term mean: 8.05°C). This last record of highest mean April temperature was broken by more than 1°C in April 2009, to the level of 13.22°C.

After looking at the record of a mean 12-month temperature in 2006/2007 in Potsdam, an analysis was extended to the whole of Germany, most of Europe and the Northern Hemisphere, where records were also detected (Kundzewicz *et al.* 2008).

3. CHANGES IN VALUES OF SEASONAL TEMPERATURE EXTREMES

The long time series of daily minimum and maximum values of temperature in Potsdam were analyzed in the context of seasonal properties for all four seasons. Seasons are defined as MAM (March, April, May) for spring, JJA (June, July, August) for summer, SON (September, October, November) for autumn, and DJF (December, January, February) for winter. Results illustrating seasonal mean of maximum and minimum temperatures for 1893-2008 (in case of winter – including 2009, for 1893 only January and February) are presented in Fig. 2. It shows seasonal warming for all seasons; on average about 1°C/100 years, *i.e.*, 0.1°C/decade. The slope of regression lines for the whole 116-year period varies from 0.0074 (an increase of the maximum temperature for autumn) to 0.0121 (an increase of the minimum temperature for summer). In three seasons, the average increase of the minimum temperature is higher than the average increase of the maximum temperature, except for winter, where the maximum temperature grows slightly faster than the minimum temperature, hence the seasonal amplitude grows. In four cases (spring minimum, summer minimum, summer maximum and autumn minimum) changes are statistically significant at the 0.01 level, while in one case (spring maximum) at the 0.05 level. For three indices (autumn maximum, winter minimum, and winter maximum), changes are not statistically significant at the 0.05 level, while one of them (winter maximum) is nearly significant (0.052).

Climatic time series show strong natural variability (irregular oscillations), which is superimposed on a gradual warming trend. There is a strong random component, so that some record-warm extremes, that occurred a long time ago, have not been exceeded to-date. Cold extremes, even if they occur now, are getting considerably less frequent and less severe. In 1917, the highest monthly mean of daily maximum temperature of June (27.15°C) was observed, even if the climate was then clearly colder than now. Only six years later, in 1923, the ever lowest monthly mean of daily maximum temperature of June (15.62°C) was observed (Kundzewicz and Józefczyk 2008). The warmest spring day ever observed (T_{max} = 34.0°C) occurred on 24 May 1922, the warmest autumn night and day (T_{min} = 18.6°C and T_{max} = 34.0°C) occurred on 4 September 1895. All these "warm" records were set a long time ago, while one of "cold" records – the coldest autumn day observed on 21 November 1993 (T_{max} = –6°C) – was set relatively recently.

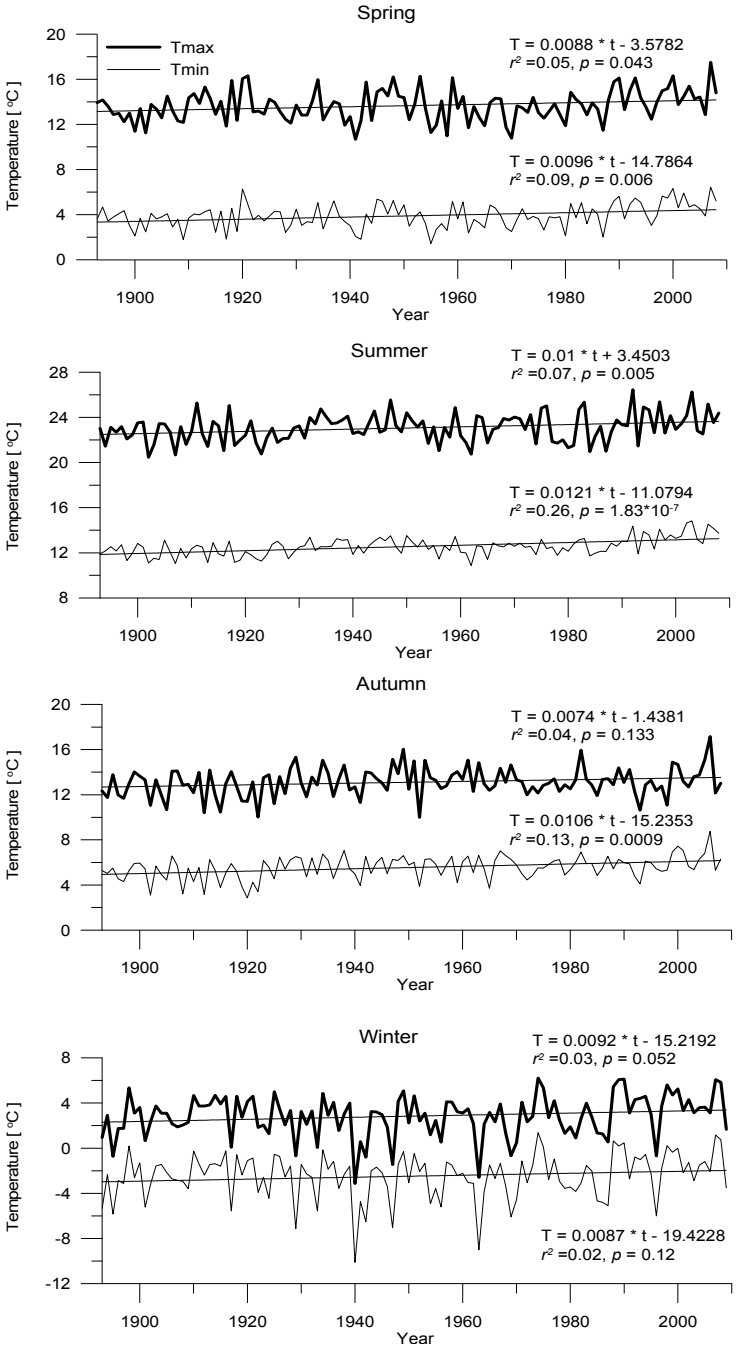


Fig. 2. Changes in seasonal mean of daily minimum and maximum values of temperature for all seasons, in Potsdam from 1893 to 2008.

It is clear that an occurrence of a record-high mean monthly temperature does not necessarily mean that the highest daily maximum temperature in this month is a record high. For example, July 2006 was the warmest July on record, as far as the monthly mean temperature is concerned (23.69°C). However, the highest daily maximum temperature during this month was

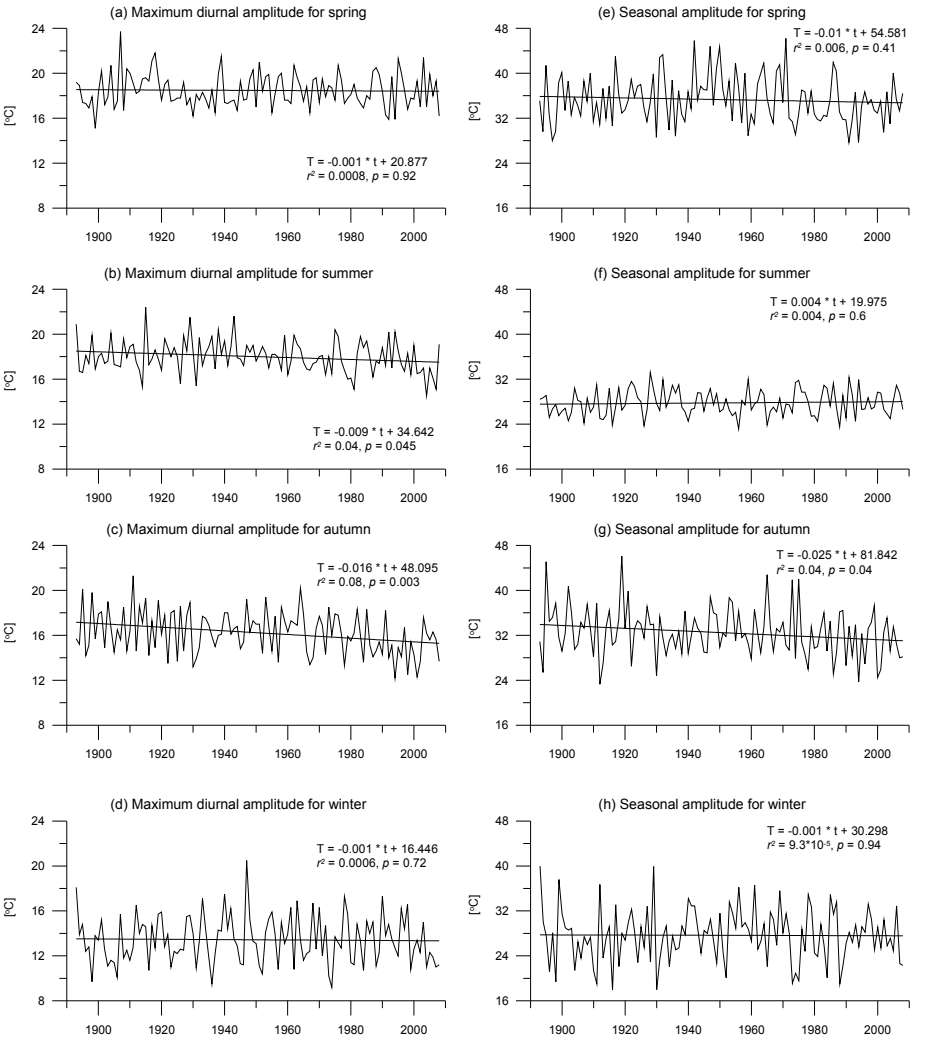


Fig. 3. Time series of temperature amplitudes – maximum diurnal amplitude (difference between maximum and minimum temperature for the same day) for (a) spring, (b) summer, (c) autumn, (d) winter; and seasonal amplitude (difference between maximum and minimum temperature for the same season) for (e) spring, (f) summer, (g) autumn, (h) winter.

35.9°C, that is, below the highest daily maximum temperature of 36.8°C, observed during a much less warm July 2007, when the mean monthly temperature was only 18.05°C.

Most of the intra-seasonal temperature amplitude (understood as the highest difference between daily maximum and minimum temperature in one season, *cf.* Fig. 3a,b,c,d) and the difference between a seasonal maximum and minimum (Fig. 3e,f,g,h) have decreased. The regression slope in all but one of these diagrams is negative, yet in four cases the slope is smaller than 0.005. The steepest slope is for autumn (–0.025 for a seasonal amplitude and –0.016 for maximum diurnal amplitude). Only for seasonal summer amplitude has the slope been positive.

This decreasing tendency is especially strong for the maximum diurnal amplitude in autumn (significant at 0.01 level). However, most (five out of eight) changes illustrated in Fig. 3 are not statistically significant, except for the maximum diurnal amplitude for summer (0.05 significance level) and autumn (0.01) and the seasonal amplitude for autumn (0.05). In the past, the temperature range was much higher. For instance, the lowest and highest autumn temperature values ever observed in 1911-1925 spanned the range from –12.4°C (on 29 November 1925) to +34.7°C (on 3 September 1911).

4. CHANGES IN SEASONAL NUMBERS OF "COLD" AND "WARM" EXTREMES

The numbers of cold and warm days and nights were determined for each season, based on subjective definition of seasonal „cold" and „warm". Alternatively, one could use the percentile-based definitions of these notions, but here impact-based definitions are found more meaningful and easier to interpret by the readership. For instance, the number of excursions of daily minimum temperature below 0°C is far more meaningful than a percentile-based index. Also excursions under the levels of –10°C and +10°C can be intuitively expected as thresholds for a cold night in winter and summer, respectively. The impact interpretation of such thresholds is quite natural. Frosts in spring and autumn jeopardize the traffic (slippery roads), while frosts in spring cause detrimental effects to sensitive crops (*e.g.*, blooming peaches and apricots, walnut, grapes). During a frosty night, some people (in particular the homeless and those under influence of alcohol) may freeze to death and sensitive plants may severely suffer. Cold summer nights adversely impact tourism (*e.g.*, people camping in tents), while heat waves of longer duration adversely affect human health (hyperthermia) and crops.

Figures 4-7 present temporal changes in the numbers of cold nights and days for each season, for the interval 1893-2008 (1893-2009 for winter). The seasonal thresholds for cold nights were selected as –10°C for winter, 0°C for spring and autumn, and +10°C for summer (Fig. 4). The seasonal thre-

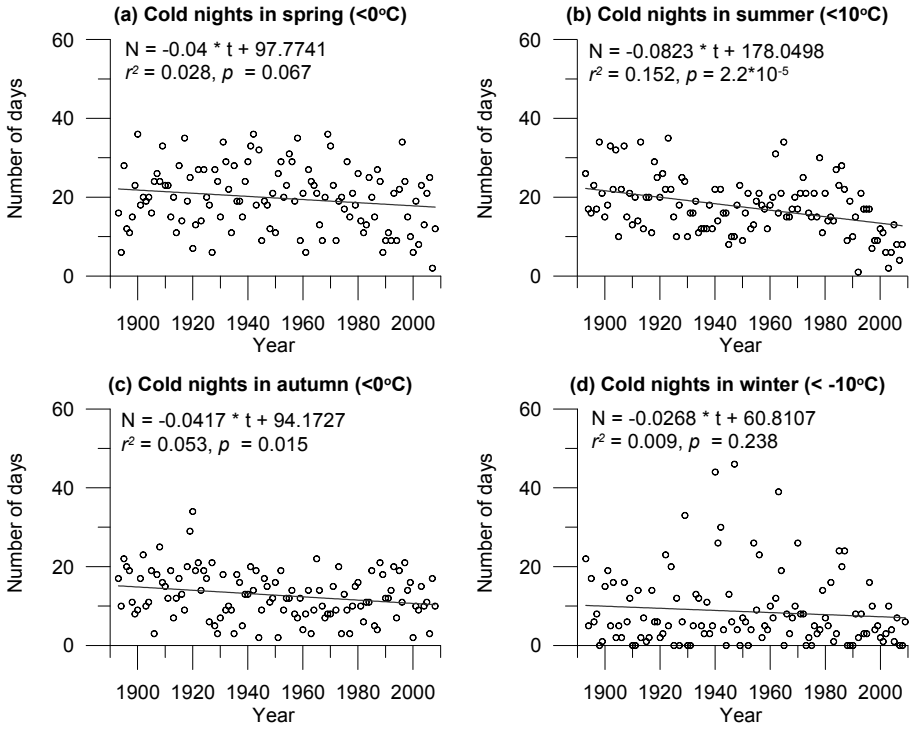


Fig. 4. A time series of the number of cold nights for each season in Potsdam for 1893-2008 (1893-2009 for winter).

sholds for cold days were selected as 0°C for winter, +10°C for spring and autumn, and +20°C for summer (Fig. 5). The seasonal thresholds for warm nights were selected as 0°C for winter, +10°C for spring and autumn, and +15°C for summer (Fig. 6). The seasonal thresholds for warm days were selected as +10°C for winter, +20°C for spring and autumn, and the threshold for hot summer days was +30°C (Fig. 7).

The warm-extreme indicators, such as the number of hot days (with maximum daily temperature exceeding 30°C) were found to increase. In agreement with the warming of winter temperatures, the cold-extreme indicators, such as the number of frost nights (assumed, for simplicity, to be equivalent to minimum daily temperature below 0°C) and of ice days (with maximum daily temperature below 0°C) have been decreasing. In 8 (out of 16) cases presented in Figs. 4-7, changes are statistically significant, at either the 0.01 level or the 0.05 level. In five categories (hot summer days, cold summer nights, cold autumn days, warm summer and spring nights), changes are significant at the 0.01 level and in three categories (warm winter days, cold and warm autumn nights) at the 0.05 level. In the remaining eight categories

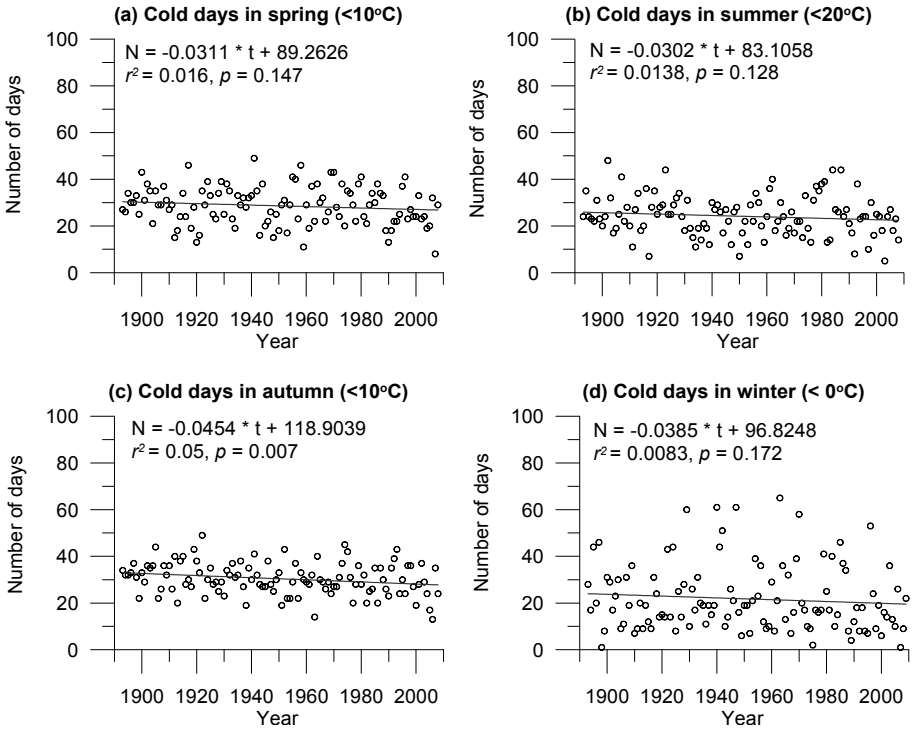


Fig. 5. A time series of the number of cold days for each season in Potsdam for 1893-2008 (1893-2009 for winter).

(cold spring and winter nights, cold days in spring, summer, and winter, warm winter nights, and warm days in spring and autumn) changes are not statistically significant at the 0.05 level. However, low correlation coefficient, r^2, and huge scatter illustrate strong random component (natural variability) of the data points in Figs. 4-7. Seasonal values of temperature indices for a particular year may strongly depart from the mean long-term relation, such as linear regression. This is strongest in winter, and in particular for winter temperature minima, whose drop from the long-term trend in a single year can be very abrupt.

Frost in autumn occurred as early on 2 October (in 1957), while the last spring frost occurred as late on 20 May (in 1952). That is, based on the observations made so far, the absolute frost-free period extends from 21 May to 1 October (132 days). Frost has never been noted on the Potsdam Station in the months of June, July, August, and September. The first hot day, in absolute terms, *i.e.*, a day with T_{max} in excess of 30°C, occurred as early on 22 April (31.8°C in 1968) and as late on 20 September (32.9°C in 1947). That is, in the light of the observations, over a couple of weeks, from

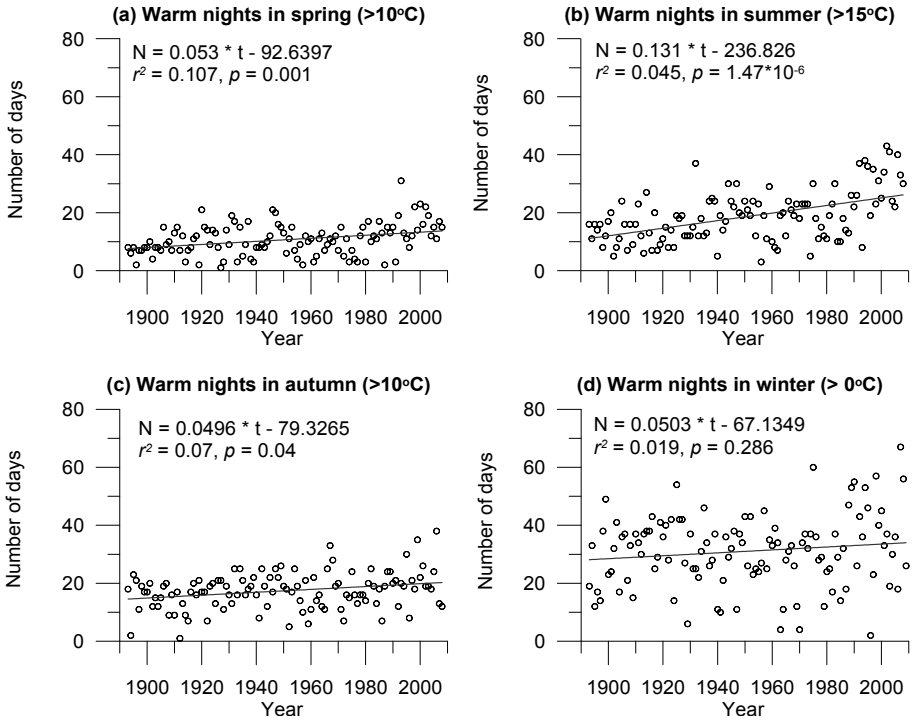


Fig. 6. A time series of the number of warm nights for each season in Potsdam for 1893-2008 (1893-2009 for winter).

22 April (first hot day) until 20 May (last frost), air temperature in Potsdam may as well go down below 0°C (minimum) or rise above 30°C (maximum).

Indicators related to frost and hot days are also illustrated in Figs. 8-10. Figure 8 presents the ordinal number of the last spring frost day ($T_{min} < 0$°C) and of the first autumn frost day. The regression slopes show that the last frost day has been occurring earlier than before in spring but the change is not statistically significant, while in autumn, frosts have been starting later (significance level 0.05). The increasing length of a frost-free interval also indicates a statistically significant (at 0.05 level) warming tendency (Fig. 9) – every decade, the frost-free interval grows, on average, by one day. However, in individual years, departures from the overall trends are very strong. For example, within the last 13 years (1996-2008) both the highest value of the annual number of frost days (133 days in 1996) and the lowest value (52 days in 2007) on record have been observed (Kundzewicz and Józefczyk 2008).

The warming is also accompanied by the increasing tendency of the time span of occurrence of hot days (Fig. 10), but the changes are not statistically significant.

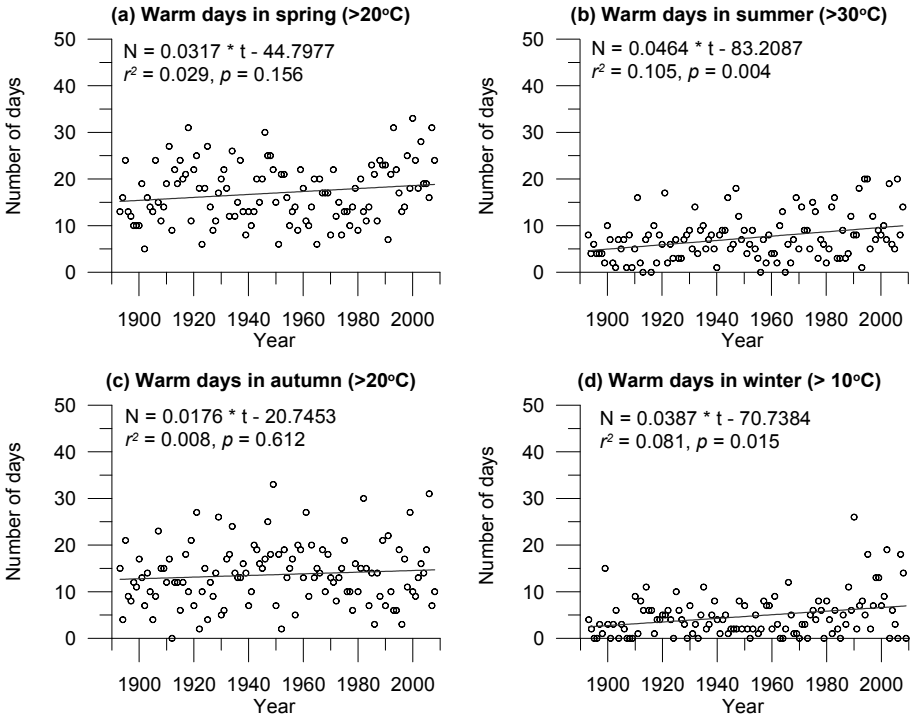


Fig. 7. A time series of the number of warm days for each season in Potsdam for 1893-2008 (1893-2009 for winter).

5. INTERPRETATION OF CHANGES

This paper illustrates a high year-to-year variability of temperature indices, superimposed on a warming trend, based on an analysis of a long time series of high-quality records. One may try to explain the sources of the substantial warming in recent decades, and prior to this, the lack of warming, and even some cooling in the 1950s and 1960s. In IPCC parlance, one needs to address a complex issue of change detection and attribution. Detection is a process of demonstrating that observed change is significantly different (in a statistical sense) from what can be explained by natural internal variability. Once a change is detected, attribution is a process of demonstration that:

- the detected change is consistent with a combination of external forcing including anthropogenic changes in the composition of the atmosphere and natural internal variability; and
- it is not consistent with alternative, physically-plausible explanations of recent climate change that exclude important elements of the given combination of forcings.

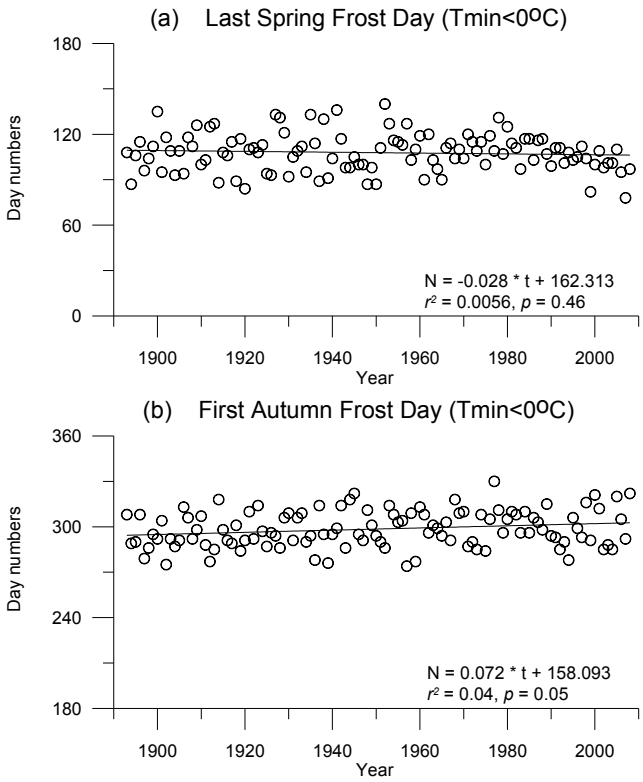


Fig. 8. The ordinal number of the last spring frost day (T_{min} < 0°C) (a) and of the first autumn frost day (b) in individual years. The New Year day is interpreted as day number 1.

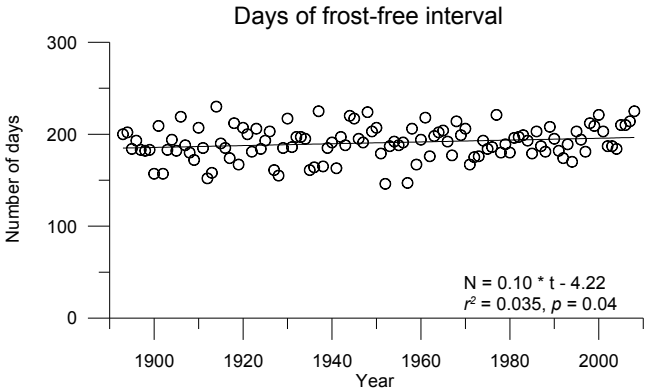


Fig. 9. The number of days of a frost-free interval (for each year last-spring-frost-day-number in Fig. 8 was subtracted from first-autumn-frost-day-number).

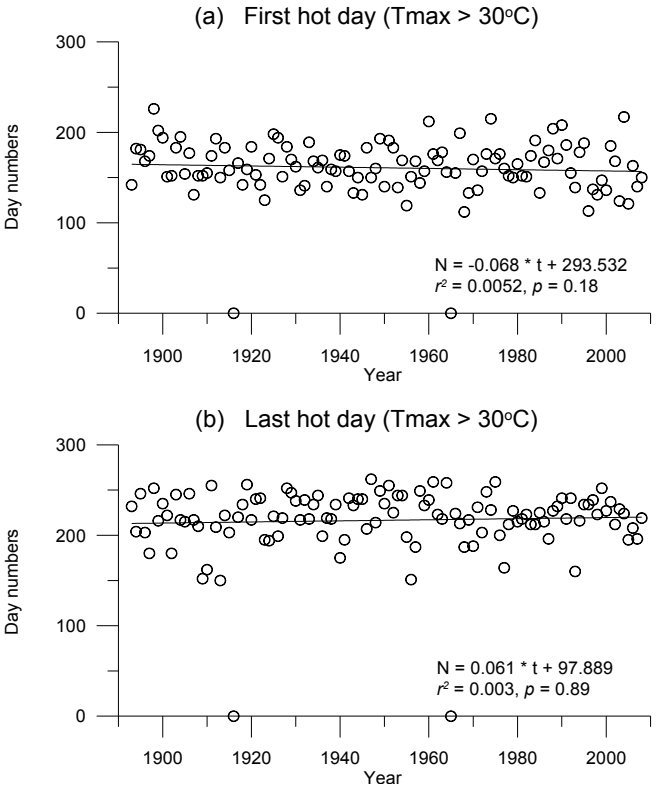


Fig. 10. The ordinal number of the first hot day (T_{max} > 30°C) (a) and of the last hot day (b) for individual years. It should be noted that in two years, 1916 and 1965, there was not even a single day with maximum temperature above 30°C.

A formal process of detection and attribution cannot be carried out for records on a single station. However, the possible mechanisms of change could, and should, also be discussed.

The climate of our planet has been changing globally many times in the Earth's history – there have been many warmer and many colder intervals. Mechanisms of climate change can be divided into the following four groups: (i) changes in the solar activity (*cf.* sunspots number); (ii) changes in orbital parameters (in time scale of tens of millennia, irrelevant to the present climate change); (iii) changes in the composition of the Earth's atmosphere (greenhouse gases – water vapour, carbon dioxide, methane, and nitrous oxide; aerosols; dust); and (iv) changes in the properties of the Earth's surface (albedo, water storage). The first two mechanisms above are purely natural and mankind has no influence on them. The latter two mechanisms can be influenced by both natural and anthropogenic factors. The global increases

of concentrations of greenhouse gases (IPCC 2007), which are real and strong, are not sufficient to explain the details of the observed temperature change. The rapid temperature increase observed over Europe (and also at the Potsdam station) in the last three decades is considerably stronger than the mean global warming and the temperature rise expected from anthropogenic greenhouse gas increases. Variability of temperature indices can be partly explained by the oscillations in the system of ocean and atmosphere (notably North Atlantic Oscillations).

Several authors (*e.g.*, Makowski *et al.* 2008, Ruckstuhl *et al*. 2008) found that the aerosol and cloud-induced radiative forcing could explain a portion of the recent changes in temperature indices in Europe.

Solar irradiance measurements on the Earth's surface illustrate considerable changes. An interval of global solar dimming and subsequently – an interval of global solar brightening (continuing to-date) have been noted that cannot be explained by variations of the Sun's activity. The explanation is sought in the changes of atmospheric transmittance due to increases and subsequent decreases in anthropogenic aerosol concentrations, cloud-mediated aerosol effects, and direct cloud effects. In Europe, sulphurous emissions have grown since the 1950s, then peaked in the early 1970s in Western Europe, and in the late 1980s and the early 1990s – in Eastern Europe, and decreased since then. A reduction in anthropogenic aerosol concentrations in Western Europe, since the late 1970s and the early 1980s, resulted from considerable efforts undertaken in many countries to curb air pollutant emissions. The decrease in Eastern Europe can be partly associated with the economic collapse of the communist system, dominated by heavy industry, responsible for the high level of air pollution.

Aerosols affect atmospheric transmittance and hence temperature via the direct aerosol effect (scattering and absorption of sunlight by aerosol particles). There also exist, however, cloud-mediated indirect aerosol effects, such as the cloud albedo effect (enhancement of cloud albedo due to smaller droplets) or the cloud lifetime effect (extension of cloud lifetime due to smaller droplets and less precipitation loss), *cf.* Ruckstuhl *et al.* (2008). Clouds simultaneously affect solar shortwave and thermal longwave radiation but with opposite sign. Hence, the total cloud effect is the sum of the negative shortwave cloud effect and the longwave cloud effect (in which water vapour functions as a greenhouse gas), that partly compensate each other.

The findings of this present paper, illustrated in Fig. 3a, b, c, d, show that the maximum diurnal amplitude observed in Potsdam has decreased with time for all seasons, but only for summer and autumn are the changes statistically significant (at levels 0.05 and 0.01, respectively). However, there is a very strong variability around decreasing trends of seasonal maximum diurnal amplitude.

These findings can be indirectly compared to the results of Makowski *et al.* (2008), who investigated annual mean diurnal temperature range (DTR) for the period 1950-2005 for 23 different countries and regions in Europe as well as Europe as a whole. They demonstrated that the long-term trend of DTR has reversed from a decrease to an increase during the 1970s in Western Europe and during the 1980s in Eastern Europe. For the 16 out of 23 regions studied, as well as for the European mean, there was a statistically significant period of decrease and a subsequent increase in DTR. Of the remaining seven regions, two show a non-significant increase, three show a significant decrease and two reflect no significant trend (therein the eastern part of Germany, where Potsdam is located).

The diurnal temperature range is a suitable measure to investigate the counteracting effects of longwave and shortwave radiative forcing, because the diurnal minimum is closely related to the longwave radiative flux, while the diurnal maximum is predominantly determined by shortwave radiation. Makowski *et al*. (2008) find that the long-term trends in DTR are strongly affected by changes in incoming shortwave radiation (undergoing a dramatic change from dimming to brightening), presumably largely influenced by the direct and indirect effects of aerosol from SO_2 emissions.

6. CONCLUSIONS

Besides conducting the studies of change detection in mean temperature data, the research community has been carefully watching temperature extremes in different categories, such as the maximum and minimum daily, monthly, seasonal, and annual temperatures. The present paper indicates that global and general findings of ubiquitous warming are in general agreement with temperature extremes in a specific, long-term, high-quality observation record. However, it shows that the natural variability at a single station is very strong and that extremes in a single year may largely differ from the dominating tendency. Absolute record values of maximum or minimum temperature do not necessarily match the trend present in the long-term time series. It can be clearly seen that high values of „warm“ extremes (such temperature-related indicators as seasonal maximum and minimum temperatures, number of hot days) may have occurred many decades ago, when the level of warming (as indicated by the linear regression) was much lower. Similarly, despite the warming, cold extremes may have occurred in recent decades, largely differing from the value corresponding to the decreasing tendency.

Hence, one has to be careful with the interpretation of warming. Rather than re-iterating the global warming statement with every exceptional warm spell and questioning it with every exceptional cold spell (*e.g.*, January 2010), as often done in the media, one needs to take a more balanced view

with consideration of old records and natural variability. Contrary to common interpretation, climate vagaries have always been strong. This should be remembered even if there is a tendency for "cold" extremes to become less frequent and less severe and for "warm" extremes to become more frequent and more severe.

A formal process of change detection and attribution in temperature indices at the Potsdam station cannot be carried out. However, possible mechanisms of change were discussed, including the link between air pollution and warming. The analysis of data at a baseline station where long time series of records are available allowed the authors to contribute to a more general debate, in which the data series are typically much shorter and of lower quality.

Acknowledgements. The authors express their gratitude to Professor Peter Werner, for kind data support. The German Weather Service is commended for the continuation of observations at the baseline Potsdam station and for making data freely available on the web (online). The authors acknowledge the constructive remarks of three anonymous referees who made it possible to improve this contribution. Special thanks go to Professor Zbigniew Sorbjan, who proposed a constructive way of enriching the material.

References

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett, and P.D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.* **111**, D12106, DOI: 10.1029/2005 JD006548.

IPCC (2007), Summary for policymakers. **In:** S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.), *Climate Change 2007: The Physical Science Basis*, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, USA.

Kundzewicz, Z.W., and D. Józefczyk (2008), Temperature-related climate extremes in the Potsdam Observation Record, *Geografie* **113**, 4, 372-382.

Kundzewicz, Z.W., D. Józefczyk, and H. Österle (2007), Warmest 12 consecutive months on record at the Potsdam meteorological station, Germany, *Weather* **62**, 10, 284-286.

Kundzewicz, Z.W., F.-W. Gerstengarbe, H. Österle, P.C. Werner, and W. Fricke (2009), Recent anomalies of mean temperature of 12 consecutive months –

Germany, Europe, Northern Hemisphere, *Theor. Appl. Climatol.* **95**, 3-4, 417-422, DOI: 10.1007/s00704-008-0013-9.

Lehmann, A., and M. Kalb (1993), 100 Jahre meteorologische Beobachtungen an der Säkularstation Potsdam 1893-1992, Selbsverlag des DeutschenWetterdienstes Offenbach am Main, 32 pp.

Makowski, K., M. Wild, and A. Ohmura (2008), Diurnal temperature range over Europe between 1950 and 2005, *Atmos. Chem. Phys. Discuss*. **8**, 7051-7084, also: http://www.atmos-chem-phys-discuss.net/8/7051/2008/.

Ruckstuhl, C., R. Philipona, K. Behrens, M. Collaud Coen, B. Dürr, A. Heimo, C. Mätzler, S. Nyeki, A. Ohmura, L. Vuilleumier, M. Weller, C. Wehrli, and A. Zelenka (2008), Aerosol and cloud effects on solar brightening and the recent rapid warming, *Geophys. Res. Lett*. **35**, L12708, DOI: 10.1029/2008GL034228.

Trenberth, K.E., P.D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J.A. Renwick, M. Rusticucci, B. Soden, and P. Zhai (2007), Observations: Surface and atmospheric climate change. **In:** S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.), *Climate Change 2007: The Physical Science Basis*. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK and New York, USA.

Received 20 May 2009
Received in revised form 7 April 2010
Accepted 21 April 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1134-1146
DOI: 10.2478/s11600-010-0019-4

On the Origin of Seasonal Variation of Aerosol Optical Thickness in UV Range over Belsk, Poland

Janusz JAROSŁAWSKI and Aleksander PIETRUCZUK

Institute of Geophysics, Polish Academy of Sciences, Warszawa, Poland
e-mails: januszj@igf.edu.pl, alek@igf.edu.pl

Abstract

Aerosol optical thickness (AOT) and seasonal variation of AOT over Belsk, Poland, in the UV wavelength range (310-380 nm) have been analysed using results of measurements by Brewer spectrophotometer No. 064 and Cimel sunphotometer data for the 2002-2007 period. The comparison of AOT derived from direct Sun measurements by Brewer spectrophotometer in the 310-320 nm range and retrieved from Cimel measurements at longer wavelengths shows good correlation ($R = 0.96$), with overestimation of retrieved values compared to the measured ones by about 6%. Basing on aerosol microphysical properties taken from almucantar retrievals and Mie theory, optical properties of aerosol in the UV range has been calculated. Analysis of seasonal variation of AOT at Belsk reveals two maxima: in April and July-August. Analysis of back-trajectories in conjunction with analysis of fire maps from Fire Information For Resource Management System shows that these seasonal maxima are connected with seasonal biomass burning in Eastern and Southern Europe.

Key words: atmospheric aerosol, UV, Brewer spectrophotometer, Cimel sunphotometer.

1. INTRODUCTION

Atmospheric aerosols can modify substantially the amount of solar UV radiation reaching the Earth's surface (e.g., Liu *et al*. 1991, Kylling *et al*. 1998,

WMO 2003). Their role in attenuation of solar UV radiation can be as important as the atmospheric ozone during high aerosol levels episodes (Zerefos 1997, Krzyścin and Puchalski 1998, Jarosławski and Krzyścin 2005). Up to 50% of solar UV radiation can be absorbed and scattered during such episodes (Acosta and Evans 2000). Optical characteristics of aerosol particles in the visible part of solar radiation are widely and routinely measured throughout the world in respect of their role in climate modification through absorption and scattering of solar light (IPCC 2007). On the other hand, measurements of optical properties of atmospheric aerosol in the shorter UV wavelengths (below 340 nm) are less frequent, mainly because of difficulties connected with the measurement itself (Cheymol and de Backer 2003, Gröbner and Meleti 2004, Jarosławski *et al.* 2003, Kazadzis *et al*. 2007)

In this work we analyze aerosol optical thickness (AOT) of atmospheric aerosol over Belsk, Poland, in the UV wavelength range (310-380 nm). We use direct Sun observations from Brewer spectrophotometer as well as the retrievals from the almucantar scans collected by the Cimel sunphotometer during the 2002-2007 period using the method presented in our previous paper (Pietruczuk and Jarosławski 2008). We compare the statistical characteristics of AOT data set obtained by the direct (direct Sun measurements from Brewer spectrophotometer) and indirect (calculated by the Mie theory from Cimel almucantar measurements) methods. A modified Brewer calibration algorithm to reduce possible systematic errors, especially for the wavelengths shorter than 320 nm, is proposed.

2. METHOD AND INSTRUMENTATION

Measurements of aerosol optical properties reported in this paper were taken in the Central Geophysical Observatory of the Institute of Geophysics, Polish Academy of Sciences, at Belsk. The observatory is located in the rural site about 50 km to the SSW from Warsaw. It could be qualified as a background station weakly affected by local pollutions. Measurements of atmospheric aerosol by means of sunphotometers, Brewer spectrophotometer and lidar, as well as in-situ techniques are performed at Belsk. Various standard meteorological observations are also carried out, complementarily to aerosol measurements. In this work we examine Sun-photometer data obtained by Brewer spectrometer and collocated CIMEL instrument.

Aerosol optical thickness (AOT) data in the UV range (310-320 nm) has been retrieved from the total ozone measurements (the so-called direct Sun observations) performed by the Brewer ozone spectrophotometer No. 064. The instrument is calibrated on regular yearly basis against the travel standard (No. 017 maintained by the IOS company). The basic constants (extra terrestrial constants, simulating instrumental response to solar radiation,

when spectrophotometer is located outside the atmosphere – ETC) of the Brewer spectrophotometer necessary for AOT calculation have been obtained by the Langley plot method applied for the direct Sun radiation measurements performed at Belsk. Detailed description of measurement technique and calibration of the instrument have been made by Jarosławski *et al.* (2003). According to the later findings of Arola and Koskela (2004), the procedure of ETC calculation has been modified to minimize the possible systematical errors that might occur for wavelengths shorter than 320 nm. The range of airmass used for ETC calculation by the Langley method has been limited to 3 which corresponds to solar zenith angles smaller than 70 degrees. Moreover, days with negligible daily variation of total ozone content measured in situ by the Brewer spectrophotometer as well as stable (within ±0.02 – estimated instrument uncertainty for AOT measurements) and low (typically below 0.25 at 340 nm) aerosol optical depth from the Cimel sunphotometer have been selected for ETC calculation. The AOT data obtained at airmasses above 3 have been also rejected from further analysis to minimize the influence of the diffuse light on calibration and measurements results. This requirement practically excludes winter months (November, December, January) from analysis.

AOTs taken form Brewer measurements were compared with those obtained by CIMEL instrument. Belsk's CIMEL CE-318 sunphotometer, instrument No. 318, is part of AERONET network (Holben *et al*. 1998). This well known and widely used instrument employs photodiode detector to measure direct Sun radiation and Sky radiance. Direct Sun measurements taken at wide range (340-1020 nm) are used to determine aerosol optical depths, Angström exponent, and columnar water vapour concentration. The sky radiance measurements are also taken at 440, 670, 870, and 1020 nm by means of almucantar and principal plane scans. Nonlinear numerical procedure (Dubovik and King 2000) is applied to the almucantar scans providing microphysical and radiative properties of aerosol at wavelengths of measurements. Those products are called almucantar retrievals.

Basing on aerosol microphysical properties, taken from almucantar retrievals, and Mie theory, we calculated aerosol optical properties in UV range. Our previous paper (Pietruczuk and Jarosławski 2008) shows that AOTs calculated in this way well agree with AOT derived from Brewer direct Sun total ozone measurements. Correlation coefficients of about 0.98-0.99 were obtained for wavelengths of 340 and 380 nm, when our calculations were compared to direct CIMEL measurements. In case of comparison with Brewer direct Sun measurements correlation coefficients were about 0.96. This shows that the method presented in our previous paper correctly estimates AOT in UV range, in spite of the fact that values obtained by this method are slightly overestimated in UV as compared to direct Sun measurements.

3. RESULTS

During the five years of the reporting period, 130 000 direct Sun measurements by means of Brewer spectrometer, over 16 000 direct Sun measurements by means of CIMEL photometer, and 850 CIMEL almucantar scans have been taken. We used AERONET level 2.0 data which are cloud screened, recalibrated and manually inspected. AERONET direct Sun data were available up to May 2007, the date of last calibration of Belsk CIMEL, but almucantar retrievals for spherical particles were available before 2006.

Figure 1 presents comparison of AOT at 320 nm from Brewer direct Sun measurements and from the Mie theory applied to CIMEL almucantar retrievals. Brewer AOTs at 320 nm have been averaged over five minutes around the time of almucantar scan (typically 5 points). The Brewer and CIMEL retrievals correlate very well, correlation parameter $R = 0.96$, but AOTs retrieved from almucantar scans seem to be overestimated by 6% comparing to Brewer ones; the regression coefficient is 1.06. The overestimation could increase with decreasing wavelength, as shown in our previous paper (Pietruczuk and Jarosławski 2008).

Aerosol Optical Thickness

Frequency distribution of aerosol optical thickness is typically skewed and the median value of AOT is a better descriptor than the mean value for these

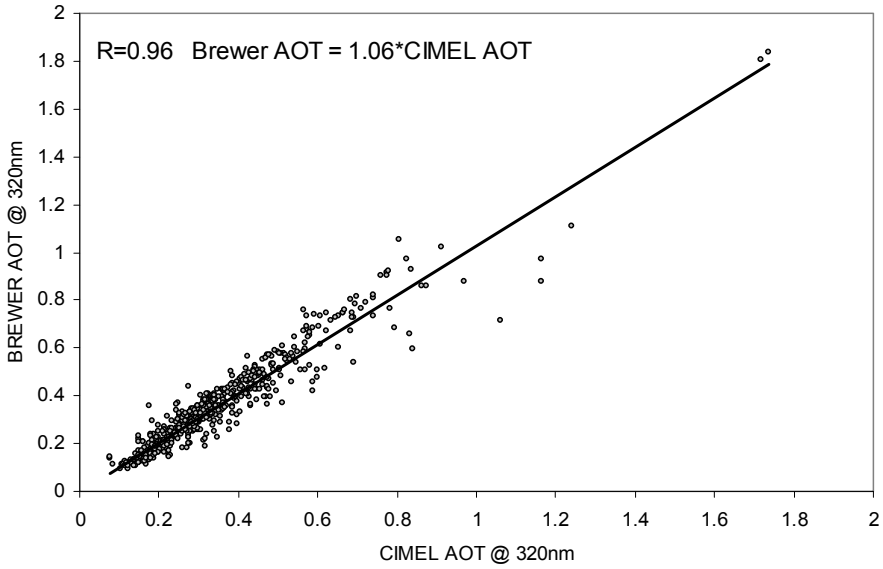


Fig. 1. Correlation between AOT at 320 nm measured by Brewer spectrometer and retrieved form CIMEL almucantar products.

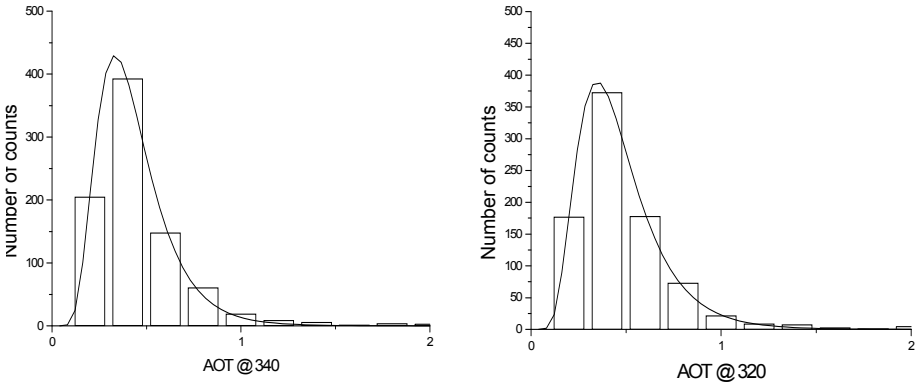


Fig. 2. Frequency distributions of AOT retrieved from microphysical properties of aerosol.

average conditions (see Fig. 2). Such a distribution follows log-normal distribution (O'Neill *et al.* 2000, Behnert *et al*. 2007) given by the following equation:

$$N(\tau) = A \exp\left(-\frac{\ln^2(\tau/\tau_C)}{2\omega^2}\right),$$

where $N(\tau)$ is the number of counts of AOT between τ and $(\tau + \mathrm{d}\tau)$. That means, the logarithm of AOT follows Gaussian distribution with mean $\ln(\tau_C)$ and standard deviation ω. The value of τ_C is the median and modal value of the distribution. Parameters of log-normal distributions fitted to the frequency distribution of AOT at 310, 320, 340, and 380 nm compared with direct Sun results are listed in Table 1.

The difference between AOTs at 310 and 320 nm is much smaller than that between AOTs at 320 nm and 340 nm which has a consequence in Ångström exponent decreasing in UV range. The difference between AOTs

Table 1

Parameters of log-normal distribution fitted to frequency distribution of AOT derived from all single almucantar retrievals (Mie calculations) and direct Sun measurements: Brewer (310-320 nm) and CIMEL (340-380 nm)

		310 nm	320 nm	340 nm	380 nm
Almucantar retrievals	τ_C	0.346	0.332	0.305	0.260
	ω	0.573	0.577	0.584	0.595
Direct Sun measurements	τ_C	0.296	0.304	0.303	0.262
	ω	0.536	0.486	0.646	0.643

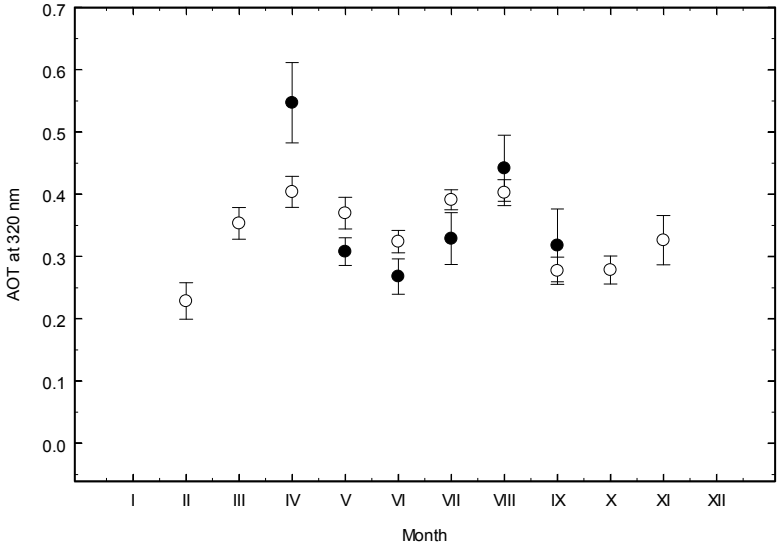


Fig. 3. Time series of aerosol optical thickness at 320 nm derived from Brewer measurements (white circles) and CIMEL retrievals (black circles).

derived from almucantar retrievals and direct Sun Brewer measurements increases with decreasing wavelength. This effect was noticed in our previous paper (Pietruczuk and Jarosławski 2008). It is probably caused by uncertainties connected to extrapolation of refractive index into UV range or with uncertain size distribution of fine mode of aerosol. Small particles influence AOTs in UV range more than coarse ones.

Seasonal variability of aerosol optical properties is connected mainly with atmospheric circulation and aerosol source of origin. Climatology of AOT in the visible range performed by many authors (Holben *et al.* 2001, Behnert *et al*. 2007) shows maximum mainly in the summer season. Similar results are obtained for AOT in UV range (Kazadzis *et al.* 2007). This summer maximum, observed also in our results, is probably connected with the accumulation of aerosol in the boundary layer. Slow air motion over urban/industrial areas of the continent favor accumulation of the pollutants in the moving airmass. Our measurements show two maxima of AOT, the one during summer season (July and August) and the other in April. Figure 3 presents monthly averaged modal value of AOT distribution. We used τ_C parameter, of lognormal distribution of AOTs in case of Brewer measurements and modal values of AOTs in case of CIMEL retrievals. We decided to use monthly averages obtained from daily averaged values to avoid situation when hundreds or even thousands of single measurements of low values of AOT taken during a cloud-free day dominate the distribution for the whole

month. Parameter ω of log-normal distribution in case of Brewer data and the width of distribution between 25 and 75% of counts in case of CIMEL retrievals divided by square root of the number of observation serves as an error of AOT characteristic for each month.

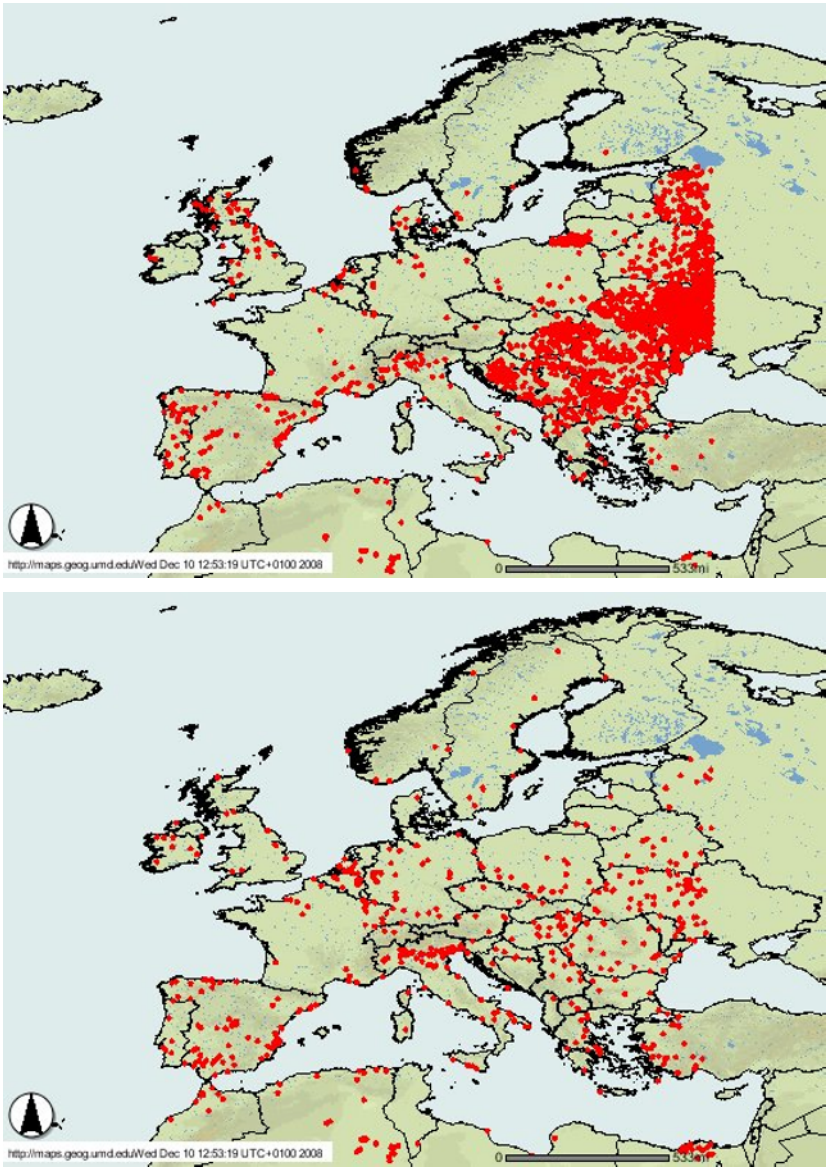


Fig. 4. Maps of fire activity in Europe for March 2007 (upper) and June 2007 (lower) according to the MODIS fire product. Colour version of this figure is available in electronic edition only.

Both maxima of AOT are probably connected with seasonal fires in Eastern (spring) and Southern Europe (summer) and accumulation processes. Analysis of MODIS fire products (Justice *et al*. 2002, NASA 2002), available at http://maps.geog.umd.edu, shows seasonal activity of fires in these regions in the respective months. An example of the fires occurrence during and outside the biomass burning season and an advection of polluted air over Belsk resulting in the very high levels of AOT is shown in Figs. 4 and 5. Similar configuration of meteorological conditions has been observed for the majority of cases of elevated AOT observed at Belsk in March-April and July-

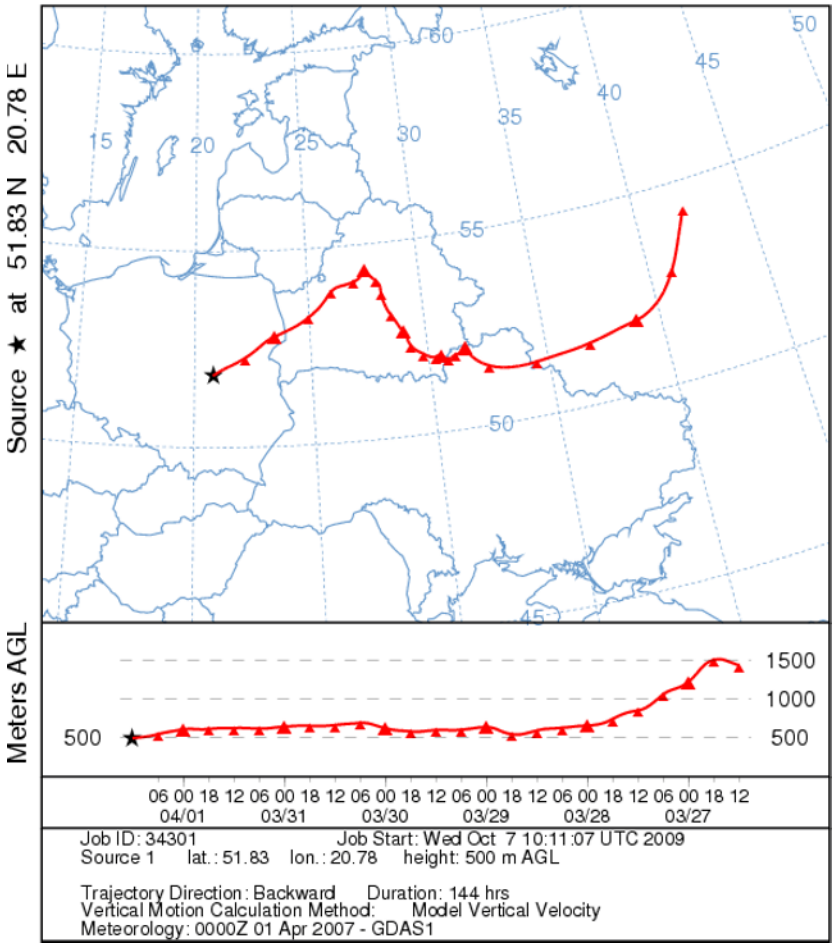


Fig. 5. Backward trajectory ending at Belsk on 1 April 2007 at 12 GMT. Mean AOT at 340 nm measured at Belsk was equal 1.42, which is 300% of the monthly mean. The trajectory starts in the region of intense biomass burning (see Fig. 4, upper map). Colour version of this figure is available in electronic edition only.

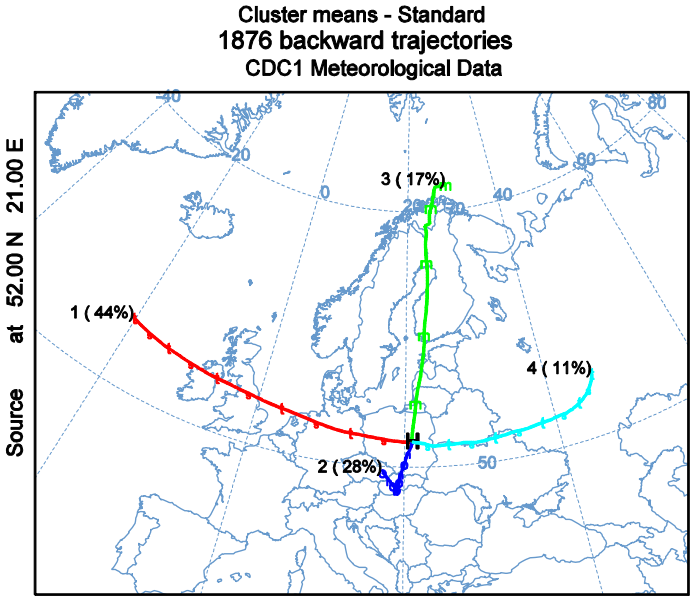


Fig. 6. Mean trajectories of air arriving over Belsk according to the results of cluster analysis of backward trajectories. Colour version of this figure is available in electronic edition only.

August. To verify this hypothesis the backward trajectories of air coming over Belsk in the period 2002-2007 have been computed using the HYSPLIT model available at the web page: http://www.arl.noaa.gov/HYSPLIT.php. We used trajectories ending at Belsk at noon at 500 m above ground level as most representative for pollution transport in this region. We neglected transport in free troposphere because pollution like mineral dust, which is typically transported over boundary layer is relatively rarely observed over Central Europe. Trajectories were subsequently classified in groups by the cluster analysis – a tool built in HYSPLIT software package (Jorba *et al*. 2004). Similar cluster analysis was previously made for Belsk (Pietruczuk and Podgórski 2009) and other stations and is considered to be a useful tool for finding the aerosol source of origin.

Mean trajectories for each cluster are shown in Fig. 6. One can distinguish four main directions of air advection corresponding roughly to North, East, South and West. Then the seasonal variability of AOT in airmass classified to each cluster has been calculated (see Fig. 7). That means we used daily means of AOT corresponding to trajectory form certain cluster to calculate monthly mean of AOT for each cluster.

The highest values of monthly means occur in spring and summer season for clusters 2 and 4, which corresponds to the airmasses coming from east

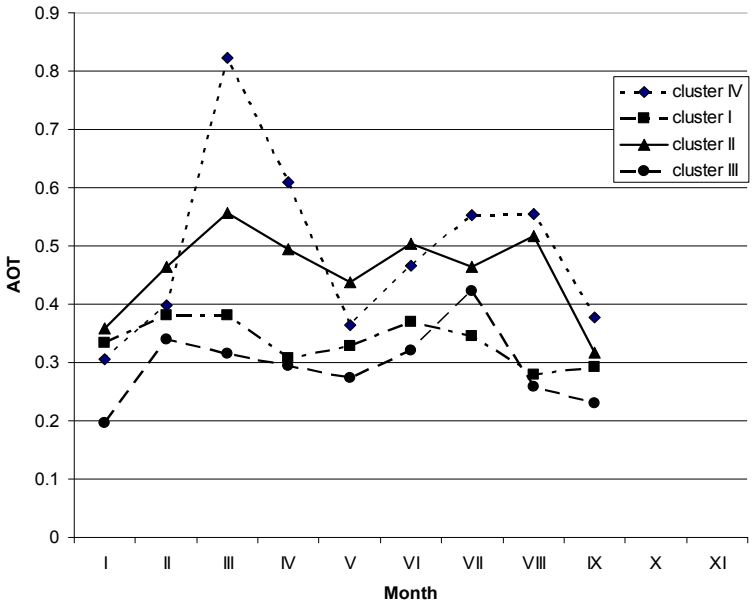


Fig. 7. Seasonal variation of AOT in air arriving over Belsk from the different directions according to the classification shown in Fig. 6.

and south during the biomass burning season. The variability of AOT in case of clusters 1 and 3 is smoother. Biomass burning products from this region are also detected in Central and Northern Europe by means of remote techniques (e.g., Pietruczuk and Chaikovsky 2007, Arola *et al*. 2007) and influence of such kind of the aerosol on surface UV radiation was also studied (Pietruczuk and Krzyścin 2008). Another possible source of aerosol during spring and late summer could be rural activity producing erosive fraction of the aerosol (soil aerosol).

4. CONCLUSIONS

The AOT values in the UV range 310-320 nm calculated by the modified method of calibration and measurement do not show significant wavelength dependence, as concerns the mean values for the whole dataset. Direct Sun measurement results from Brewer spectrophotometer correlate well with those derived from almucantar retrievals (correlation coefficient $R = 0.96$). The annual variation of AOT shows two maxima (April and August). Analysis of the seasonal variation of AOT for the air masses arriving over Belsk from different sectors revealed from clustering of backward trajectories, connected with analysis of the intensity and location of biomass burning in

Europe, lead to the conclusion that seasonal variability of AOT at Belsk is caused by the advection of aerosol from remote regions of biomass burning and from accumulation of aerosol in the atmosphere as a result of the specific meteorological conditions during summer. Practically all episodes with AOT significantly higher than the seasonal mean occur as a result of advection of the air from the regions of biomass burning.

References

Acosta, L.R., and W.F.J. Evans (2000), Design of the Mexico City UV monitoring network: UV-B measurements at ground level in the urban environment, *J. Geophys. Res.* **105**, D4, 5017-5026, DOI: 10.1029/1999JD900250.

Arola, A., and T. Koskela (2004), On the sources of bias in aerosol optical depth retrieval in the UV range, *J. Geophys. Res*. **109**, D08209, DOI: 10.1029/2003JD004375.

Arola, A., A. Lindfors, A. Natunen, and K.E.J. Lehtinen (2007), A case study on biomass burning aerosols: effects on aerosol optical properties and surface radiation levels, *Atmos. Chem. Phys*. **7**, 16, 4257-4266.

Behnert, I., V. Matthias, and R. Doerffer (2007), Aerosol climatology from ground-based measurements for the southern North Sea, *Atmos. Res.* **84**, 3, 201-220, DOI: 10.1016/j.atmosres.2006.05.006.

Cheymol, A., and H. De Backer (2003), Retrieval of the aerosol optical depth in the UV-B at Uccle from Brewer ozone measurements over a long time period 1984-2002, *J. Geophys. Res.* **108**, D24, 4800, DOI: 10.1029/2003JD003758.

Dubovik, O., and M.D. King (2000), A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, *J. Geophys. Res.* **105**, D16, 20673-20696, DOI: 10.1029/2000JD900282.

Gröbner, J., and C. Meleti (2004), Aerosol optical depth in the UVB and visible wavelength range from Brewer spectrophotometer direct irradiance measurements: 1991-2002, *J. Geophys. Res.* **109**, D09202, DOI: 10.1029/2003JD004409.

IPCC (2006), Third Assessment Report: Climate Change 2007: Synthesis Report by R.T. Watson and the Core Writing Team (eds.), Intergovernmental Panel on Climate Change Geneva, Switzerland, 184.

Holben, B.N., T.F. Eck, I. Slutsker, D. Tanré, J.P. Buis, A. Setzer, E. Vermote, J.A. Reagan, Y.J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov (1998), AERONET – A federated instrument network and data archieve for aerosol characterization, *Remote Sens. Environ*. **66**, 1, 1-16, DOI: 10.1016/S0034-4257(98)00031-5.

Holben, B.N., D. Tanré, A. Smirnov, T.F. Eck, I. Slutsker, N. Abuhassan, W.W. Newcomb, J.S. Schafer, B. Chatenet, F. Lavenu, Y.J. Kaufman, J. Vande Castle, A. Setzer, B. Markham, D. Clark, R. Frouin, R. Halthore, A. Karneli, N.T. O'Neill, C. Pietras, R.T. Pinker, K. Voss, and G. Zibordi (2001), An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, *J. Geophys. Res.* **106**, D11, 12067-12097, DOI: 10.1029/2001JD900014.

Jarosławski, J.P., and J.W. Krzyścin (2005), Importance of aerosol variations for surface UV-B level: Analysis of ground-based data taken at Belsk, Poland, 1992-2004, *J. Geophys. Res.* **110**, D16201, 10.1029/2005JD005951.

Jarosławski, J., J.W. Krzyścin, S. Puchalski, and P. Sobolewski (2003), On the optical thickness in the UV range: Analysis of the ground-based data taken at Belsk, Poland, *J. Geophys. Res.* **108**, D23, 4722, DOI: 10.1029/2003JD003571.

Jorba, O., C. Pérez, F. Rocadenbosch, and J.M. Baldasano (2004), Cluster analysis of 4-day back trajectories arriving in the Barcelona area (Spain), from 1997 to 2002, *J. Appl. Meteorol*. **43**, 6, 887-901, DOI: 10.1175/1520-0450(2004)043<0887:CAODBT>2.0.CO;2.

Justice, C.O., L. Giglio, S. Korontzi, J. Owens, J.T. Morisette, D. Roy, J. Descloitres, S. Alleaume, F. Petitcolin, and Y. Kaufman (2002), The MODIS fire products, *Remote Sens. Environ.* **83**, 1-2, 244-262, DOI: 10.1016/S0034-4257(02)00076-7.

Kazadzis, S., A. Bais, V. Amiridis, D. Balis, C. Meleti, N. Kouremeti, C.S. Zerefos, S. Rapsomanikis, M. Petrakakis, A. Kelesis, P. Tzoumaka, and K. Kelektsoglou (2007), Nine years of UV aerosol optical depth measurements at Thessaloniki, Greece, *Atmos. Chem. Phys*. 7, 2091-2101.

Krzyścin, J.W., and S. Puchalski (1998), Aerosol impact on the surface UV radiation from ground-based measurements taken at Belsk, Poland, 1980-1996, *J. Geophys. Res.* **103**, D13, 16175-16181, DOI: 10.1029/98JD00899.

Kylling, A., A.F. Bais, M. Blumthaler, J. Schreder, C.S. Zerefos, and E. Kosmidis (1998), Effect of aerosols on solar UV irradiances during the photochemical activity and solar ultraviolet radiation campaign, *J. Geophys. Res.* **103**, D20, 26051-26060, DOI: 10.1029/98JD02350.

Liu, S.C., S.A. McKeen, and S. Madronich (1991), Effect of anthropogenic aerosols on biologically active ultraviolet radiation, *Geophys. Res. Lett*. **18**, 12, 2265-2268, DOI: 10.1029/91GL02773.

NASA/University of Maryland (2002) MODIS Hotspot/Active Fire Detections. Data set. MODIS Rapid Response Project, NASA/GSFC (producer), University of Maryland, Fire Information for Resource Management System (distributors), available on-line: http://maps.geog.umd.edu.

O'Neill, N.T., A. Ignatov, B.N. Holben, and T.F. Eck (2000), The lognormal distribution as a reference for reporting aerosol optical depth statistics; empirical

tests using multi-year, multi-site AERONET sunphotometer data, *Geophys. Res. Lett*. **27**, 20, 3333-3336, DOI: 10.1029/2000GL011581.

Pietruczuk, A., and A.P. Chaikovsky (2007), Properties of fire smoke in Eastern Europe measured by remote sensing method, *Proc. SPIE* 6745, 67451T, DOI: 10.1117/12.740916.

Pietruczuk, A., and J. Jarosławski (2008), An alternative method for aerosol optical thickness retrieval in the UV range, *J. Atmos. Sol.-Terr. Phys*. **70**, 7, 973-979, DOI: 10.1016/j.jastp.2008.01.017.

Pietruczuk, A., and J.W. Krzyścin (2008), Variability of aerosols forcing on the surface UV radiation: Analysis of data taken at Belsk, Poland, in spring 2007, *Publs. Inst. Geophys. Pol. Acad. Sc*. **D-72**, 403, 61-76.

Pietruczuk, A., and J. Podgórski (2009), The lidar ratio derived from sun-photometer measurements at Belsk Geophysical Observatory, *Acta Geophys.* **57**, 2, 476-493, DOI: 10.2478/s11600-009-0006-9.

WMO (2003), Scientific Assessment of Ozone Depletion: 2002, Global Ozone Research and Monitoring Project – Report No. 47, World Meteorological Organization, Geneva, 498 pp.

Zerefos, C. (1997), Factors influencing the transmission of the solar ultraviolet irradiance through the Earth's atmosphere. **In:** C.S. Zerefos and A.F. Bais (eds.), *Solar Ultraviolet Radiation, Modelling, Mesurements and Effects*, NATO ASI Ser. I, Vol. 52, Springer Verlag, New York, 133-141.

Received 2 April 2009
Received in revised form 29 January 2010
Accepted 26 February 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1147-1163
DOI: 10.2478/s11600-010-0034-5

A Multiyear Analysis of Aerosol Optical Thickness over Europe and Central Poland Using NAAPS Model Simulation

Aneta E. MACISZEWSKA[1], Krzysztof M. MARKOWICZ[2], and Marcin L. WITEK[2]

[1]Institute of Meteorology and Water Management, Warszawa, Poland
e-mail: aneta.maciszewska@imgw.pl (corresponding author)

[2]Institute of Geophysics, University of Warsaw, Warszawa, Poland
e-mails: kmark@igf.fuw.edu.pl, mwit@igf.fuw.edu.pl

A b s t r a c t

This study contains a comparative analysis of aerosol optical thickness (AOT) between numerical calculations obtained from the Navy Aerosol Analysis and Prediction System (NAAPS) model and direct observations from the AERONET robotic network and the Saharan Aerosol over WArsaw (SAWA) field campaign. AOT was calculated for 500 nm wavelength. The comparison shows underestimation of the total aerosol optical thickness simulated by NAAPS. The correlation coefficients between model and observation oscillates between 0.57 and 0.72. Results of seven-year (1998-2004) NAAPS simulation of aerosol components (sea salt, mineral dust, sulphate, and smoke) show large temporal and spatial variability of the aerosol optical thickness over Europe. The least polluted region is the Iberian Peninsula, while the highest aerosol burdens occurred in Central Europe, mostly due to anthropogenic sulphate particles. Finally, the analysis of mineral dust transport shows frequent episodes of Saharan dust inflow over Central Europe. There are about 20 days a year (4 days in May) when instantaneous AOT associated with mineral dust aerosol increases over 0.1.

Key words: aerosol, optical properties, pollution transport model.

1. INTRODUCTION

Atmospheric aerosols play an important role in regulating the Earth's radiative balance (IPPC 2007, Charlson *et al.* 1992). Aerosol influences on climate are commonly divided into direct, indirect and semi-direct aerosol effects. The direct aerosol effect (Coakley and Cess 1985) originates from scattering and absorption of the short- and long-wave solar radiation by aerosol particles in the atmosphere (McCormick and Ludwig 1967). These radiative processes lead to changes in the planetary albedo and reduction of the solar energy reaching the Earth's surface (Grassl and Newiger 1982). Aerosol radiative effects partially balance the positive forcing induced by greenhouse gases. However, large uncertainties remain in evaluation of the aerosol influence on climate. These uncertainties originate mostly from heterogeneous distribution of aerosol emission sources, highly variable aerosol optical properties and their short lifetime in the atmosphere. Significant progress has been made in the recent decade to reduce these deficiencies. Still, our knowledge on regional aerosol variability and aerosol influence on local climate remains low. This is especially pertinent to the territory of Poland and Central Europe, which lack comprehensive studies and analyses.

In this work we analyze seasonal variability of the aerosol optical thickness over Europe. To do this we use numerical simulations of the Navy Aerosol Analysis and Prediction System global pollutions transport model. To validate model results we employ observational data from the AERONET robotic network (Holben *et al.* 2001) as well as direct optical depth observations of the Saharan Aerosol over WArsaw (SAWA) experiment (Markowicz 2008).

2. METHODOLOGY

The Navy Aerosol Analysis and Prediction System (NAAPS) long-range pollution transport model was developed by the Naval Research Laboratory in Monterey (NRL). Detailed description of the model can be found in Christensen (1997) and Witek *et al.* (2007). The NAAPS model uses meteorological fields obtained from the Navy Operational Global Atmospheric Prediction System (NOGAPS) global circulation model (Hogan and Rosmond 1991).

The current version of NAAPS includes gaseous SO_2 and four aerosol components: mineral dust, sea salt, particulate sulphates (SO_4) and smoke. For each of the aerosol species, source areas and values of emission from the Earth's surface are parameterized. Emitted aerosol undergoes advection by large scale flow and diffusion by turbulent processes, and it is removed from the atmosphere through dry deposition and precipitation processes.

In NAAPS, the emission sources are defined separately for each of the aerosol components. Mineral dust emission areas are characterized by the U.S. Geological Survey (USGS) Land Cover Characteristic Database (Anderson *et al.* 1976). Dust is lifted from the surface whenever the friction velocity exceeds a threshold value (60 cm/s) and the surface moisture is less than 30%. The employed emission parameterization is proportional to friction wind velocity raised to the 4th power (Westphal *et al.* 1988). The sea salt emission parameterization follows the Monahan formulation (Monahan *et al.* 1986) and is proportional to surface wind velocity raised to the 3.41. The sulfur dioxide emission in NAAPS is based on the Annual SO_x/NO_x Global Emissions Inventory (GEIA), version 1A, for the year 1985 with a seasonal variability and two-level vertical distribution (Benkowitz and Scholtz 1996). The anthropogenic emissions are converted to 95% SO_2 and 5% particulate sulfate, whereas the natural emission of dimethylosulfate (DMS) is converted to 43% SO_2 and 5% sulfates. The GEIA inventory has 1×1 degree resolution. The sulfur dioxide SO_2 is oxidized in NAAPS model to SO_4^- using linear first-order oxidation rate. This relationship is a function of the solar zenith angle, latitude and time of year (Christensen 1997). Finally, the smoke emission in the NAAPS model is based on geostationary satellite data products (Wildfire ABBA) and MODIS fire products (http://rapidfire.sci. gsfc.nasa.gov/firemaps/). Both biomass burning products are generated under the Fire Locating and Modeling of Burning Emissions (FLAMBE) project (Reid *et al.* 2009) (http://www.nrlmry.navy.mil/flambe/).

Prognostic variables in NAAPS are aerosol mass mixing ratios q_i. From q_i the number of aerosol particles at each model level j is inferred using the relation

$$N_i(j) = \frac{3}{4} \frac{q_i(j)\,\rho_{\text{air}}(j)}{\pi r_i^3 \rho_i} \quad , \tag{1}$$

where $q_i(j)$ is the aerosol mass mixing ratio at level j, $\rho_{\text{air}}(j)$ is the air density at level j, r_i is the aerosol radius, and ρ_i is the aerosol density. In eq. (1) a monodisperse size distribution is assumed. Having the number of aerosol particles, optical properties of aerosol mixture can be obtained assuming characteristic optical properties within each aerosol class. Here the class optical properties are incorporated after the optical properties of aerosols and clouds database (Hess *et al.* 1998).

The horizontal resolution used in NAAPS simulations is 1×1 degree and vertical 25 sigma-coordinate levels are employed. The numerical calculations of aerosol optical thickness are compared with measurements of the Aerosol Robotic Network – AERONET (Holben *et al.* 2001) and direct observations performed during the SAWA (SAharan dust over WArsaw) field

campaign. The AERONET robotic network uses CIMEL sunphotometers (CE 318) – automatic sun and sky scanning radiometers which measure direct solar radiation at eight spectral channels and angular distribution of scattered radiation. The AERONET robotic network delivers a long-term and easily available database of atmospheric aerosol optical and microphysical properties.

The SAWA measurement campaign (Markowicz *et al.* 2008) was carried out in spring of 2005 in Radiation Transfer Observatory at the Institute of Geophysics, University of Warsaw. The main goal was to study optical properties of nonspherical dust aerosol particles transported from Sahara desert over Poland. Aerosol optical thickness was measured using a 5-channel microtops sunphotometer (Morys *et al.* 2001) and a Multi-Filter Rotating Shadowband Radiometer (Model MFR-7) (Harrison *et al.* 1994). Lidar measurements were also performed to obtain vertical aerosol distribution in the troposphere.

3. COMPARISON OF NUMERICAL RESULTS WITH VALUES MEASURED BY SUN PHOTOMETERS

The NAAPS model was validated in marine conditions by Witek *et al.* (2007). However, no studies are available to compare NAAPS with observations over regions with strong influence of continental aerosol. In this section we validate the AOT obtained from NAAPS *versus* observations performed within AERONET (at 500 nm and 2.0 data level) and measurements performed during the SAWA campaign. For this purpose, data from two AERONET stations, the Polish Academy of Sciences station at Belsk (60 km south from Warsaw) and the Leipzig station (Germany) together with data from the SAWA campaign were analyzed. Investigated time periods covered 2001-2004 for the Leipzig station, 2002-2004 for Belsk, and 23 March 2005 to 23 May 2005 for the SAWA campaign. The AERONET observations were additionally averaged over ±3 hours time interval to facilitate comparison with model results.

Figure 1 presents temporal variability of the AOT (at 500 nm) in Belsk and Leipzig in 2002 as well as the AOT derived from the NAAPS model. Comparison between the observations and model results shows reasonably good agreement. However, NAAPS usually underestimates modeled AOT's and does not resolve some aerosol events accurately. Figure 2a, b, c presents scatter plots of the model results and observations from the Belsk and Leipzig stations and the SAWA campaign. Calculated correlation coefficients are 0.62 for Leipzig, 0.57 for Belsk, and 0.72 for the SAWA measurements. A higher value of the correlation coefficient in the SAWA case could be related to different data collecting and processing technique. During SAWA, photo-

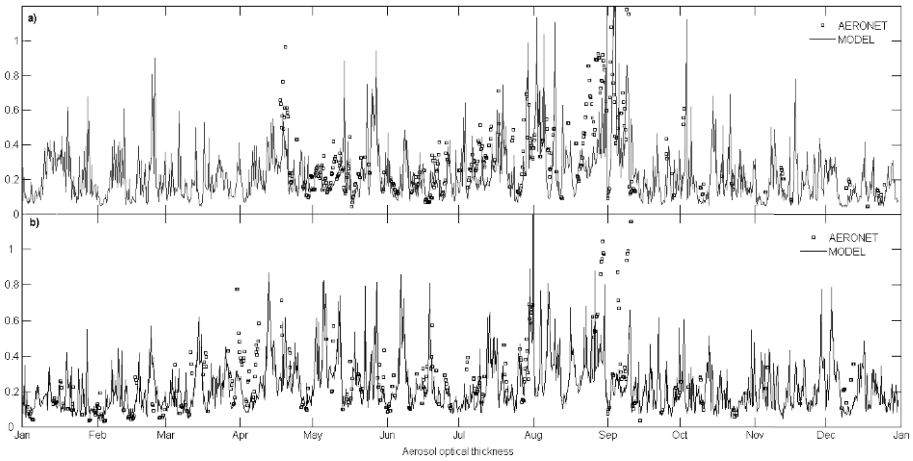


Fig. 1. Temporal variability of the aerosol optical thickness (at 500 nm and 2.0 data level) measured (squares) at Belsk (upper panel), and Leipzig (bottom pannel), and simulated by NAAPS (solid line) in 2002.

metric measurements were performed only when the Sun was not obscured by clouds. In addition to visual inspection, a whole sky camera was used to determine sky conditions close to the Sun. For AERONET measurements, the cloud filtering is based on temporal and spectral signal variability (Smirnov *et al.* 2000). In this method, some thin clouds can still influence the final AOT results.

Figure 2 also shows systematic difference between the observations and NAAPS results, with NAAPS underestimating aerosol optical depths. These differences can result from model deficiencies as well as from assumed aerosol optical properties. A coarse model resolution (1°×1°), simplified parameterizations of turbulent mixing and aerosol deposition processes, or a lack of chemical transformation mechanisms may influence the model accuracy. In addition, some aerosol types, for example anthropogenic nitrates, are not included in the simulations. Another source of NAAPS uncertainty can result from meteorology not being properly modeled by NOGAPS. We used NOGAPS analyzed fields (after data assimilation) that still can contain errors and biases that in turn can affect aerosol emission (*e.g.*, sea salt, mineral dust) and transport processes. However, it extends beyond the scope of this study to analyze the impact of these model errors on long term AOT statistics in Central Europe.

Other simplifications include the assumed monodisperse aerosol size distribution in mass-to-number conversion required for optical properties computations and the use of prescribed optical characteristics for each to the aerosol types.

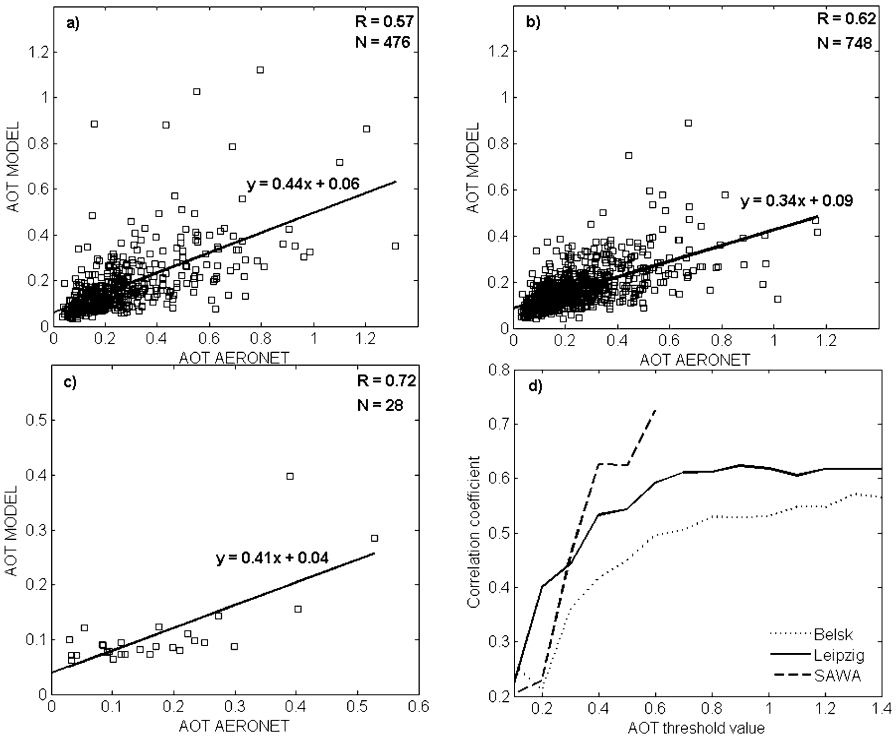


Fig. 2. Comparison of the total optical thickness derived from NAAPS and measured by AERONET station at: (a) Belsk, (b) Leipzig, and (c) during SAWA experiment. Panel (d) shows variability of the correlation coefficient as a function of the threshold AOT below which the correlation coefficient was derived. Data sets from Belsk, Leipzig, and SAWA measurements are presented.

Figure 2d presents variability of the correlation coefficient as a function of the AOT cutoff value. For a given cutoff threshold the correlation coefficient was calculated based on dataset limited to AOT's smaller than this threshold. In all cases the correlation coefficient increases with growing AOT threshold. This indicates that model predictions improve for higher aerosol events and also that the model has difficulties in representing low continental aerosol concentrations, as denoted by small correlation values for AOT's below 0.3. Those relatively clean continental conditions are often affected by anthropogenic emissions which are not fully resolved in NAAPS.

4. SEASONAL VARIABILITY OF AEROSOL OPTICAL THICKNESS OVER EUROPE

In this section the spatial and seasonal variability of the AOT at 500 nm obtained from the NAAPS re-analysis are discussed. The results of 7-year

(1998-2004) model simulations are averaged to define typical seasonal cycles.

Figure 3 presents the annual mean of AOT distribution over Europe. The average AOT value for Europe is about 0.16 at 500 nm. The maximum AOT is located in Central Europe. In that region, average AOT exceeds 0.26. The less polluted areas include Spain, southern France and central Turkey, with the average annual AOT of about 0.1. A strong spatial variability of the AOT is a result of heterogeneous pollution source distribution and a complex atmospheric circulation. The values simulated by NAAPS correspond to AOTs given by other global models (Chin *et al.* 2002, Ginoux *et al.* 2006), annual AOT given by Schaap *et al.* (2004), and PM10 discussed by Matthias (2008). Tombette *et al.* (2008) showed based on Size-Resolved Aerosol Model SIREAM (Debry *et al*. 2007) a similar spatial distribution and absolute values of the total AOT. Large differences were found only in the Po Valley in Northern Italy, where SIREAM showed higher AOTs (about 0.3 at 550 nm) comparing to NAAPS.

In Figure 4, the annual mean AOT distributions of SO_4 (a), dust (b), soot (c), and sea salt (d) are presented. Sulfate particles dominate in Central Europe where average AOT of SO_4 reaches 0.15. This result agrees with some previous studies of sulfate aerosols in Europe (*e.g*., Redington *et al.* 2002). Mineral dust dominates in southern parts of the Mediterranean Sea (AOT exceeds 0.12), which is associated with relatively short distance from the Sahara and Middle-East deserts. The highest values of sea salt AOT (up to 0.1, see Fig. 4d) are found over Atlantic Ocean and the North Sea. This is a result of strong sea salt emission in those regions associated with intense

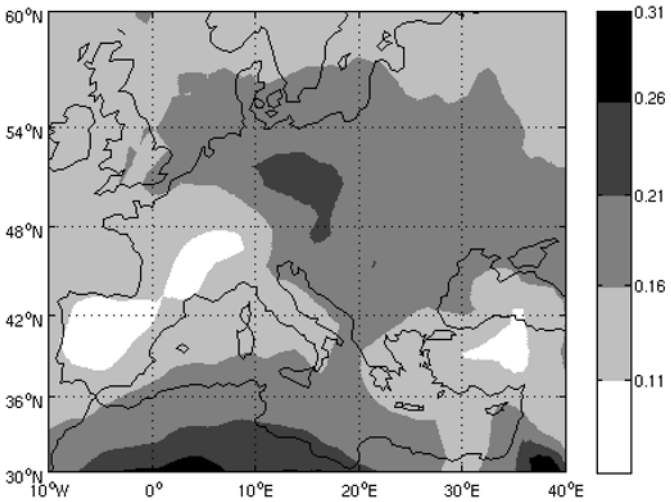


Fig. 3. Annual average aerosol optical thickness (AOT) distribution over Europe.

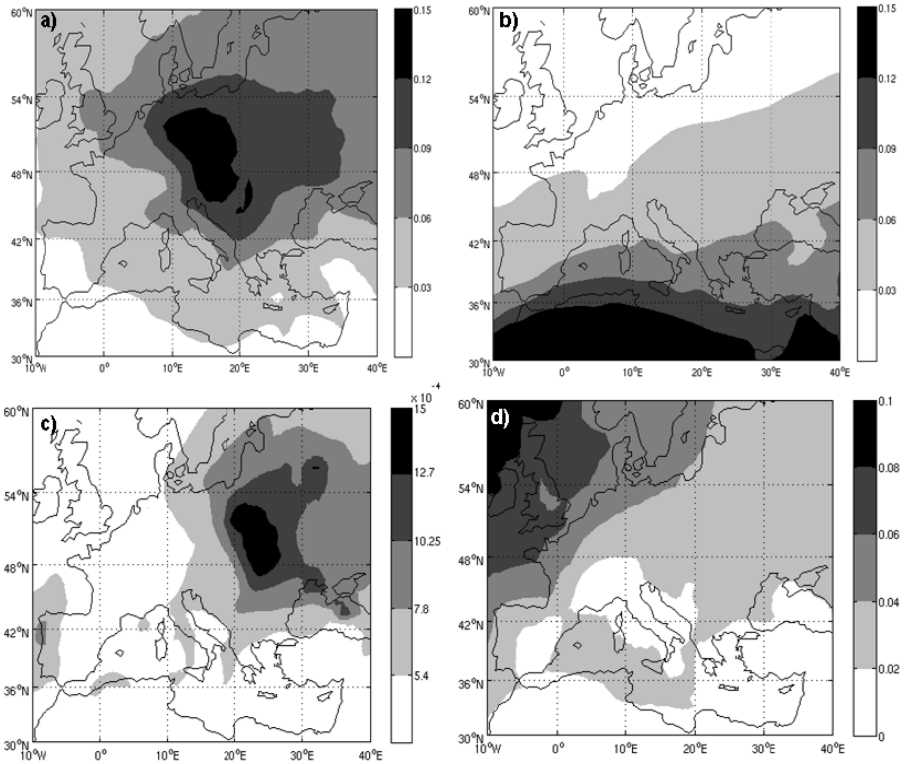


Fig. 4. Annual mean of aerosol optical thickness distribution over Europe for SO_4 (a), dust (b), soot (c), and sea salt particles (d).

cyclone activity and high surface wind speeds. Soot aerosol dominates in Eastern Europe with annual average AOT's reaching 1.5×10^{-3}. These results are consistent with satellite observation of fires activity in Europe (Giglio *et al.* 2005, 2006). Soot emissions are a consequence of forest and agricultural fires in those regions. A similar activity but of a smaller magnitude is observed in the western Iberian Peninsula. However, a seasonal and localized nature of such fire events is responsible for small annual averages of soot AOT. Despite such small values this aerosol has an important climate effect due to its large absorption coefficient. NAAPS simulations of the single scattering albedo (not presented here) show a significant contribution from soot particles.

Figure 5 shows a spatial distribution of the NAAPS derived total mean AOT during spring (a), summer (b), autumn (c), and winter (d). The mean values of AOT for Europe are: 0.13, 0.21, 0.16, and 0.14 for winter, spring, summer, and autumn, respectively. The spring season is characterized by a strong dust activity in the Saharan region (Moulin 1997). Therefore, the

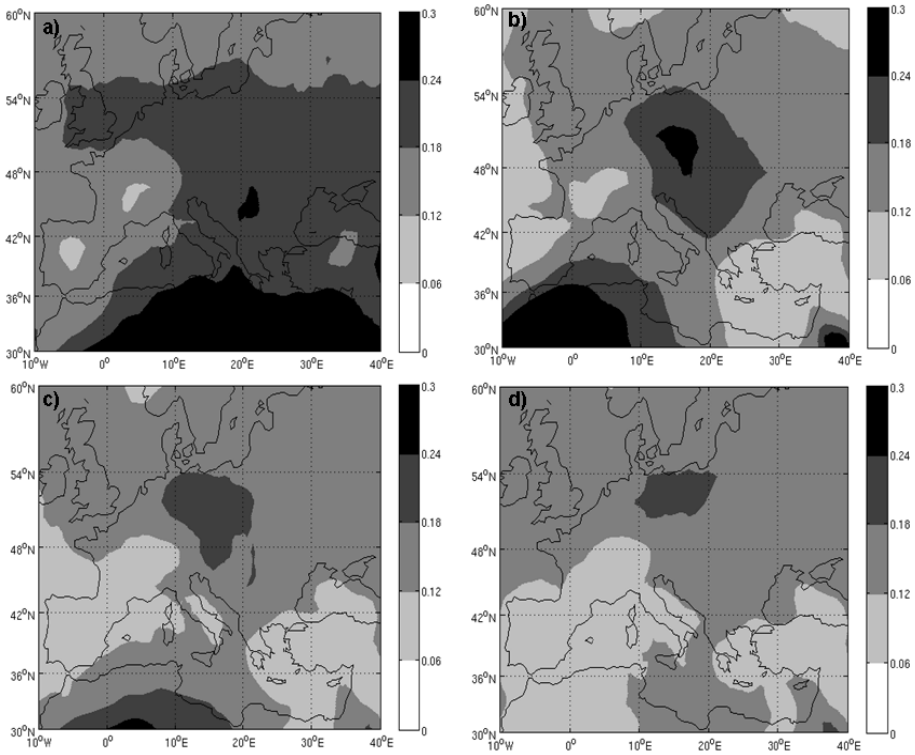


Fig. 5. Seasonal mean aerosol optical thickness in spring (a), summer (b), autumn (c), and winter (d).

largest AOTs are located in the southern part of Mediterranean Sea. During summer, the dust activity subsides and shifts towards western parts of Sahara, causing Mediterranean Sea to be less dust-polluted. Another summer maximum is modeled in Central Europe (Fig. 5b), mostly due to anthropogenic sulfates pollutions. Spatial distribution of AOT in autumn is similar to the summer one, but the magnitude of AOT is reduced in autumn. In winter, the modeled AOT's are the lowest of all seasons. A minor maximum not exceeding 0.24 is observed in Central Europe, whereas in many parts of Europe AOT is below 0.12 (Fig. 5d).

5. TEMPORAL VARIABILITY OF AEROSOL OPTICAL THICKNESS IN THE PROXIMITY OF WARSAW, POLAND

In this section we discuss seasonal variability of the AOT obtained from the NAAPS re-analysis. Figure 6 presents monthly means of the AOT in the proximity of Warsaw, Poland. The highest AOT's are modeled for spring and summer seasons, with pronounced maxima exceeding 0.2 in April and

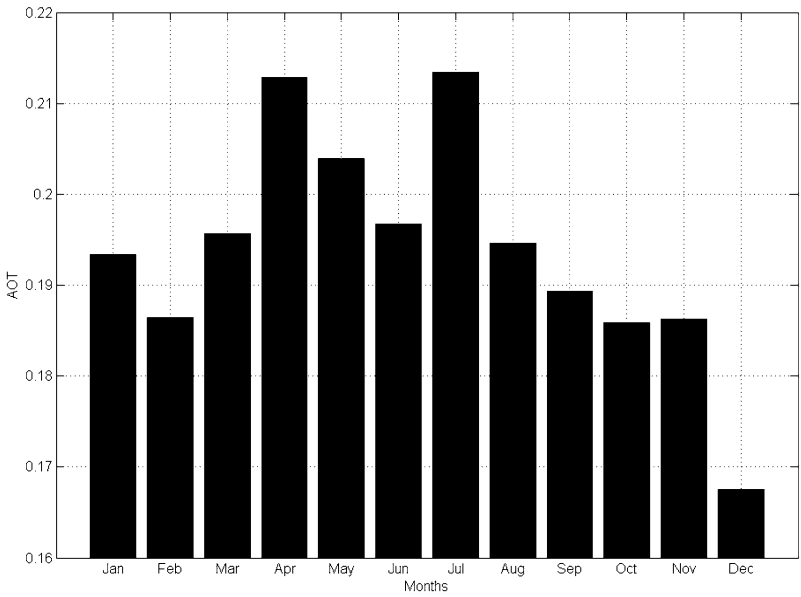


Fig. 6. Monthly values of the aerosol optical thickness over Central Poland averaged over the period 1998-2004.

July. The minimum AOT occurs in December and is below 0.17. This annual cycle is mostly caused by variability in sulfate particles concentration and seasonal occurrence of mineral dust episodes.

Figure 7 shows seasonal contribution of the aerosol components to the total AOT over Central Poland. In all seasons, anthropogenic sulfates comprise more than 50% of the total AOT, peaking in summer (71%) and having a minimum in winter (51%). A similar contribution (40-60%) of sulfate AOT to the total AOT over Central Europe has been shown by Goddard Chemistry Aerosol Radiation and Transport (GOCART) model (Chin *et al.* 2002). Substantial impact of this aerosol results from continuous emissions in the region, as opposed to intermittent occurrence of other aerosol types. Another large contributor is the mineral dust, which accounts for over 20% AOT during spring and more than 10% in all other seasons. The dust aerosol is most often transported over Poland during spring time due to frequent dust storms over Sahara and prevailing large-scale circulation patterns. Smoke aerosol is decomposed in Fig. 7 into soot and organic soluble aerosols, as these co-products of biomass burning have different optical characteristics. The total biomass burning mass (smoke class) is converted to 95% organic and 5% soot aerosols. Organic soluble aerosol amounts to up to 10% of the total AOT during summer, which is caused by frequent natural and human induced fires during this season. In other seasons the influence of smoke

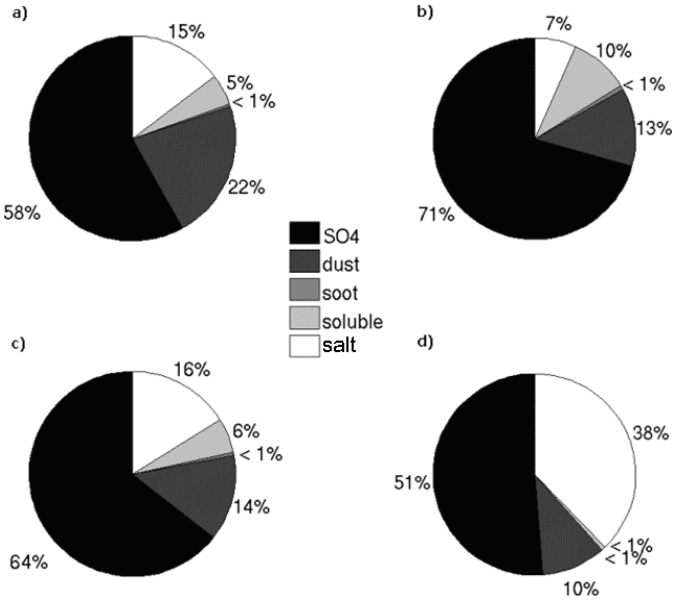


Fig. 7. Seasonal contribution to AOT (in percents) of each of the modeled aerosol components over Central Poland in spring (a), summer (b), autumn (c), and winter (d).

aerosol drops below 5%. These values are somehow smaller than GOCART results; Chin *et al.* (2002) showed contribution of organic and black carbon between 20-40%. The last, sea salt aerosol class contributes from 15% during polluted springs up to 38% during clean winters to the total AOT modeled over Central Poland. In all cases it is caused by advection of air masses loaded with sea salt particles from Northern Atlantic and North Sea. It is expected that the actual sea salt concentrations over Central Poland would be somehow smaller due to chemical reactions and transformation of those particles during transport in the continental boundary layer. Many of such microphysical and chemical mechanisms affecting aerosol life-time are not included in the NAAPS model.

The results of NAAPS simulations show frequent occurrence of mineral dust over Poland. Saharan dust events were observed during the SAWA experiment in 2005 (Markowicz *et al.* 2008). Results of more than two months of lidar and sun-photometer observations confirm several dust events over Warsaw. During those episodes the total AOT raised over 0.7 at 500 nm wavelength. In order to assess the frequency of dust transport over Poland we defined a mineral dust AOT threshold, equal to 0.1 and analyzed the NAAPS output in terms of this threshold value. Figure 8 shows monthly averaged percentage of days with dust AOT exceeding the threshold over Central Poland. One can distinguish two dust seasons (spring and autumn) in

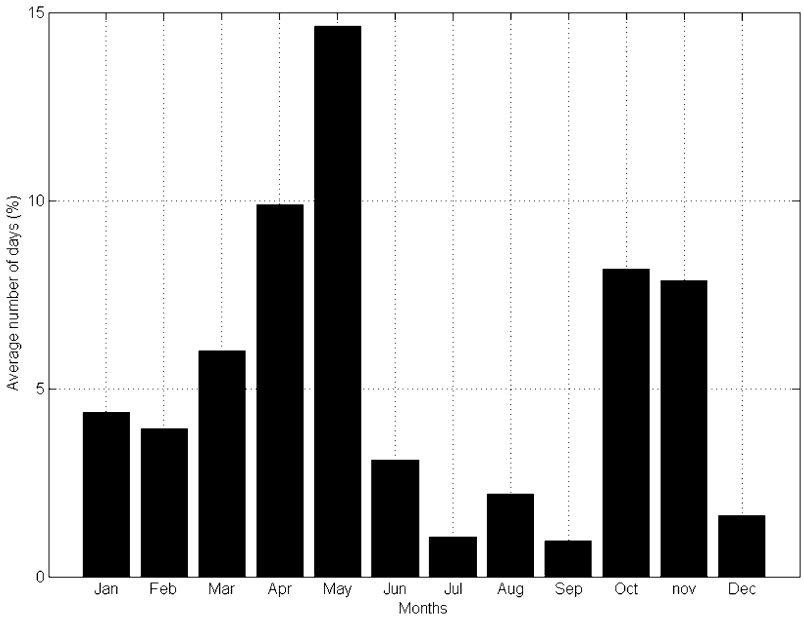


Fig. 8. Average number of days (%) with mineral dust aerosol inflow over Central Poland with the optical thickness exceeding 0.10.

Poland. Spring events are more frequent, especially in May when during 15% of days dust AOT's exceed 0.1. On the other hand, the transport of mineral aerosol over Poland in summer is negligible (less than 2%). The seasonal cycle of mineral dust events obtained from NAAPS agrees well with the Dust REgional Atmospheric Model (DREAM) (Nickovic *et al.* 2001) and lidar observations in frame of the European Aerosol Research Lidar Network (EARLINET) (Bösenberg *et al.* 2003). EARLINET measurements show somehow smaller frequency of dust events in Europe comparing to the models. Papayannis *et al*. (2009) suggested this could be due to the intermittency of lidar measurements, caused by unfavorable conditions such as the presence of low clouds, rain, or technical problems. Observations performed during the SAWA experiment confirm mostly cloudy condition during Saharan dust events because of the advection of warm and wet air masses from the Mediterranean Sea.

6. SUMMARY

In this study, spatial and temporal variability of aerosol optical thickness over Europe is analyzed based on NAAPS global aerosol transport model simulations. Average characteristics are derived for the period 1998-2004. NAAPS model results are compared against AERONET AOT observations

as well as results from the SAWA experiment. Correlation coefficients vary between 0.72 and 0.57 for the analyzed datasets. Model deficiencies influence simulations accuracy, especially insufficient representation of anthropogenic aerosols and simplified parameterization of aerosol optical properties. Despite these drawbacks, the model results provide valuable information on aerosol distribution over Europe.

The most polluted region in Europe is its central part, where annually averaged AOT exceeds 0.21. Those pollutions are mostly anthropogenic sulphates that contribute more than 50% to the total AOT. The least aerosol burdened regions include Iberian Peninsula, Southern France and Turkey, with AOT below 0.11. Mineral dust aerosol dominates in Southern Europe and over the Mediterranean Sea due to the proximity of dust-emitting regions in Northern Africa and Middle East. Wild and agricultural fires in Central and Eastern Europe increase soot aerosol concentrations in those regions, but the average soot AOT does not exceed 0.01 due to sporadic nature of such fires. Finally, sea salt aerosol dominates over stormy Northern Atlantic (average AOT reaching 0.1), but model results also indicate effective sea salt advection over continental Europe. The most polluted season in Europe is summer and the lowest aerosol concentrations are modeled in winter.

In Poland, monthly averaged aerosol AOT's vary between 0.16 and 0.21. The highest values are modeled in spring and summer. Dust transport events are most frequent in spring (10-15% of days) and autumn (8% of days) over Poland. In total there is about 20 days a year (4 days in May) when instantaneous AOT associated with mineral dust aerosol increases over 0.1.

Acknowledgements. The authors wish to thank D. Westphal and P. Flatau from Naval Research Laboratory in Monterey for providing NAAPS source code and model input. We also thank the NASA Goddard AERONET team led by B. Hoblen and station managers Albert Ansmann and Piotr Sobolewski for producing surface observations at Leipzig and Belsk stations, respectively.

References

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer (1976), A land use and land cover classification system for use with remote sensor data, *Geol. Surv. Prof. Paper* 964, 1-28.

Benkowitz, C.M., M.T. Scholtz, J. Pacyna, L. Tarrasón, J. Dignon, E.C. Voldner, P.A. Spiro, J.A. Logan, and T.E. Graedel (1996), Global gridded invento-

ries of anthropogenic emissions of sulfur and nitrogen, *J. Geophys. Res.* **101**, D22, 29239-29253, DOI: 10.1029/96JD00126.

Bösenberg, J., V. Matthias, A. Amodeo, V. Amoiridis, A. Ansmann, J.M. Baldasano, I. Balin, C. Böckmann, A. Boselli, G. Carlsson, A. Chaykovski, G. Chourdakis, A. Comeron, F. DeTomasi, R. Eixmann, V. Freudenthaler, H. Giehl, I. Grigorov, A. Hågård, M. Iarlori, A. Kirsche, G. Kolarov, L. Komguem, S. Kreipl, W. Kumpf, G. Larchevêque, H. Linne, R. Matthey, I. Mattis, A. Mekler, I. Mironova, V. Mitev, L. Mona, D. Müller, S. Music, S. Nickovic, M. Pandolfi, A. Papayannis, G. Pappalardo, J. Pelon, C. Peres, R.M. Perrone, R. Persson, D.P. Resendes, V. Rizi, F. Rocadenbosch, J. Rodriguez, L. Sauvage, L. Schneidenbach, R. Schumacher, V. Shcherbakov, V. Simeonov, P. Sobolewski, N. Spinelli, I. Stachlewska, D. Stoyanov, T. Trickl, G. Tsaknakis, G. Vaughan, U. Wandinger, X. Wang, M. Wiegner, M. Zavrtanik, and C. Zerefos (2003), EARLINET: A European Aerosol Research Lidar Network to establish an aerosol climatology, MPI-Report, Max-Planck-Institute for Meteorology, Hamburg, Germany, 348 pp.

Charlson, R.J., S.E. Schwartz, J.M. Hales, R.D. Cess, J.A. Coakley, Jr., J.E. Hansen, and D.J. Hofmann (1992), Climate forcing by anthropogenic aerosols, *Science* **255**, 5043, 423-430, DOI: 10.1126/science.255.5043.423.

Chin, M., P. Ginoux, S. Kinne, O. Torres, B.N. Holben, B.N. Duncan, R.V. Martin, J.A. Logan, A. Higurashi, and T. Nakajima (2002), Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements, *J. Atmos. Sci*. **59**, 3, 461-483, DOI: 10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2.

Christensen, J.H. (1997), The Danish Eulerian Hemispheric Model – a three-dimensional air pollution model used for the Arctic, *Atm. Env.* **31**, 24, 4169-4191, DOI: 10.1016/S1352-2310(97)00264-1.

Coakley, J.A., Jr., and R.D. Cess (1985), Response of the NCAR Community Climate Model to the radiative forcing by the naturally occurring tropospheric aerosol, *J. Atmos. Sci.* **42**, 16, 1677-1692, DOI: 10.1175/1520-0469(1985)042<1677:ROTNCC>2.0.CO;2.

Debry, E., K. Fahey, K. Sartelet, B. Sportisse, and M. Tombette (2007), Technical note: A new SIze REsolved Aerosol Model (SIREAM), *Atmos. Chem. Phys*. **7**, 1537-1547, DOI: 10.5194/acp-7-1537-2007.

Giglio, L., G.R. van der Werf, J.T. Randerson, G.J. Collatz, and P. Kasibhatla (2005), Global estimation of burned area using MODIS active fire observations, *Atmos. Chem. Phys*. *Discuss*. **5**, 6, 11091-11141.

Giglio, L., I. Csiszar, and C.O. Justice (2006), Global distribution and seasonality of active fires as observed with the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, *J. Geophys. Res*. **111**, G02016, DOI: 10.1029/2005JG000142.

Ginoux, P., L.W. Horowitz, V. Ramaswamy, I.V. Geogdzhayev, B.N. Holben, G. Stenchikov, and X. Tie (2006), Evaluation of aerosol distribution and

optical depth in the Geophysical Fluid Dynamics Laboratory coupled model CM2.1 for present climate, *J. Geophys. Res*. **111**, D22210, DOI: 10.1029/2005JD006707.

Grassl, H., and M. Newiger (1982), Changes of local planetary albedo by aerosol particles, *Sci. Total Environ*. **23**, 313-320, DOI: 10.1016/0048-9697(82)90148-6.

Harrison, L., J. Michalsky, and J. Berndt (1994), Automated multifilter rotating shadow-band radiometer: an instrument for optical depth and radiation measurements, *Appl. Optics* **33**, 22, 5118-5125, DOI: 10.1364/AO.33.005118.

Hess, M., P. Koepke, and I. Schult (1998), Optical Properties of Aerosols and Clouds: The software package OPAC, *Bull. Am. Meteorol. Soc*. **79**, 5, 831-844, DOI: 10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2.

Hogan, T.F., and T.E. Rosmond (1991), The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model, *Monthly Weather Rev*. **119**, 8, 1786-1815, DOI: 10.1175/1520-0493(1991)119<1786:TDOTNO>2.0.CO;2.

Holben, B.N., D. Tanré, A. Smirnov, T.F. Eck, I. Slutsker, N. Abuhassan, W.W. Newcomb, J.S. Schafer, B. Chatenet, F. Lavenu, Y.J. Kaufman, J.V. Castle, A. Setzer, B. Markham, D. Clark, R. Frouin, R. Halthore, A. Karneli, N.T. O'Neill, C. Pietras, R.T. Pinker, K. Voss, and G. Zibordi (2001), An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, *J. Geophys. Res*. **106**, 12067-12098, DOI: 10.1029/2001JD900014.

IPCC (2007), Summary for Policymakers. **In:** S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK, and New York, NY, USA, http://www.ipcc.ch/SPM2feb07.pdf.

Markowicz, K.M., P.J. Flatau, A.E. Kardas, J. Remiszewska, K. Stelmaszczyk, and L. Woeste (2008), Ceilometer retrieval of the boundary layer vertical aerosol extinction structure*, J. Atmos. Ocean. Tech*. **25**, 928-944, DOI: 10.1175/2007JTECHA1016.1.

Matthias, V. (2008), The aerosol distribution in Europe derived with the Community Multiscale Air Quality (CMAQ) model: comparison to near surface in situ and sunphotometer measurements, *Atmos. Chem. Phys. Discuss.* **8**, 1457-1503.

McCormick, R.A., and J.H. Ludwig (1967), Climate modification by atmospheric aerosols, *Science* **156**, 3780, 1358-1359, DOI: 10.1126/science.156.3780.1358.

Monahan, E.C., D.E. Spiel, and K.L. Davidson (1986), A model of marine aerosol generation via whitecaps and wave disruption. **In:** E.C. Monahan and

G. MacNiocaill (eds.), *Oceanic Whitecaps and Their Role in Air-Sea Exchange Processes*, D. Reidel Publ. Co., Dordrecht, New York, 167-174.

Morys, M., F.M. Mims III, S. Hagerup, S.E. Anderson, A. Baker, J. Kia, and T. Walkup (2001), Design, calibration, and performance of MICROTOPS II handheld ozone monitor and Sun photometer, *J. Geophys. Res.* **106**, D13, 14573-14582, DOI: 10.1029/2001JD900103.

Moulin, C., F. Dulac, C.E. Lambert, P. Chazette, I. Jankowiak, B. Chatenet, and F. Lavenu (1997), Long-term daily monitoring of Saharan dust load over ocean using Meteosat ISCCP-B2 data 2. Accuracy of the method and validation using Sun photometer measurements, *J. Geophys. Res.* **102**, D14, 16959-16969, DOI: 10.1029/96JD02598.

Nickovic, S., G. Kallos, A. Papadopoulos, and O. Kakaliagou (2001), A model for prediction of desert dust cycle in the atmosphere, *J. Geophys. Res.* **106**, D16, 18113-18129, DOI: 10.1029/2000JD900794.

Papayannis, A., V. Amiridis, L. Mona, G. Tsaknakis, D. Balis, J. Bösenberg, A. Chaikovski, F. De Tomasi, I. Grigorov, I. Mattis, V. Mitev, D. Müller, S. Nickovic, C. Pérez, A. Pietruczuk, G. Pisani, F. Ravetta, V. Rizi, M. Sicard, T. Trickl, M. Wiegner, M. Gerding, R.E. Mamouri, G. D'Amico, and G. Pappalardo (2008), Systematic lidar observations of Saharan dust over Europe in the frame of EARLINET (2000-2002), *J. Geophys. Res.* **113**, D10204, DOI: 10.1029/2007JD009028.

Papayannis, A., R.E. Mamouri, V. Amiridis, S. Kazadzis, C. Pérez, G. Tsaknakis, P. Kokkalis, and J.M. Baldasano (2009), Systematic lidar observations of Saharan dust layers over Athens, Greece in the frame of EARLINET project (2004-2006), *Ann. Geophys*. **27**, 3611-3620.

Redington, A.L., and R.G. Derwent (2002), Calculation of sulphate and nitrate aerosol concentrations over Europe using a Lagrangian dispersion model, *Atmos. Environ.* **36**, 4425-4439, DOI: 10.1016/S1352-2310(02)00420-X.

Reid, J.S., E.J. Hyer, E.M. Prins, D.L. Westphal, J. Zhang, J. Wang, S.A. Christopher, C.A. Curtis, C.C. Schmidt, D.P. Eleuterio, K.A. Richardson, and J.P. Hoffman (2009), Global monitoring and forecasting of biomass-burning smoke: Description of and lessons from the Fire Locating and Modeling of Burning Emissions (FLAMBE) Program, *IEEE JSTARS* **2**, 3, 144-162, DOI: 10.1109/ JSTARS.2009.2027443.

Schaap, M., M. van Loon, H.M. ten Brink, F.J. Dentener, and P.J.H. Builtjes (2004), Secondary inorganic aerosol simulations for Europe with special attention to nitrate, *Atmos. Chem. Phys*. **4**, 857-874, DOI: 10.5194/acp-4-857-2004.

Smirnov, A., B.N. Holben, T.F. Eck, O. Dubovik, and I. Slutsker (2000), Cloud-screening and quality control algorithms for the AERONET database, *Remote Sens. Environ.* **73**, 3, 337-349, DOI: 10.1016/S0034-4257(00)00109-7.

Tombette, M., P. Chazette, B. Sportisse, and Y. Roustan (2008), Simulation of aerosol optical properties over Europe with a 3-D size-resolved aerosol model:

Comparisons with AERONET data, *Atmos. Chem. Phys*. **8**, 7115-7132, DOI: 10.5194/acp-8-7115-2008.

Westphal, D.L., O.B. Toon, and T.N. Carlson (1988), A case study of mobilization and transport of Saharan dust, *J. Atmos. Sci.* **45**, 15, 2145-2175, DOI: 10.1175/1520-0469(1988)045<2145:ACSOMA>2.0.CO;2

Witek, M.L., P.J. Flatau, P.K. Quinn, and D.L. Westphal (2007), Global sea-salt modeling: Results and validation against multicampaign shipboard measurements, *J. Geophys. Res.* **112**, D08215, DOI: 10.1029/2006JD007779.

Received 2 September 2009
Received in revised form 23 February 2010
Accepted 17 June 2010

VERSITA

Acta Geophysica
vol. 58, no. 6, pp. 1164-1192
DOI: 10.2478/s11600-010-0008-7

On Ionospheric Phenomena During Pre-Storm and Main Phase of a Very Intense Geomagnetic Storm

Victor U. CHUKWUMA

Department of Physics, Olabisi Onabanjo University, Ago-Iwoye, Nigeria
e-mail: victorchukwuma@yahoo.com

A b s t r a c t

An investigation to elucidate the mechanisms responsible for the pre-storm and main phase ionospheric phenomena during November 20-21, 2003, is presented using heliophysical, interplanetary, geomagnetic, and global ionospheric data. The results show that the ionospheric responses in the main phase do not indicate prompt penetration electric fields as the main ionospheric driver. The results also show that the pre-storm phenomena do not originate from a local time effect. The simultaneous occurrence of *fo*F2 enhancements at two widely separated longitudinal zones appeared to suggest a role played by the magnetospheric electric field. However, the analysis of *hm*F2 at the stations could not confirm the notion that these fields are the main drivers of pre-storm phenomena. An investigation of flare effects on the pre-storm phenomena also revealed that solar flares are not the main drivers. The present results appear to suggest that the pre-storm ionospheric phenomena could be a result of some underlying mechanisms that are working together with varying degree of importance.

Key words: geomagnetic storm, solar X-rays, solar wind, shock gas, ionosphere, pre-storm phenomena.

1. INTRODUCTION

In recent years, intense storms of March 13-15, 1989 (*Dst* = –600 nT), October 20-21, 1989 (*Dst* = –266 nT), and April 1-2, 1973 (*Dst* = –211 nT) accompanied by the appearance of a positive storm before the beginning of

a geomagnetic disturbance in the mid-latitudes and a strong negative phase at the equator (termed pre-storm phenomena) have been studied by Chukwuma (2007a) using *fo*F2 data obtained from a global network of ionosonde stations in the East Asian, Euro-African, and American sectors. In the case of the October 20-21, 1989 storm, the phenomena under investigation have been reported at about 16% of the stations already on October 20. These ionospheric storms were caused by the southward turning of *Bz* at ~21:00 UT on October 18 which got to a change in *Bz* of δBz = 12.2 nT at 23:00 UT on that day. In the case of the storm of April 1-2, 1973, the phenomena have been reported at about 69% of the stations; these ionospheric storms appear to be caused by the southward turning of *Bz* at ~15:00 UT which got to a change in *Bz* of δBz = 13.6 nT at 19:00 UT on March 31.

In the light of above-mentioned results, and in regards to the suggestion by Danilov (2001) that the pre-storm phenomena have still some unsolved problems, Chukwuma (2007a) asserted that the difficulty with explaining 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, thereby foreclosing the explanation of any aspect of the morphology of ionospheric storms whose origin precedes the onset reference time. It is important to note that the use of SSC as a reference time is a bad choice (Prolss 1995) because these impulse-like disturbances of the magnetic field are not associated with any significant energy deposition and are also observed after the onset of a magnetic storm, as indicated, for example, by the decrease in the *Dst* index (Akasofu 1970). Also the use of the MPO for fixing the beginning of magnetic and ionospheric storms is fraught with problems that render a determination of the exact onset time difficult (Prolss 1995).

In regards to the true nature of the pre-storm, Mikhailov and Perrone (2009) have suggested that there are no convincing arguments that pre-storm phenomena at middle and sub-auroral latitudes bear a relation to magnetic storms. According to those authors, the *Nm*F2 pre-storm enhancements were due to the previous geomagnetic storm, moderate auroral activity or they represented the class of positive quiet time events and as such there is no such effect as the pre-storm *Nm*F2 enhancement as a phenomenon inalienably related to the geomagnetic storms. In their investigation Mikhailov and Perrone (2009) assumed a criterion for selecting pre-storm phenomenon which is that a pre-storm *fo*F2 enhancement should precede the magnetic storm onset and take place within a reasonable time interval (say, within 24 h before the SC) and develop under quiet geomagnetic conditions; if an observed *fo*F2 increase does not satisfy this requirement, there is no reason to consider it a pre-storm enhancement. However, it is reasonable to add that

though Mikhailov and Perrone (2009) considered intervals of moderately enhanced auroral activity as disturbed periods, they did not ascertain whether such a pre-storm auroral activity is related to the following magnetic storm.

The literature shows that the main features of ionospheric storm phenomena are explained on the basis of the principal concept: during a geomagnetic disturbance there is an input of energy into the polar atmosphere, which changes thermospheric parameters, such as composition, temperature and circulation. Composition changes directly influence the electron concentration in the 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. This concept appears not to explain most ionospheric phenomena. Attempts have also been presently made at explaining the spatial and temporal variations of ionospheric storms using ionospheric electric fields which have two primary sources: the convection electric field mapped from the magnetosphere and the neutral wind dynamo. The electric field of magnetospheric origin is normally confined to high latitudes in steady state by the shielding effects of the region 2 currents. However, because of the slow reaction of these region 2 currents to changes in solar wind/interplanetary magnetic field (IMF) conditions, the region 2 currents cannot effectively provide the necessary shield when there is a rapid change in solar wind conditions, as such electric fields in the high latitude promptly penetrate into middle and low latitudes (e.g., Kelley *et al*. 1979, Huang *et al*. 2005).

This prompt electric field is eastward during the daytime to dusk sector and westward in the midnight to dawn sector (Jaggi and Wolf 1973, Spiro *et al*. 1988, Fejer *et al*. 1990) and results in a positive and negative ionospheric storm, respectively. The neutral wind dynamo becomes significantly enhanced during geomagnetic storms, producing the so-called disturbance dynamo (Blanc and Richmond 1980). The mechanism for the disturbance dynamo is that the enhanced energy and momentum inputs during storms greatly change the global neutral winds, which in turn cause major changes in the dynamo electric fields. It is important to observe that prompt penetrating electric fields and disturbance electric fields have opposite polarity local time dependences; on the one hand, a prompt penetrating electric field would cause dayside enhancement while, on the other hand, a disturbance dynamo electric field would cause a nightside enhancement. It is plausible then, on the basis of the local time dependency of prompt penetrating electric and dynamo electric fields, to argue that these electric fields may not be able to account for the same form of ionospheric phenomena within the same longitudinal sector in the same time interval.

According to Gonzalez *et al*. (1994), the primary causes of geomagnetic storms at the Earth are strong interplanetary electric fields associated with

the passage of southward directed magnetic fields B_S past the Earth for a sufficiently long interval of time. The electric field is composed of two factors: the solar wind velocity V_{SW} and the southward IMF. Tsurutani *et al*. (1992) demonstrated that it is the extraordinarily high southward *Bz* rather than high V_{SW} that is the dominant part of the electric field. When *Bz* is strongly negative, magnetic reconnection between the IMF and the geomagnetic field produces open field lines which allow mass, energy and momentum to be transferred from the solar wind to the Earth's magnetosphere (Davis *et al*. 1997). Therefore, it could be argued that the results of Chukwuma (2007a) are predicated on interplanetary electric field and it is critical to understand how much of the interplanetary electric field can penetrate to the ionosphere and how long penetration electric fields exist (Huang 2008).

In a recent paper, Chukwuma (2009) investigated the heliophysical and interplanetary phenomena that was responsible for the major storm of Nov. 20-21, 2003. Presently, in this related paper, the solar activity in the period Nov. 18-21, 3003, the nature and development of that same storm and the pre-storm and main phases of the ionospheric storm during Nov. 20-21, 2003, have been examined in order to explain the ionospheric phenomena during this major storm. My interest mainly lies in elucidating the mechanisms responsible for the pre-storm phenomena as well as revealing any possible relationship between such mechanisms and those responsible for the main phase ionospheric phenomena. The solar activity during the period under investigation was mostly dominated by the active region 501 which produced M3.9 and M4.5 flares on Nov. 18, of which the M3.9, with the heliophysical coordinates at N00 E18, produced the CMEs that caused the intense storm of Nov. 20-21, 2003. AR 501 also produced the M1.3, M9.6, and M5.8 flare of Nov. 20 that is associated with the ionospheric storm on this day.

2. DATA AND METHOD OF ANALYSIS

The solar activity data used in this study consists of GOES 5-minute values of Solar X-rays: 1-8 A (Fig. 1a) obtained from Space Physics Interactive Data Resource (SPIDR) (http://spidr.ngdc.noaa.gov). The X-ray flare classes are A, B, C, M and X, corresponding to increasing levels of the measured Solar X-ray flux. The levels are: $A \geq 10^{-8}$ W/m^2, $B \geq 10^{-7}$ W/m^2, $C \geq 10^{-6}$ W/m^2, $M \geq 10^{-5}$ W/m^2, $X \geq 10^{-4}$ W/m^2.

The geomagnetic index and solar wind data consist of hourly values of the low-latitude magnetic index, *Dst*, the total magnetic field *B*, interplanetary electric field E_y, the proton number density N_{sw}, the solar wind flow speed V_{sw}, the solar wind dynamic pressure P_{sw} and the IMF *Bz* component in the GSM coordinates. These data were obtained from NSSDC's OMNIWeb Sevice (http://nssdc.gsfc.nasa.gov/omniweb) and SPIDR (http://spidr.ngdc.noaa. gov).

(Note: in the GSM coordinate system, $\hat{x}$ is towards the sun, $\hat{y}$ is defined by $\hat{n}\times\hat{n}/|\hat{n}\times\hat{x}|$, where $\hat{n}$ is in the south magnetic pole direction, and $\hat{z}$ forms a right-hand system.)

The ionospheric data used in this study consists of hourly values of *fo*F2 obtained from SPIDR's network of ionosonde stations located in the East Asian, Australian, European–African, and American sectors. These stations are listed in Table 1. It is important to note that *fo*F2 data from SPIDR could have errors due to the presence of spread F and many other factors. If necessary, automatically scaled traces have been manually corrected. However, scrutiny of these would need an access to the original, detailed data from individual stations, which is a laborious. According to Kane (2005), these cannot be considered in detail for a general analysis like the present one, and this kind of scrutiny is not done because where they probably occur, errors are expected (hopefully) to be minimized in averaging. Furthermore, according to Zhizhin

Table 1

Ionosonde stations located in four sectors

Stations and their codes	Geographic coordinates		Geomagnetic coordinates		Difference between LST and UT (in hours)
	φ	λ	φ	λ	
East Asian sector					
Manzhouli (MAN)	49.60°N	117.50°E	35.62°N	207.3°E	+8
Chongqing (CHO)	29.50°N	106.40°E	29.72°N	206.7°E	+7
Guangzhou (GUA)	23.10°N	113.40°E	25.70°N	206.7°E	+8
Hainan (HAI)	18.30°N	109.30°E	15.45°N	196.4°E	+7
Australia sector					
Darwin (DAR)	–12.50°N	131.00°E	3.57°N	191.1°E	+9
Learmonth (LEA)	–21.90°N	114.00°E	–32.58°N	187.10°E	+8
Mundaring (MUN)	–32.00°N	116.40°E	–42.61°N	190.20°E	+8
European–Africa sector					
Juliusruh/Rugen (JUL)	54.60°N	13.40°E	56.89°N	103.80°E	0
Rome (ROM)	41.80°N	12.50°E	47.12°N	92.80°E	+0
Athens (ATH)	38.00°N	23.50°E	36.09°N	104.20°E	+1
Grahamstown (GRA)	–33.30°N	26.50°E	–34.31°N	91.70°E	+1
American sector					
Goosebay (GOO)	53.30°N	–60.40°E	56.68°N	353.0°E	–4
Milestone Hill (MIL)	42.60°N	–71.50°E	48.93°N	318.2°E	–5
Wallops Island (WAL)	37.80°N	–75.50°E	42.32°N	302.4°E	–5
Puerto Rico (PUE)	18.50°N	–67.20°E	29.29°N	6.1°E	–4
Jicamarca (JIC)	–12.10°N	–77.00°E	–0.98°E	355.70°E	–5

et al. (2008), the data held in SPIDR can be quality controlled by peer-matching techniques where stations are compared to the nearest neighbours to see if the observations are "similar" (in a fuzzy mathematical sense); however, in the absence of known parallel stations the SPIDR data are used as they are, with removal of off-range values, using consistency of data over two solar cycle. Presently there was no need for elimination of any data point.

The present study of global ionospheric response to heliophysical and interplanetary forcing is concerned with variations in *fo*F2 during Nov. 20-21, 2003. However, the F2 region response to geomagnetic storms is most conveniently described in terms of *D*(*fo*F2), that is, the normalized deviations of the critical frequency *fo*F2 from the reference

$$D(foF2) = \frac{foF2 - (foF2)_{ave}}{(foF2)_{ave}} \times 100\% . \quad (1)$$

The *D*(*fo*F2) variations are described in terms of percentage of the critical frequency *fo*F2 from the reference and, following Liu *et al*. (2008) and references therein, positive and negative storms occur when the absolute maximum value of *D*(*fo*F2) exceeds 20%. According to Burešová and Laštovička (2007), this limit is sufficiently large to prevent inclusion of random perturbations and disturbances of neutral atmospheric origin (gravity waves, etc.), thereby making the indicated positive and negative storms represent real changes in electron density not simply redistribution of the existing plasma.

The *D*(*fo*F2) data for the present study are derived from respective hourly values of *fo*F2 on Nov. 20-21, 2003, while the reference for each hour is the average value of *fo*F2 for that hour calculated from the four quiet days, Nov. 26-29, 2003. An important criterion used in choosing the reference period is these days must be devoid of not only of any significant geomagnetic activity but also there must be an absence of any considerable solar activity; this follows from the fact that, as Ben'kova *et al*. (1983) and Blagoveshchensky *et al*. (2006) and references therein have shown, the high solar flares activity results in ionospheric disturbances due to their effects on thermospheric neutral density (Sutton *et al*. 2006). Presently, Fig. 1a shows that the X-ray fluxes for the flares during Nov. 26-29, 2003 do not exceed 10^{-3} W/m^2 which indicates that the reference period is devoid of any significant solar activity, and Fig. 1b shows that with values of $Dst > -25$, the reference days are characterized by the absence of any significant ring current activity all (Tsurutani *et al*. 1995).

3. RESULTS

3.1 Heliophysical observations

Recently, Chukwuma (2009) in a related paper showed that the solar activity during November 18-21, 2003, was mainly due to active region 501, which

produced M1.8, M3.2, M3.9, and M4.5 flares on Nov. 18 and M1.7 flare on Nov. 19 and the M1.4, M9.6, and M5.8 flares of Nov. 20. Furthermore, using the results of Gopalswamy *et al*. (2005), Chukwuma (2009) concluded that the CMEs that caused the intense storm of Nov. 20-21, 2003 most probably resulted from the M3.9 flare. Consideration for flare M3.9 as source flare is further supported by the results of Ben'kova *et al*. (1983) which showed that the high flare activity of July 9, 1982, resulted in sudden ionospheric disturbance against a background of quiet geomagnetic activity but triggered an intense geomagnetic storm, two days after, on July 11, 1982. I will also suggest, basing on the results of Ben'kova *et al*. (1983), that the M1.4 and M9.6 flares that occurred, respectively, at 02:12 and 07:47 on Nov. 20 (Chukwuma 2009) are most likely to be associated with ionospheric storm on this day.

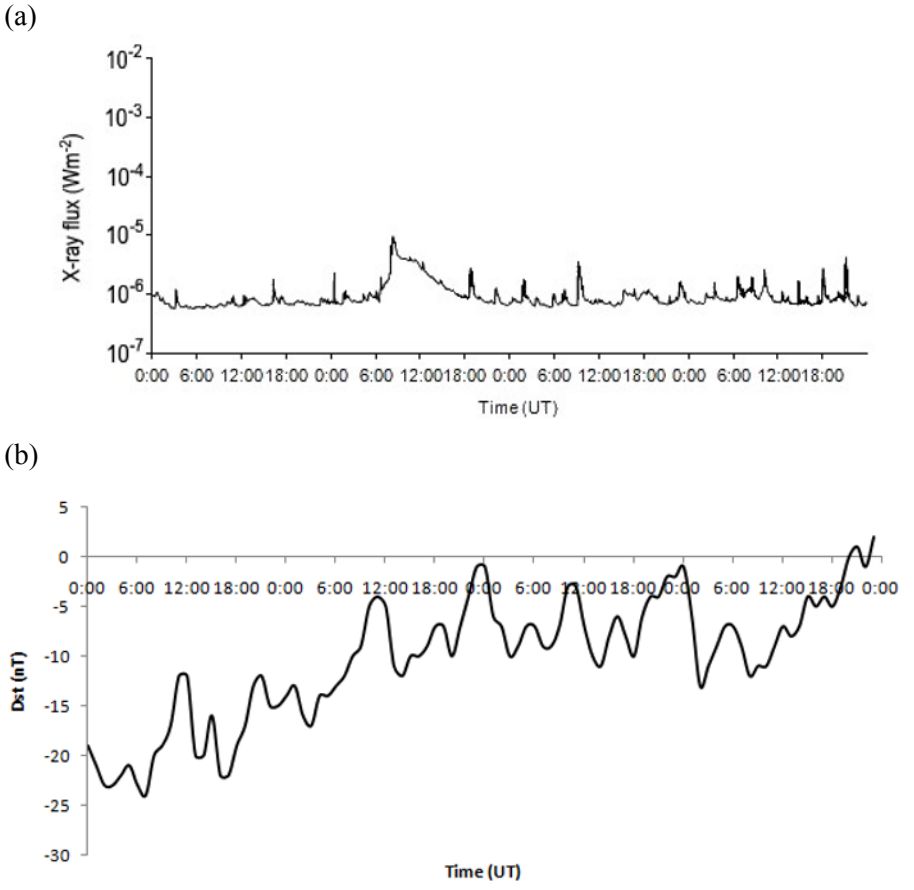


Fig. 1. The X-ray fluxes for the flares (a) and the *Dst* index variations (b) for the reference days: Nov. 26-29, 2003.

3.2 Interplanetary and geomagnetic observations

Figure 2a is the *Dst* plot for 00:00 UT, November 20, to 23:00 UT, Nov. 21, and shows *Dst*, on Nov. 20, decreasing rather steadily from its value of –4 nT at 00:00 to weak storm values of –32 and –34 nT at 06:00 and 07:00 UT, respectively. *Dst* recovered to –17 nT at 08:00 UT but shortly decreased to –38 nT at 09:00 UT before depressing sharply to –68 nT at 10:00 UT, indicating the existence of a moderate storm at this hour. *Dst* recovered to a value of –49 nT at 12:00 UT but thereafter depressed very strongly to –396 nT at 18:00 UT and got to the minimum value of –422 nT at 21:00 UT on Nov. 20.

Given that this paper involves an investigation into the ionospheric phenomena at pre-storm, it is necessary to clearly identify the storm commencement in order to delineate the early phase from the preceding pre-storm phase. Note that the present storm is characterized by the absence of sudden storm commencement (SSC) in which case one would be presented with the option of using the main phase onset as a defining parameter. However, as we already noted, the reliability of using the SSC as a storm onset has been argued for a long time and as a result some investigators choose the main phase onset (MPO) of the storm instead of SSC as the start time of a storm (Liu *et al*. 2008). The use of SSC as a reference time constitutes a bad choice (Prolss 1995). Also the use of the MPO for fixing the beginning of magnetic and ionospheric storms is fraught with problems that render a determination of the exact onset time difficult (Prolss 1995). It was observed that on November 20 proton density N_{sw} increased abruptly to 17.6 cm^{-3} at 09:00 UT (Chukwuma 2009). The large increase in the proton number density at this hour signals the arrival of a shock in the interplanetary medium (Nielsen and Honary 2000, Strickland *et al*. 2001). Coincident with the large increase in the proton number density at about 09:00 UT is the abrupt rise in temperature; T_p rose to peak value of 534,783 K (Fig. 2d). The temperature increase at this hour appears to confirm the arrival of a shock in the interplanetary medium. According to Hargreaves (1995) and Kelley (1989), the arrival of a shock in the interplanetary medium is indicated, in part, by the rapid increase in plasma density and plasma temperature. It is also noteworthy that Chukwuma (2009) has shown that at about 09:00 UT on Nov. 20, there was an abrupt increase in V_{sw} with the coming of a fast solar wind stream with $V_{sw} \approx 700$ km/s, which was followed by a decrease of V_{sw} to 687 km/s at 10:00 UT − an event which, according to Hargreaves (1995) and Kelley (1989), appears to confirm the presence of a shock in the interplanetary medium at ~09:00 UT.

Furthermore, according to Kamide *et al*. (1998), the majority (67%) of intense storms undergo two-step development during the main phase. That is, after the first $Dst_{\min}$ is reached, a brief partial recovery intervenes, to be

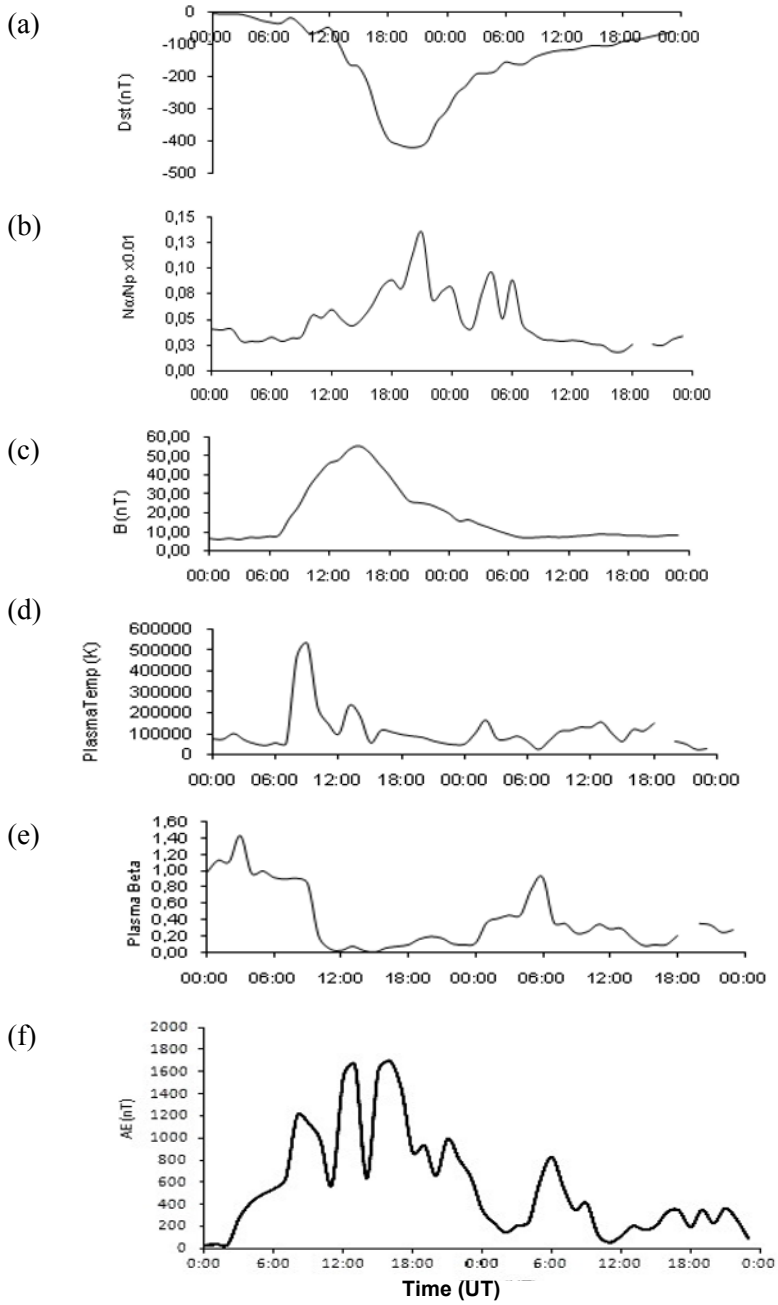


Fig. 2. Composition of interplanetary and geomagnetic observations for Nov. 20-21, 2003: (a) low-latitude magnetic index *Dst*, (b) alpha/proton density ratio $N\alpha/Np$, (c) total magnetic field *B*, (d) plasma temperature *Tp* , (e) plasma beta β, (f) *AE* index.

followed by another Dst_{min}. Presently, the storm of Nov. 20-21, 2003, is an intense geomagnetic storm but its *Dst* profile does not appear to situate it in the class of two-step storms with regards to storm classification (e.g., Farrugia *et al*. 2006, Gonzalez *et al*. 1994, Kamide *et al*. 1998, Vieira *et al*. 2001): the *Dst* profile appears to suggest the existence of a weak storm with Dst_{min} at 07:00 UT, a moderate storm with Dst_{min} at 10:00 UT, and a great storm with Dst_{min} at 21:00 UT, which imply a triple step storm. However, an application of the Kamide *et al*. (1998) criteria for two-step storms on the present *Dst* data shows that the intense storm of Nov. 20-21, 2003, is a two-step storm consisting of a moderate storm with $Dst_{min} = -68$ nT at 10:00 UT, and a major storm with $Dst_{min} = -422$ nT at 20:00 UT. I want to note that given the fact that the present *Dst* profile (Fig. 2a) does not present a precise main phase onset, the time of shock arrival, i.e., 09:00 UT, which is clearly identified, is hereby used as the reference time for the onset of the intense geomagnetic activity.

According to Gonzalez *et al*. (2001, 2002), two interplanetary structures are important for the development of intense storms: the sheath region just behind the forward shock, and the coronal mass ejecta (driver gas) itself. In the recent study of some heliophysical and geophysical phenomena during October-November 2003, Chukwuma (2009) showed that in regards to the present storm, the sheath was the most geoeffective element and occupied the interval 09:00-15:00 UT while the ejecta contained a magnetic cloud in the period 15:00-23:00 UT. This result is due to the occurrence of the average values: $\beta = 0.10$ (Fig. 2e), $N\alpha/Np = 0.08$ (Fig. 2b), $Tp = 83{,}389$ K (Fig. 2d), and $B = 36$ nT (Fig. 2c), which meet the criteria of magnetic cloud in the aforementioned interval. According to Gonzalez *et al*. (1994, 2001), there are numerous mechanisms that lead to the intensification of southward *Bz* in the high-speed stream sheath region which, like in the case of the present storm, resulted in the sheath being the most geoeffective structure in the development the very intense storm; for example, if there is a preexisting southward *Bz* upstream of the shock, shock compression will intensify this component. As these fields convect toward the driver gas, the draping effect will intensify the fields further. Chukwuma (2009) has shown that for the Nov. 20-21, 2003, storm there was a pre-existing southward IMF upfront of the shock which was intensified by compression by the shock, whereby southward *Bz* was increasing steadily from a value of –1.1 nT at 02:00 UT to –7.1 nT at 09:00 UT. It is probable that during this interval the normal shielding effect of the region 2 current sheet broke down. But according to Forbes (1989), the breaking down of the region 2 current sheet and the penetration to middle and low latitudes of electric fields generated by magnetosphere-ionosphere coupling processes is generally accepted to take place in

the early phases of a magnetic storm. So the existence of ionospheric storm prior to the early phase of a magnetic storm proposes the existence among other drivers of prompt penetration electric field.

Having identified the shock, I want to further note that the development of a storm can usually be divided into the initial phase and the main phase which are characteristically different in form. The initial phase, according to Lei *et al*. (2008), is mainly controlled by the shock and is typified by *Dst* showing a positive change responding to a ram pressure increase in the solar wind (Kamide *et al*. 1998), and in this regards Chukwuma (2009) showed that solar wind dynamic pressure P_{sw} for Nov. 20-21, 2003 increased abruptly from 3.05 nPa at 07:00 to 16.29 nPa at 09:00 UT, while the main phase is characterized by significant depression in *Dst* resulting from the enhancement of trapped particle population in the magnetosphere. However, for the present storm there is the absence of the positive change in *Dst* profile that characterizes the initial phase, which indicates the merging of the initial phase into a main phase that is not initially rapid and monotonic.

According to Gonzalez *et al.* (1994), the principal characteristic of the *Dst* representation of a geomagnetic storm is the main phase. It does transpire from this statement that the nature of the main phase of a geomagnetic storm provides a definitive explanation of the mechanisms of the ring current energization that are responsible for development of the storm. Observe that the main phase of the storm in the period 12:00-18:00 UT is characterized by a depression of the *Dst* index by −347 nT which indicates strong ring current intensification. It does appear from the results of Chukwuma (2009) that the large-scale interplanetary electric field driven by a period of prolong southward IMF is the major cause of the ring current buildup. It has also been shown that previous ring current particles can get energized to higher energies by new large-scale electric fields and can diffuse radially to lower *L* values, thus leading to a more energetic ring current (Gonzalez *et al*. 2001). Results of Chukwuma (2009) show that the interplanetary electric field *Ey* which is proportional to the solar wind energy input (Burton *et al*. 1975) indicates the existence of large electric field (>10 mV/m) and a large energy input between 16:00 and 17:00 UT.

It is pertinent to observe that the main phase is marked by a single extended rapid ring current injection in the intervals 12:00-18:00 UT, causing the rate of change of *Dst* to exceed −50 nT/h. Measurements of Wygant *et al.* (1998) on the CIRRES satellite during the March 24, 1991, geomagnetic storm had shown that large scale magnetospheric electric field penetrated into the inner magnetosphere between $L = 2$ and $L = 4$ when the rate of change of *Dst* was of the order of −50 nT/h. This would result in an instantaneous penetration of electric field from high latitude to the middle and equatorial

ionosphere. According to Huang (2008), the penetration electric field lasts for 8-10 hours without obvious attenuation during the main phase of a magnetic storm when the IMF remains southward. This implies that the shielding process of the region 2 field-aligned current does not work effectively during the storm main phase. Note that the low latitude magnetic index *Dst* variation (Fig. 2a) shows that the IMF remained southward in the period 12:00-18:00 UT. This result is supported by the results of Chukwuma (2009) which show that at about 12:00 UT, *Bz* had rotated sharply southward with a value of –24.0 nT indicating a change in *Bz* of $\delta Bz = 29.4$ nT in an hour. Thereafter, *Bz* undergoes a slight variation before increasing strongly from a value of –25.6 nT at 13:00 UT to a peak value –50.9 nT at 15:00 UT, which indicates a further change in *Bz* of $\delta Bz \approx 12.65$ nT in an hour. These *Bz* values are indicative of strong penetration electric fields which have profound effects on the redistribution of the global ionospheric plasma that its field uplifts the evening equatorial ionosphere (Huang *et al*. 2008). Observe also that the *B* plot in the period 09:00-20:00 UT (Fig. 2c) on Nov. 20 presented evidence of the existence of very strong interplanetary magnetic fields in the geospace.

Figure 2f is the auroral electrojet index, *AE,* plot for 00:00 UT, Nov. 20, to 23:00 UT, Nov. 21. The *AE* index serves as proxy for Joule heating in the auroral region. Figure 2f shows moderate auroral activity characterized by *AE* = 400-600 nT during 04:00-07:00 UT. Mikhailov and Perrone (2009) had related the pre-storm *Nm*F2 enhancement to such auroral activity. A sudden enhancement of *AE* exceeding 1000 nT was recorded at 08:00 UT. Observe from Fig. 2a that with the value of *Dst* > –25 at 08:00 UT, the sudden enhancement of *AE* occurred at a time associated with the absence of any significant ring current activity (Tsurutani *et al*. 1995). The *AE* plot also shows that the main phase is characterized by rapid bursts in energy injection rate with *AE* index recording 1500 nT and 1600 nT, respectively, during the intervals 12:00-13:00 and 15:00-16:00 UT. It is known that during magnetic storms, ionospheric disturbance dynamo electric fields, which are westward during the day and eastward at night, are driven by enhanced energy deposition into high latitude ionosphere (Blanc and Richmond 1980). The Blanc and Richmond theory predicts that the disturbance dynamo is slow to develop because of the gradual build up of the zonal winds and this results in time delays; relatively fast (occurring about 2-3 hours after major increases in convection) disturbance dynamo fields (e.g., Scherliess and Fejer 1997) are most likely due to dynamo action of fast travelling equatorward wind surges (e.g., Fuller-Rowell *et al.* 2002), while slower disturbances (occurring 3-12 hours later) are believed to be driven mostly by the electrodynamic action of storm-enhanced high latitude equatorward winds (Blanc and Richmond 1980). Note that the energy injections during the main phase are coincident

with the single extended rapid ring current injection in the intervals 12:00-18:00 UT, during which the rate of change of *Dst* exceeded −50 nT/h. Also, it has been noted that the *Bz* values during this period are indicative of strong penetration electric fields.

3.3 Ionospheric response

Figure 3 presents the *D*(*fo*F2) variations at the East Asian stations. The ionospheric response for Manzhouli indicates that the interval 00:00-12:00 UT on Nov. 20 was rather quiet; the interval 14:00-15:00 UT shows that *fo*F2 was enhanced by ~46% above the reference level. This positive storm was immediately followed by an intense negative storm resulting from the abrupt depletion of *fo*F2 at 16:00 UT. The *D*(*fo*F2) *versus* UT plot for Chongqing also shows that the ionosphere at this station was rather quiet in the interval 00:00-12:00 UT on Nov. 20. Thereafter, the ionosphere recorded a positive storm between 13:00 and 15:00 UT with 45% enhancement in peak electron density at 14:00 UT. At Guangzhou, the main ionospheric feature began at ~12:00 UT when *fo*F2 became abruptly enhanced, resulting in an intense storm with 43% increase in *fo*F2. The ionosphere also recorded a positive storm between 16:00 and 19:00 UT. At ~20:00 UT, the *fo*F2 suffered 66% depletion resulting in a very intense negative storm that lasted for two hours. The *D*(*fo*F2) variations at Hainan reveal the ionosphere recording of positive storms at 07:00 and 13:00 UT, a moderate storm at 18:00-19:00 UT, with a 40% peak electron density enhancement at ~19:00 UT followed by a strong and abrupt depletion in *fo*F2 that resulted in 75% in reduction in peak electron density at 21:00 UT.

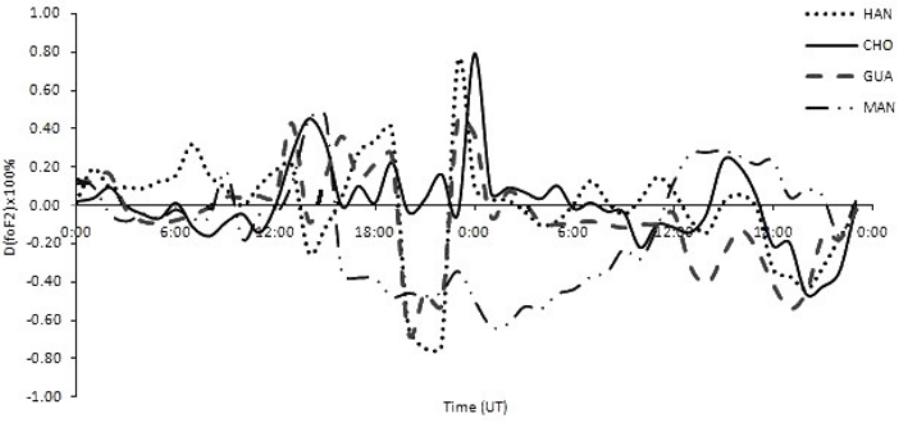


Fig. 3. Variations in *D*(*fo*F2) for East Asian sector for November 20-21, 2003.

Figure 4 presents the *D*(*fo*F2) variations at the Australian stations. The ionosphere at Darwin was quiet in the period 00:00-10:00 UT; at ~12:00 UT *fo*F2 reduced sharply to ~30% from the reference level. From ~14:00 UT *fo*F2 depressed strongly, resulting in an intense negative storm ~70% peak electron density depletion at ~17:00-18:00 UT on Nov. 20. The ionosphere at Learmonth, with the exception of a positive ionospheric storm at ~08:00 UT, was largely quiet in the period 00:00-14:00 UT on Nov. 20. However, at ~16:00 UT there commenced a sharp depletion in *fo*F2 which led to ~72% reduction in peak electron density at 19:00 UT. The ionospheric variations at Mundaring show a positive storm at ~08:00 UT followed by an intense negative storm during which the ionosphere recorded maximum depletion of 63% in peak electron density at ~18:00 UT on November 20.

Figure 5 shows the ionospheric response to the interplanetary forcing in the European–African zone. At Juliusruh–Rugen, a positive storm occurred at ~12:00 UT on Nov. 20 whereas the interval 01:00-16:00 UT was mostly defined by the existence of negative storm. Note that from ~19:00 UT, there came into existence a very intense positive storm whose peak electron density was enhanced to a value of 93% at 22:00 UT. The ionosphere above Rome was relatively quiet in the hours 00:00-11:00 UT on Nov. 20; at ~12:00 UT, the ionosphere recorded a 29% increase in *fo*F2 which resulted in a positive storm that lasted until 15:00 UT. This positive stormed was immediately followed a sharp depletion of *fo*F2 leading to an intense negative storm at 17:00 UT with *D*(*fo*F2) recording a value of –30%. The ionosphere at Athens exhibited quiet conditions until ~12:00 UT when a positive storm commenced and lasted until 15:00 UT. The positive stormed was also directly followed a sharp depletion of *fo*F2 leading to a negative storm at 18:00 UT.

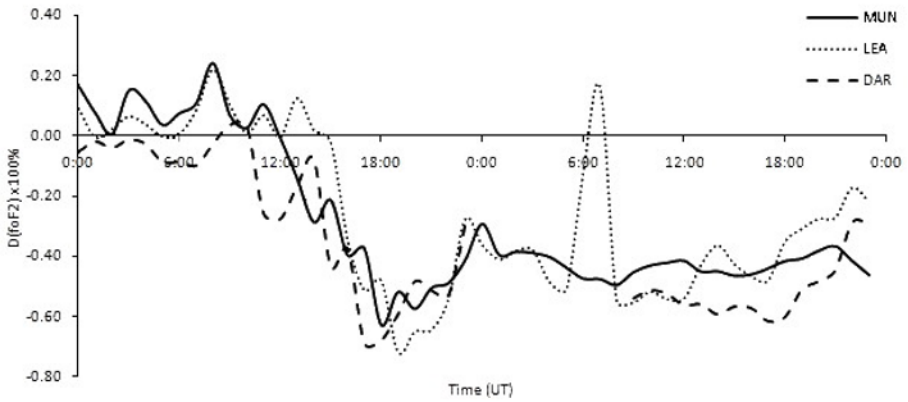


Fig. 4. Variations in *D*(*fo*F2) for Australian sector for November 20-21, 2003.

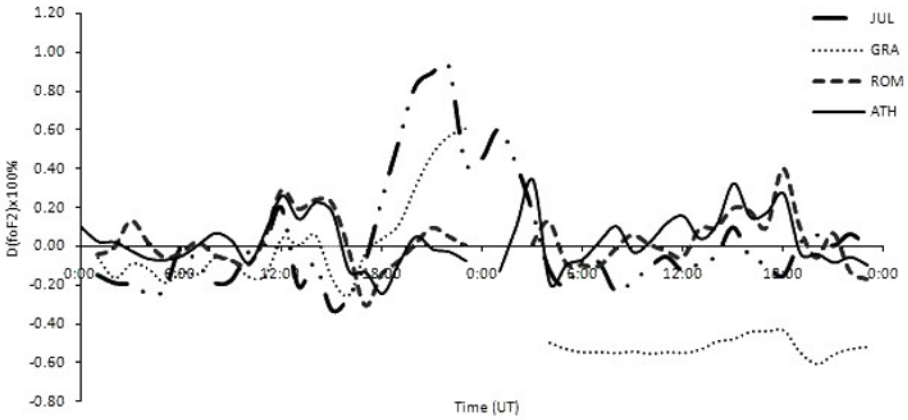


Fig. 5. Variations in $D(fo\mathrm{F2})$ for European–African sector for November 20-21, 2003.

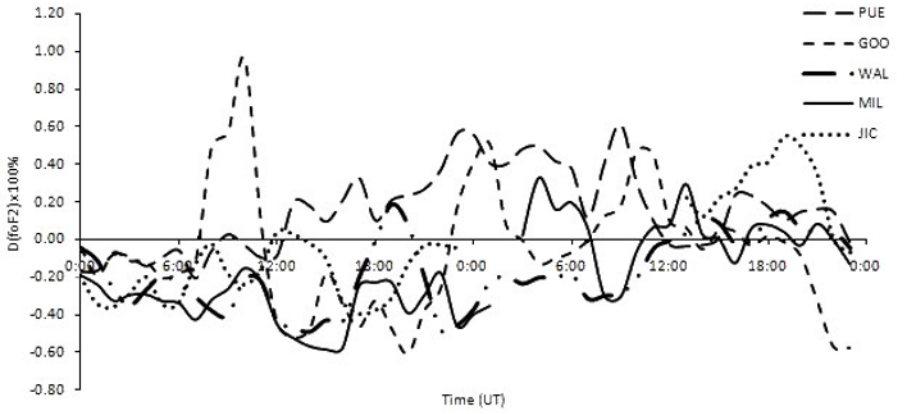


Fig. 6. Variations in $D(fo\mathrm{F2})$ for American sector for November 20-21, 2003.

The ionosphere at Grahamstown was quiet in the hours 00:00-18:00 UT, Nov. 20. At ~19:00 UT, foF2 began to experience increased enhancement in peak electron density.

Figure 6 presents the ionospheric conditions for the stations of interest in the American zone. The ionosphere at Goosebay shows that between 08:00 and 10:00 UT the peak electron density was strongly enhanced with the ionosphere recording ~50% ≤ foF2 ≤ 96%. However, between 13:00 and 20:00 UT, foF2 suffered strong depletion that is indicative of intense negative storm with values mostly in the range –33% ≤ foF2 ≤ –61%. The ionosphere at Milestone Hill was largely defined by the existence of intense negative with peak deple-

tion in *fo*F2 of ~ –60% occurring at 16:00 UT on Nov. 20. The ionospheric *fo*F2 variations at Wallops Island show that the ionosphere appears largely characterized by the existence of intense negative storm during the period under study. The ionosphere at Puerto Rico with the exception of a negative storm at 07:00 UT was largely quiet in the period 00:00-12:00 UT on Nov. 20. Nevertheless, beginning from ~13:00 UT the ionosphere started to experience what appears as modulated progressive enhancement in peak electron density. The ionosphere at Jicamarca on Nov. 20 presents the existence of intense negative storm in the period 02:00-19:00 UT, Nov. 20, with peak depletion occurring at 17:00 UT with *fo*F2 reduced to 40% below the reference level.

4. DISCUSSION

An analysis of the interplanetary and geomagnetic observations shows that the pre-storm phase occurred between 00:00 and 09:00 UT, and the main phase between 09:00 and 21:00 UT, as well as the merging of the initial phase of the present storm into its main phase. The interest of this paper mainly lies in revealing the mechanisms responsible for the pre-storm phenomena as well as any possible relationships between such mechanisms and those responsible for main phase ionospheric phenomena, especially the role of penetration electric fields. According to Huang (2008), strong penetration electric fields during intense storms have profound effects on redistribution of global ionospheric plasma.

It was observed in Section 3.2 that the main phase is marked by a single prolong rapid ring current injection in the intervals 12:00-18:00 UT on November 20, causing the rate of change of *Dst* to exceed –50 nT/h and this would result in an instantaneous penetration of electric field from high latitude to the middle and equatorial latitudes. According to Huang (2008), the penetration electric field lasts for 8-10 hours without obvious attenuation during the main phase of a magnetic storm when the IMF remains southward. This implies that the shielding process of the region 2 field-aligned current does not work effectively during the storm main phase. Note that Fig. 2a, as well as the results from IMF*Bz* variations for the present storm of Chukwuma (2009) show that the IMF remained southward in the period 12:00-18:00 UT during the main phase of the present storm. Furthermore, Chukwuma (2009) has shown that *Bz* values are indicative of strong penetration electric fields which have profound effects on the redistribution of the global ionospheric plasma that its field uplifts the evening equatorial ionosphere (Huang 2008).

It is known that prompt penetration electric field is eastward during the daytime to the dusk sector and westward in the midnight to the dawn sector (Jaggi and Wolf 1973, Spiro *et al*. 1988, Fejer *et al*. 1990). For example,

during the Bastille Day storm, an eastward penetration electric field between 20:00 and 21:00 UT on 15 July 2000 associated with a sharp decrease in the *Dst* index lifted the ionosphere over the South Atlantic and Brazilian sectors where dusk conditions prevailed (Basu S. *et al*. 2001). At the same time, in the post midnight sector, the penetration of a westward electric field caused the ionospheric F region over India to drift downward and collapse through recombination (Sastri *et al*. 2002). It is pertinent to note that the main phase was characterized in the intervals 12:00-13:00 UT and 15:00-16:00 UT by strong *AE* enhancements which will probably set-off relatively fast ionospheric disturbance dynamo electric fields (e.g., Scherliess and Fejer 1997) during the main phase. Thus, there might be two competing mechanisms that control the ionospheric F layer.

Presently, in both the East Asian and Australian sectors, the dusk conditions prevailed in the period 12:00-15:00 UT in which case the ionosphere should be lifted and enhanced if the controlling mechanism is the prompt penetration electric field. Available results indicate that there is some level of simultaneity in ionospheric response in regards to the existence of positive storm in the interval 13:00-14:00 UT at Manzhouli, Guangzhou, Chongqing and Hainan; however, the positive storm does not appear to reflect a latitudinal dependence. In the midnight/postmidnight sector of 16:00-18:00 UT, the penetration of a westward electric field would cause the ionospheric F region over this longitudinal zone to drift downward and collapse through recombination: the *D*(*fo*F2) variations at Manzhouli recorded an abrupt depletion of *fo*F2 at 16:00 UT and an intense negative storm at 18:00 UT. At Darwin, the ionosphere recorded an intense negative storm at ~17:00-18:00 UT. The ionosphere at Learmonth shows that at ~16:00 UT there commenced a sharp depletion in *fo*F2 which led to intense reduction in peak electron density at 19:00 UT. The ionospheric variations at Mundaring also recorded an intense negative storm at ~18:00 UT.

In the European–African zone, the daytime conditions prevailed in the interval 12:00-18:00 UT on Nov. 20 in which case the ionosphere is expected to be lifted by prompt penetration electric field and enhanced; note that at Juliusruh–Rugen the dominant response feature to penetration electric field appears to be the very intense positive storm that started from ~18:00 UT. The ionosphere above Rome and Athens, respectively, recorded a positive storm that lasted from 12:00 until 15:00 UT. Also, the 12:00-18:00 UT interval in the American zone coincides with daytime and it is expected that the prevailing prompt penetration electric field will be eastward, and it is observed that the ionosphere at Puerto Rico started to experience from ~13:00 UT what appears as modulated progressive enhancement in peak electron density with *fo*F2 recording 33% enhancement at 17:00 UT.

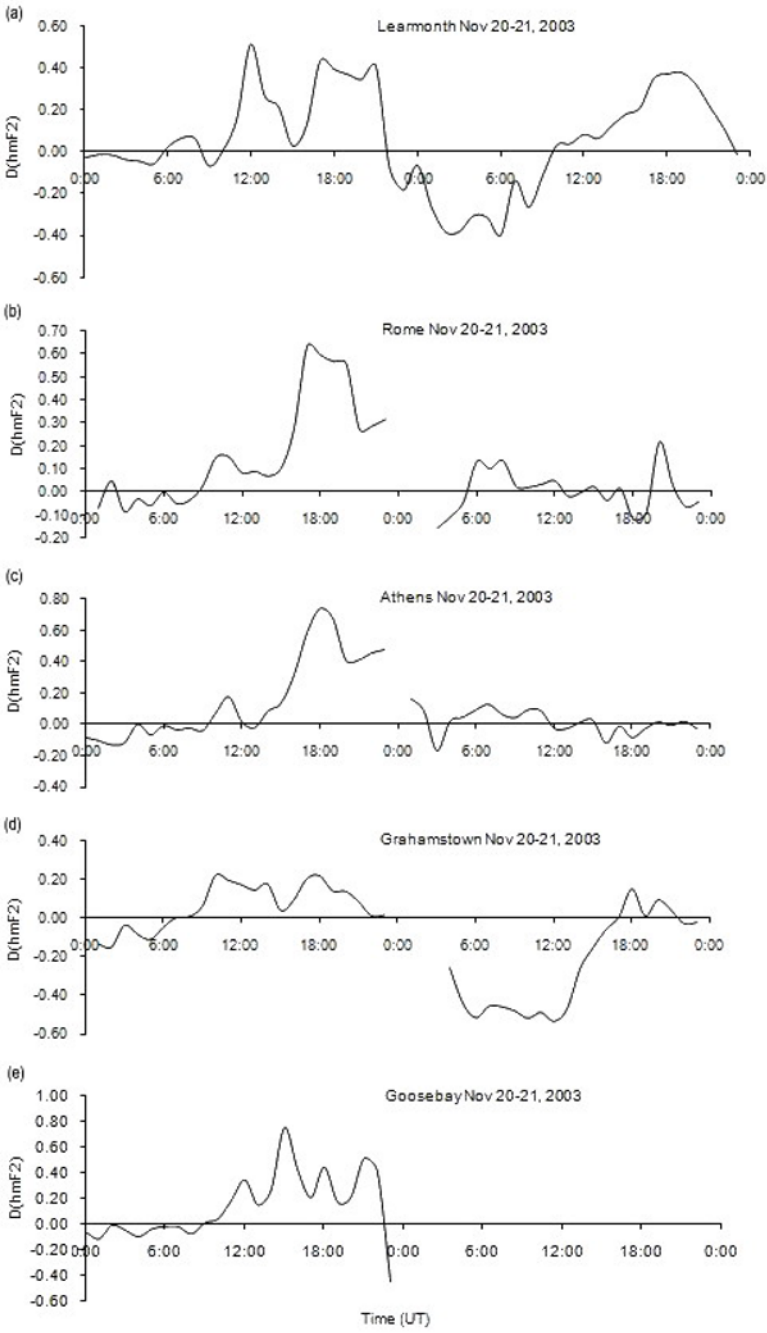


Fig. 7. Variations in $D(hmF2)$ for selected stations with available data for November 20-21, 2003.

The above results notwithstanding, there is a depletion in peak electron density at Hainan at 14:00 UT; at Darwin, *fo*F2 reduced sharply at ~12:00 UT with a depletion in peak electron density occurring at 15:00 UT. Also the ionosphere at Mundaring recorded depletion in *fo*F2 at 15:00 UT. Furthermore, in the midnight/post midnight sector, a strong positive storm was recorded at Guangzhou and Hainan. It is noteworthy that the interval 13:00-15:00 UT at Juliusruh–Rugen was mostly defined by the existence of a negative storm while the ionosphere at Rome also recorded an intense negative storm at 17:00 UT. Furthermore, at Athens the ionosphere recorded a sharp depletion of *fo*F2 leading to a negative storm at 18:00 UT, while at Grahamstown the ionosphere recorded a negative storm at 18:00 UT. At Goosebay, in the hours between 13:00 and 20:0 UT, *fo*F2 suffered strong depletion that is indicative of intense negative storm. The ionosphere at Milestone Hill during this period was largely defined by the existence of intense negative storm with peak depletion in *fo*F2 occurring at 16:00 UT. The ionospheric *fo*F2 variations at Wallops Island are also characterized by the existence of intense negative storm during the period 12:00-18:00 UT. Peak depletion of ionospheric *fo*F2 at this station occurred at 14:00. The ionosphere at Jicamarca on Nov. 20 was quiet in the interval 12:00-15:00 UT and later recorded a negative storm at 17:00 UT.

These ionospheric responses that are reported in the last paragraph appear to suggest the existence of ionospheric disturbance dynamo electric fields at these stations at the various times and in that regards do not strongly show the prompt penetration electric fields as the solitary main phase ionospheric driver. However, it is important to point out that the depletion of *fo*F2 in the European–African longitudinal zone cannot be accounted for by the disturbance dynamo electric field. Available data (Fig. 7) show that the ionosphere at these stations was lifted as follows: Rome (63%), Athen (73%), and Grahamstown (9%). These results appear to suggest that the *fo*F2 depletion at these stations may not be mainly due to changes in neutral composition resulting from neutral wind produced predominantly in the region of Joule heating in the aurora zone. It is noteworthy that Mikhailov and Schlegel (1998) also suggested that the very large *fo*F2 depletion during a geomagnetic storm is not due to changes in neutral composition. The loss of ionization at F2-region heights may be primarily due to charge transfer reactions of the type (Prolss 1995)

$$O^+ + N_2 \xrightarrow{k_1} NO^+ + N, \tag{2}$$

$$O^+ + O_2 \xrightarrow{k_2} O_2^+ + O, \tag{3}$$

where k_1 and k_2 are associated reaction rate constants. The resulting molecular ions NO^+ and O^+ are quickly destroyed by dissociative recombination. Therefore, the loss rate reduces to the simple form

$$I(N) = \beta[O^+], \tag{4}$$

where β depends on the density of the molecular gases N_2 and O_2,

$$\beta = k_1[N_2] + k_2[O_2]. \tag{5}$$

According to Prolss (1995) and references therein, during very intense activity soft particle precipitation will increase the vibrational excitation of molecular nitrogen. This will in turn increase the rate of coefficient k_1 of eqs. (2) and (5) and therefore the loss of ionization. Also, according to Mikhailov and Schlegel (1998) strong electric field during the storm of April 3, 1992, resulted in very large increases in k_1, by more than a factor of 10, and, as a consequence, this resulted in the linear loss coefficient β in eq. (5) increasing by a factor of 5 at F2-layer heights.

It is pertinent to note that in the Australian sector, the negative storm did not exhibit equatorward shift whereby it begins in the higher latitudes and then spreads to the lower latitudes; present results show the negative storm beginning earlier at the lower latitude station of Darwin. Furthermore, the negative storm in this sector does not appear to indicate any latitudinal dependence and the amplitude of the negative phase (maximum absolute value of $D(foF2)$) decreased poleward at ~17:00 UT on November 20. Furthermore, observe that the ionospheric response in European–African longitudinal zone does not present the important features of the negative storm which according to Danilov (2001) are:

– Its equatorward shift during a geomagnetic storm from the auroral latitudes to middle latitudes;

– The amplitude of the negative phase (maximum absolute value of $D(foF2)$) decreases during the equatorward shift.

These results do not also recommend ionospheric disturbance dynamo electric fields as the dominant driver of ionospheric responses in the main phase.

Having examined the ionospheric response in the main phase with regards to penetration electric fields and disturbance dynamo fields, the pre-storm phase would be discussed for the likely drivers of the ionospheric phenomena. Presently, the pre-storm ionospheric phenomena under investigation only occurred at Hainan at 07:00 UT, and Learmonth, Mundaring and Goosebay at ~08:00 UT on Nov. 20 with respective moderate to intense enhancement in peak electron density and a strong depletion of $foF2$ between 01:00 and 06:00 UT at Jicamarca on the same day. These phenomena

represent a 20% occurrence in available data. Observe that the pre-storm enhancement does not present any longitudinal dependence. It is pertinent to also note that a preliminary investigation of the longitudinal dependence by Burešová and Laštovička (2007) using European and Boulder data was inconclusive. The lack of longitudinal dependence of the pre-storm phenomena could suggest that their origin does not derive from local time effect whose explanation is based on the assumptions that positive storms are caused by meridional winds and negative storms by changes in the neutral gas composition (Prolss 1995). However, the fact that the enhancements occurred in simultaneity at two widely separated longitudinal zones may suggest a role played by the magnetospheric electric field imposed on the high latitude ionosphere penetrating almost instantaneously to middle and low latitudes on both dayside and the night-side (Basu S. *et al*. 2005, and references therein), as well as a role by the disturbance dynamo electric field- prompt penetration electric fields and disturbance dynamo electric fields have opposite polarity local time dependences. The role of magnetospheric electric field appears plausible given the analysis of the geomagnetic phenomenon which shows the arrival of the shock at 09:00 UT, and, according to Gonzalez *et al*. (1994, 2001), if there is a preexisting southward *Bz* upstream of the shock, the high-speed stream shock compression will intensify this component. In this regards, it may be important to note that results of Chukwuma (2009) show that *Bz* was very weak but positive at 1.1 nT at 0:100 UT on Nov. 20. At ~02:00 UT, *Bz* rotated southward, increasing steadily from a value of –1.1 to –5.2 nT at 08:00 UT, indicating a change in *Bz* of δBz = 6.3 nT. It is probable that during this interval the normal shielding effect of the region 2 current sheet broke down and electric fields generated by the magnetosphere-ionosphere coupling processes penetrated to ionospheric middle and low latitudes. Chukwuma (2007a) has shown in the recent study of the intense storms of March 13-15, 1989 (*Dst* = –600 nT) and April 1-2, 1973 (*Dst* = –211) that pre-storm phenomena occurred on October 20, 1989, as a result of the southward turning of *Bz* at ~21:00 UT on Oct. 18 which got to a change in *Bz* of δBz = 12.2 nT at 23:00 UT on Oct. 18 and on April 1, 1973, as a result the southward turning of *Bz* at ~15:00 UT which got to a change in *Bz* of δBz = 13.6 nT at 19:00 UT on March 31.

Under the above-mentioned scheme, a prompt penetrating electric field would cause dayside enhancement at Hainan, Learmonth and Mundaring and, on the other hand, a disturbance dynamo electric field would cause nightside enhancement at Goosebay. But Fig. 2f shows a sudden enhancement of *AE* of 1000 nT occurred at 08:00 UT. The implication of this result from Blanc and Richmond's (1980) theory is that relatively fast disturbance

dynamo fields will be delayed by about 2-3 hours and this appears to rule out disturbance dynamo electric field as the probable cause of nightside enhancement at Goosebay. Now the pre-storm phenomena, as observed, do not display a systematic latitudinal dependence; they are strong at both low latitude and higher middle latitude, and this aspect led Burešová and Laštovička (2007) to suggest that an explanation by means of the magnetospheric electric field rather unlikely. But the existence of pre-storm phenomena at Hainan, Learmonth, Mundaring, and Goosebay proposes the breaking down of region 2 current sheet and prompt penetration to middle latitudes of electric fields generated by magnetosphere–ionosphere coupling processes prior to the early phase of the present magnetic storm.

Now the 6% increase in *hm*F2 observed at Learmonth cannot totally account for enhancement in *fo*F2 at that station; *hm*F2 data were not available at Hainan and Mundaring. Also the *hm*F2 plot for Goosebay shows that the positive storm at 0800 UT at this station cannot be accounted for by the uplifting of the F2-region. The plot shows that the ionospheric plasma at this station was moved downward by 8% at this time. On the account of the *hm*F2 variations, the present results appear to suggest that both prompt penetrating electric fields and disturbance dynamo electric fields are not the main drivers of pre-storm phenomena. It is also instructive to note that both the analysis of Liu *et al*. (2008) and Burešová and Laštovička (2007) indicate that pre-storm enhancements are not accompanied by a corresponding change in *hm*F2. Blagoveshchensky *et al*. (2006) have also shown that the variations in *D*(*fo*F2) are in antiphase with *D*(*hm*F2). Now given the ionospheric phenomena during 12:00-18:00 UT on Nov. 20, where prompt penetrating electric field could not account for the observed ionospheric response despite the single extended rapid ring current injection in the period, it will be possible to strongly postulate that the prompt penetrating electric field is not a driver of pre-storm phenomena.

Table 1 lists the coordinates of the ionosonde stations and shows that the observed latitudinal extent of pre-storm phenomena is 15.45°N geomagnetic latitude. Danilov (2001) and Kane (2005) have, on the account of the cusp being the only formation that reacts to the coming geomagnetic disturbance before any geomagnetic index does, suggested the soft particle precipitation in the dayside cusp region as the possible driver of pre-storm enhancement, but in a recent paper, Burešová and Laštovička (2007) argued that the observed latitudinal extent of pre-storm phenomena is 40.0°N geomagnetic latitude and their occurrence in any time of the day practically excludes soft particle precipitation as an important mechanism of pre-storm phenomena. I would argue that the exclusion, basing on the longitudinal extent, of soft particle precipitation as an important driver of pre-storm enhancement is

inconclusive, because Maih (1989) has reported soft particle precipitation at F2 heights in the equatorial zone. Presently, Fig. 2f shows, basing on the results of Ahn *et al*. (1983), an enhanced global particle energy injection rate at 08:00 UT, which probably acted as a possible driver. Precipitating particles have also been suggested as a source of heating in the lower part of the thermosphere (Danilov 2001) which may lead to thermospheric composition changes, thereby acting as a contributing mechanism to the pre-storm phenomenon. Chukwuma (2007b) has recently hypothesised that particle precipitation as a mechanism may account for the simultaneous depletion of *fo*F2 during severe storms. Furthermore, Basu S. *et al*. (2007) in their study of the equatorial ionosphere at dusk relating to penetration electric fields during intense magnetic storms have suggested the lifting of the ionospheric plasma in the South Atlantic and Brazillian sector to be overemphasized, possibly because of enhanced conductivity gradient caused by the energetic particle precipitation.

It is known (Ben'kova *et al*. 1983, Blagoveshchensky *et al*. 2006 and references therein, Brodrick *et al*. 2005) that the effects of solar flares on Earth's ionosphere result in the appearance of sudden ionospheric disturbances (SID) which are observed in the dayside. Furthermore, in a recent study of some heliophysical and geophysical phenomena during the high solar activity period in October-November 2003, Chukwuma (2009) showed that very large X-class flares might not cause very intense geomagnetic storms, while flares of M importance would, which underscores the fact that the initiation of space weather events is a problem in the solar physics community which requires distinct efforts focusing on flares and CMEs separately. In this regards, a consideration of the flare activity is attempted to explain the pre-storm ionospheric phenomena during Nov. 20, 2003: In the presentation of the characteristics of solar X-ray flares during Nov. 18-21, 2003, Chukwuma (2009) pointed out the occurrence of M1.4, and M9.6 flares, respectively, at 02:12 and 07:47 UT on Nov. 20. I want observe that during the respective 7 and 2 hours between these flares and the arrival of the shock, the *Dst* index largely maintained values of $Dst > -25$ (Fig. 2a). These conditions provide an opportunity to analyze the direct effect of flares on changes in the ionosphere without the ambiguity of any geomagnetic effects. The times of occurrence of the flares, according to the characteristics of solar X-ray flares during Nov. 18-21, 2003 (Chukwuma 2009), on Nov. 20 coincided with daytime in both the East Asian and Australian sectors which appears to suggest that the flares could be responsible for the enhancements at mid latitude stations of Hainan, Learmonth, and Mundaring. This suggestion derives from the fact that the intensification of the X-ray emission results in an increase in the electron density in the ionospheric D, E and F region (Ben'kova *et al*. 1983, Brodrick *et al*. 2005). In the F region, an increase in

ionization density can reach 20-50% (Ben'kova *et al*. 1983, Blagoveshchensky *et al*. 2006) with the effect lasting 1.5 days against the background of quiet magnetic activity (Ben'kova *et al*. 1983). Blagoveshchensky *et al*. (2006) have also shown that the increase in F2 layer peak electron density *D*(*fo*F2) that was associated with the X17 flare of Oct. 28, 2003 was ~50% in the mid latitude stations in Europe. The results of Dymond *et al*. (2004) and references therein also show that solar flares result in about 20% increase in electron density near the F2 peak. However, the increased electron density decayed out with a time constant of approximately 1 hour at the F region heights (Dymond *et al*. 2004). The aforementioned increases in ionospheric peak ionization could be explained by the fact that thermospheric neutral density at low to mid-latitudes increases by about 50-60% in response to large solar flares with the time to peak response of the thermosphere neutral density being ~72±42 min with respect to the peak EUV flux emitted by flares (Sutton *et al*. 2006). According to Danilov (2001), composition changes directly influence the electron concentration in the F2 region though the actual peak electron density dependence on [O] and [N_2] could be more complicated than presently assumed; for example, some worker have proposed that electron concentration is proportional to $[O]_n$ (n = 0.7-0.85) and so simultaneous variation of both [O] and [N_2] with the [O]/[N_2] ratio remaining constant still leads to an increase of concentration (Danilov 2001).

However, the postulation that flare activity was likely to be responsible for the pre-storm enhancements of *fo*F2 observed in the East Asian and Australian sector becomes less probable on the following facts: the pre-storm enhancements of *fo*F2 occurred nearly simultaneously with the nighttime intense positive storm in the American station of Goosebay (nighttime events cannot be caused by flares (Burešová and Laštovička 2007), the European–African zone did not record any pre-storm ionospheric phenomena, and the occurrence of depletion of *fo*F2 at the low latitude station of Jicamarca. Furthermore, the pre-storm enhancement at Learmonth and Mundaring that occurred at 08:00 UT cannot be accounted by solar flares because the increase in electron density would probably have, according to Dymond *et al.* (2004) and references therein, decayed out. This leaves the enhancement at Hainan as a possible result of solar flares. It is therefore more plausible then to claim that solar flares are not the main drivers for the enhancements (Liu *et al*. 2008). According to Burešová and Laštovička (2007, 2008), solar flares can occasionally strengthen the pre-storm enhancements of F2-region peak electron density, but are not responsible for their occurrence, at least not as a dominant or very important source of such events.

5. CONCLUSIONS

The present investigation shows that:

- The ionospheric responses in the main phase do not indicate prompt penetration electric fields as the main ionospheric driver. Also ionospheric disturbance dynamo electric fields are not the dominant driver of ionospheric responses in this phase.
- None of the following drivers: magnetospheric electric fields, disturbance dynamo electric fields, and solar flares are the main drivers of the pre-storm phenomena.

These results appear to suggest that in the main, the pre-storm phenomena could be the result of some underlying mechanisms that are working together with varying degree of importance (Lei *et al.* 2008, Zhao *et al.* 2008). Furthermore, the present results point to the fact that the real nature of the pre-storm phenomena's characteristics and controlling mechanism still presents an unresolved problem in ionospheric physics. However, it should be noted that the very existence of pre-storm ionospheric phenomena appear to strongly indicate that preceding the arrival of a shock at the magnetosphere are some coupling mechanism between interplanetary features and the thermosphere-ionosphere system by which the solar wind continual modifies the dynamics, electrodynamics and chemistry of the Earth's ionosphere on a global level.

References

Aarons, J., and R.S. Allen (1971), Scintillation boundary during quiet disturbed magnetic conditions, *J. Geopys. Res*. **76**, 1, 170-177, DOI: 10.1029/JA076i001p00170.

Aarons, J., R.S. Allen, and H.E. Whitney (1972), Observations of scintillations of two satellite beacons near the boundary of the irregularity region, *Planet. Space Sci*. **20**, 7, 965-972, DOI: 10.1016/0032-0633(72)90209-7.

Ahn, B.-H., S.-I. Akasofu, and Y. Kamide (1983), The Joule heat production rate and the particle energy injection rate as a function of the geomagnetic indices AE and AL, *J. Geophys. Res*. **88**, 6275-6287, DOI: 10.1029/JA088iA08p06275.

Akasofu, S.-I. (1970), Diagnostics of the magnetosphere using geomagnetic, auroral and airglow phenomena, *Ann. Geophys*. **26**, 443-457.

Basu, S., Su. Basu, K.M. Groves, H.-C. Yeh, S.-Y. Su, F.J. Rich, P.J. Sultan, and M.J. Keskinen (2001), Response of the equatorial ionosphere in the South Atlantic region to the great magnetic storm of July 15, 2000, *Geophys. Res. Lett*. **28**, 18, 3577-3580, DOI: 10.1029/2001GL013259.

Basu, S., Su. Basu, K.M. Groves, E. Mackenzie, M.J. Keskinen, and F.J. Rich (2005), Near-simultaneous plasma structuring in the midlatitude and equatorial ionosphere during magnetic superstorms, *Geophys. Res. Lett*. **32**, L12S05, DOI: 10.1029/2004GL021678.

Basu, S., Su. Basu, F.J. Rich, K.M. Groves, E. Mackenzie, C. Coker, Y. Sahai, P.R. Fagundes, and F. Becker-Guedes (2007), Response of the equatorial ionosphere at dusk to penetration electric fields during intense magnetic storms, *J. Geophys. Res*. **112**, A08308, DOI: 10.1029/2006JA012192.

Ben'kova, N.P., K.N. Vasilev, V.N. Ishkov, N.S. Kaminer, E.F. Kozlov, N.M. Samorokin, L.A. Kolesnikova, E.V. Lavrova, E.V. Nepomniashchaia, A.M. Novikov, N.K. Pereyaslova, and V.V. Formichev (1983), Helio- and geophysical phenomena in July 1982, *Geomagn. Aeron.* **23**, 5, 579-582.

Blagoveshchensky, D.V., J.W. MacDugall, and A.V. Pyatkova (2006), The ionosphere of Europe and North America before the magnetic storm of October 28, 2003, *Geomagn. Aeron.* **46**, 3, 350-360, DOI: 10.1134/S001679320603 0091.

Blanc, M., and A.D. Richmond (1980), The ionospheric disturbance dynamo, *J. Geophys. Res*. **85**, A4, 1669-1686, DOI: 10.1029/JA085iA04p01669.

Brodrick, D., S. Tingay, and M. Wieringa (2005), X-ray magnitude of the 4 November 2003 solar flare inferred from the ionospheric attenuation of the galactic radio background, *J. Geophys. Res*. **110**, **A**09S36, DOI: 10.1029/2004JA 010960.

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.

Burton, R.K., R.L. McPherron, and C.T. Russell (1975), An empirical relationship between interplanetary conditions and Dst, *J. Geophys. Res*. **80**, 31, 4204-4214, DOI: 10.1029/JA080i031p04204.

Chukwuma, V.U. (2007a), On positive and negative ionospheric storms, *Acta Geod. Geophys. Hung*. **42**, 1, 1-21, DOI: 10.1556/AGeod.42.2007.1.1.

Chukwuma, V.U. (2007b), On the geomagnetic and ionospheric responses to an intense storm associated with weak IMF *Bz* and high solar wind dynamic pressure, *Acta Geophys.* **55**, 4, 469-489, DOI: 10.2478/s11600-007-0030-6.

Chukwuma, V.U. (2009), On heliophysical and geophysical phenomena during October-November 2003, *Acta Geophys.* **57**, 3, 778-800, DOI: 10.2478/s11 600-009-0003-z.

Danilov, A.D. (2001), F2 region response to geomagnetic disturbance, *J. Atmos. Sol.-Terr. Phys*. **63**, 5, 441-449, DOI: 10.1016/S1364-6826(00)00175-9.

Davis, C.J., M.N. Wild, M. Lockwood, and Y.K. Tulunay (1997), Ionospheric and geomagnetic responses to changes in IMF B_z: a superposed epoch study, *Ann. Geophys*. **15**, 2, 217-230, DOI: 10.1007/s00585-997-0217-9.

Dymond, K.F., S.A. Budzien, A.C. Nicholas, S.E. Thonnard, R.P. McCoy, R.J. Thomas, J.D. Huba, and G. Joyce (2004), Ionospheric response to the solar flare of 14 July 2000, *Radio Sci*. **39**, RS1S25, DOI: 10.1029/2002RS 002842.

Farrugia, C.J., V.K. Jordanova, M.F. Thomsen, G. Lu, S.W.H. Cowley, and K.W. Ogilvie (2006), A two-ejecta event associated with a two-step geomagnetic storm, *J. Geophys. Res*. **111**, A11104, DOI: 10.1029/2006JA011893.

Fejer, B.G., R.W. Spiro, R.A. Wolf, and J.C. Foster (1990), Latitudinal variation of perturbation electric fields during magnetically disturbed periods – 1986 Sundial observations and model results, *Ann. Geophys*. **8**, 441-454.

Forbes, J.M. (1989), Evidence for the equatorward penetration of electric fields, winds, and compositional effects in the Asian/Pacific sector during the September 17-24, 1984 ETS interval, *J. Geophys. Res*. **94,** A12, 16999-17007, DOI: 10.1029/JA094iA12p16999.

Fuller-Rowell, T.J., G.H. Millward, A.D. Richmond, and M.V. Codrescu (2002), Storm-time changes in the upper atmosphere at low latitudes, *J. Atmos. Sol.-Terr. Phys*. **64,** 12-14, 1383-1391.

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**, **A**4, 5771-5792, DOI: 10.1029/93JA02867.

Gonzalez, W.D., A.L. Clúa de Gonzalez, J.H.A. Sobral, A. Dal Lago, and L.E. Vieira (2001), Solar and interplanetary causes of very intense geomagnetic storms, *J. Atmos. Sol.-Terr. Phys*. **63**, 5, 403-412, DOI: 10.1016/S1364-6826(00) 00168-1.

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.

Gopalswamy, N., S. Yashiro, G. Michalek, H. Xie, R.P. Lepping, and R.A. Howard (2005), Solar source of the largest geomagnetic storm of cycle 23, *Geophys. Res. Lett*. **32**, L12S09, DOI: 10.1029/2004GL021639.

Hargreaves, J.K. (1995), The solar wind and the magnetosphere. **In**: *The Solar-Terrestrial Environment*, Chapter 5, Cambridge University Press, New York, 420 pp.

Huang, C-S. (2008), Continuous penetration of the interplanetary electric field to the equatorial ionosphere over eight hours during intense geomagnetic storms, *J. Geophys. Res*. **113**, A11305, DOI: 10.1029/2008JA013588.

Huang, C-S., J.C. Foster, and M.C. Kelley (2005), Long-duration penetration of the interplanetary electric field to the low-latitude ionosphere during the main phase of magnetic storms, *J. Geophys. Res*. **110**, **A**11309, DOI: 10.1029/ 2005JA011202.

Huang, C-S., K. Yumoto, S. Abe, and G. Sofko (2008), Low-latitude ionospheric electric and magnetic field disturbances in response to solar wind pressure enhancements, *J. Geophys. Res*. **113**, A08314, DOI: 10.1029/2007JA012940.

Jaggi, R.K., and R.A. Wolf (1973), Self-consistent calculation of the motion of a sheet of ions in the magnetosphere, *J. Geophys. Res*. **78**, 16, 2852-2866, DOI: 10.1029/JA078i016p02852.

Kamide, Y., N. Yokoyama, W.D. Gonzalez, B.T. Tsurutani, I.A. Daglis, A. Brekke, and S. Masuda (1998), Two-step development of geomagnetc storms, *J. Geophys. Res*. **103**, A4, 6917-6921, DOI: 10.1029/97JA03337.

Kane, R.P. (2005), Ionospheric *fo*F2 anomalies during some intense geomagnetic storms, *Ann. Geophys*. **23**, 7, 2487-2499.

Kelley, M.C. (1989), *The Earth's Ionosphere: Plasma Physics and Electrodynamics*, International Geophysics Series **43**, 10-22, Academic Press, New York.

Kelley, M.C., B.G. Fejer, and C.A. Gonzales (1979), An explanation for anomalous equatorial ionospheric electric fields associated with northward turning of the interplanetary magnetic field, *Geophys. Res. Lett*. **6**, 4, 301-304, DOI: 10.1029/GL006i004p00301.

Lei, J., W. Wang, A.G. Burns, S.C. Solomon, A.D. Richmond, M. Wiltberger, L.P. Goncharenko, A. Coster, and B.W. Reinisch (2008), Observations and simulations of the ionospheric and thermospheric response to the December 2006 geomagnetic storm: Initial phase, *J. Geophys. Res*. **113**, **A**01314, DOI: 10.1029/2007JA012807.

Liu, L., W. Wan, M-L. Zhang, B. Zhao, and B. Ning (2008), Prestorm enhancements in NmF_2 and total electron content at low latitudes, *J. Geophys. Res*. **113**, **A**02311, DOI: 10.1029/2007JA012832.

Miah, M.A. (1989), Observation of low energy particle precipitation at low altitude in the equatorial zone, *J. Atmos. Terr. Phys*. **51**, 6, 541-549, DOI: 10.1016/0021-9169(89)90094-9.

Mikhailov, A.V., and L. Perrone (2009), Prestorm NmF2 enhancements at middle latitudes: delusion or reality?, *Ann Geophys.* **27**, 3, 1321-1330.

Mikhailov, A., and K. Schlegel (1998), Physical mechanism of strong negative storm effects in the daytime ionospheric F2 region observed with EISCAT, *Ann Geophys.* **16**, 5, 602-608, DOI: 10.1007/s00585-998-0602-z.

Nielsen, E., and F. Honary (2000), Observations of ionospheric flows and particle precipitation following a Sudden Commencement, *Ann. Geophys*. **18**, 8, 908-917, DOI: 10.1007/s00585-000-0908-y.

Pi, X., M. Mendillo, W.J., Hughes, M.J. Buonsanto, D.P. Sipler, J. Kelly, Q. Zhou, G. Lu., and T.J. Hughes (2000), Dynamical effects of geomagnetic storms and substorms in the middle-latitude ionosphere: An observational campaign, *J. Geophys. Res*. **105**, **A**4, 7403-7417, DOI: 10.1029/1999JA900460.

Prolss, G.W. (1995), Ionospheric F-region storms. **In**: H. Volland (ed.), *Handbook of Atmospheric Electrodynamics*, CRC Press, Boca Raton, FL, 195-248.

Sastri, J.H., K. Niranjan, and K.S.V. Subbarao (2002), Response of the equatorial ionosphere in the Indian (midnight) sector to the severe magnetic storm of July 2000, *Geophys. Res. Lett*. **29**, 13, 1651, DOI: 10.1029/2002GL015133.

Scherliess, L., and B.G. Fejer (1997), Storm time dependence of equatorial disturbance dynamo zonal electric fields, *J. Geophys. Res*. **102**, A11, 24037-24046, DOI: 10.1029/97JA02165.

Spiro, R.W., R.A. Wolf, and B.G. Fejer (1988), Penetration of high-latitude-electric-field effects to low latitudes during SUNDIAL 1984, *Ann. Geophys*. **6**, 39-49.

Strickland, D.J., R.E. Daniell, and J.D. Craven (2001), Negative ionospheric storm coincident with DE 1-observed thermospheric disturbance on October 14, 1981, *J. Geophys. Res*. **106**, **A**10, 21,049-21,062, DOI: 10.1029/2000JA00 0209.

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**, L22101, DOI: 10.1029/2006GL027737.

Tsurutani, B.T., W.D. Gonzalez, F. Tang, and Y.T. Lee (1992), Great magnetic storms, *Geophys. Res. Lett*. **19**, 1, 73-76, DOI: 10.1029/91GL02783.

Tsurutani, B.T., W.D. Gonzalez, A.L.C. Gonzalez, F. Tang, J.K. Arballo, and M. Okada (1995), Interplanetary origin of geomagnetic activity in the declining phase of the solar cycle, *J. Geophys. Res*. **100**, **A**11, 21,717-21,733, DOI: 10.1029/95JA01476.

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.

Wygant, J., D. Rowland, H.J. Singer, M. Temerin, F. Mozer, and M.K. Hudson (1998), Experimental evidence on the role of the large spatial scale electric field in creating the ring current, *J. Geophys. Res*. **103**, **A**12, 29,527-29,544, DOI: 10.1029/98JA01436.

Zhao, B., W. Wan, K. Tschu, K. Igarashi, T. Kikuchi, K. Nozaki, S. Watari, G. Li, L.J. Paxton, L. Liu, B. Ning, J.-Y. Liu, S.-Y. Su, and H.P. Bulanon (2008), Ionosphere disturbances observed throughout Southeast Asia of the superstorm of 20-22 November 2003, *J. Geophys. Res*. **113**, **A**00A04, DOI: 10.1029/2008JA013054.

Zhizhin, M., E. Kihn, R. Redmon, D. Medvedev, and D. Mishin (2008), Space Physics Interactive Data Resource — SPIDR, *Earth Sci. Inform*. **1**, 2, 79-91, DOI: 10.1007/s12145-008-0012-5.

Received 17 May 2009
Received in revised form 28 December 2009
Accepted 28 December 2009