ISSN 1913-1844 (Print)
ISSN 1913-1852 (Online)

# MODERN APPLIED SCIENCE

Vol. 7, No. 6 June 2013

CANADIAN CENTER OF SCIENCE AND EDUCATION

# Editorial Board

# Contents

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# Generation of Seismic-Related DC Electric Fields and Lithosphere-Atmosphere-Ionosphere Coupling

Valery Sorokin[1] & Masashi Hayakawa[2]

[1] Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation (IZMIRAN), Russian Academy of Sciences, Troitsk, Moscow, Russia

[2] The University of Electro-Communications (UEC), Advanced Wireless Communications Research Center (AWCC), 1-5-1 Chofugaoka, Chofu Tokyo, Japan

Correspondence: Masashi Hayakawa, The University of Electro-Communications (UEC), Advanced Wireless Communications Research Center (AWCC), 1-5-1 Chofugaoka, Chofu Tokyo 182-8585, Japan. Tel: 81-424-44-6349. E-mail: hayakawa@hi-seismo-em.jp

Received: March 7, 2013 Accepted: April 5, 2013 Online Published: May 6, 2013

doi:10.5539/mas.v7n6p1 URL: http://dx.doi.org/10.5539/mas.v7n6p1

**Abstract**

This paper reviews modeling of the influence of earthquake (EQ) preparation processes on the ionosphere through the electric field and electric current occurring in the global atmosphere–ionosphere electric circuit. Our consideration is based on the satellite-and ground-based experimental data of electric fields, plasma and electromagnetic perturbations obtained for several days before an EQ. We have ruled out the models which are not consistent with the experimental data on the electric fields in the ionosphere and also on the ground surface. There has then been proposed a new model of the generation of electric field on the basis of injection of charged aerosols into the atmosphere, and we discuss the mechanism of lithosphere-atmosphere-ionosphere coupling. It is then shown that such changes in the electric field within the ionosphere induce a variety of plasma and electromagnetic phenomena associated with an impending EQ.

**Keywords:** earthquakes, earthquake precursors, DC electric field, lithosphere-atmosphere-ionosphere coupling

## 1. Introduction

Numerous plasma and electromagnetic anomalies observed within the ionosphere above the regions of seismic activity are found as evidence that processes of earthquake (EQ) preparation effects take place in the ionosphere for several days before an EQ. Observations of anomalous plasma and electromagnetic phenomena in the ionosphere over the zones of seismic activity were extensively discussed in many reviews and books (Gokhberg et al., 1988; Liperovsky et al., 1992; Molchanov, 1993; Buchachenko et al., 1996; Varotsos, 2001; Hayakawa & Molchanov, 2002; Pulinets & Boyarchuk, 2004; Tronin, 2006; Sorokin, 2007; Molchanov & Hayakawa, 2008; Hayakawa, 2009, 2012; Uyeda et al., 2009; Sorokin & Chmyrev, 2010; Hayakawa & Hobara, 2010), and these phenomena are considered as manifestation for the existence of lithosphere-atmosphere-ionosphere (LAI) coupling or interaction. There are ionospheric effects as a result of the simultaneous actions of various factors such as acoustic waves, electric fields, electromagnetic radiation, chemically active substances, etc. An important role in the formation of these factors is played by aerosols of the lower atmosphere, which influence its conductivity and forms an external electric charge and a current by atmosphere dynamics. Seismic activity is accompanied by the injection of soil aerosols and radioactive substances into the atmosphere, so that the enhancement of such activity in seismic regions changes the state of ionospheric plasma and electromagnetic field at the temporal scale for a few days before an EQ.

An analysis of satellite data showed the presence of electromagnetic perturbations over a wide frequency range. These perturbations are localized within the magnetic field tube conjugate with the seismic focus of an impending EQ. There are quite many papers on those satellite recordings of wave and plasma disturbances possibly associated with an individual EQ or several strong EQs (Parrot & Lefeuvre, 1985; Larkina et al., 1989; Chmyrev et al., 1989; Galperin et al., 1993; Molchanov et al., 1993; Pulinets et al., 1994; Parrot, 1994, 2009, 2011; Chmyrev et al., 1997). The presence of electron density fluctuations in the ionosphere above seismic regions was substantiated by ample satellite data (Afonin et al., 1999), and there were recorded changes in the ionic composition and temperature of the plasma in the upper ionosphere and perturbations of the height profile

of the ionospheric F region (Pulinets et al., 1994; Boskova et al., 1994). An analysis of satellite images of the Earth's surface in the infrared (IR) frequency range showed the presence of stable and unstable components of the anomalous IR radiation flux above active crust faults; this flux corresponded to an increase in the temperature of the near-Earth layer by several degrees (Qiang et al., 1999; Tronin, 1999; Tronin et al., 2002; Ouzounov et al., 2012). Simultaneously with electromagnetic and plasma phenomena in the ionosphere, there were observed an increase in the concentration of certain gases (e.g., $H_2$, $CO_2$, and $CH_4$) by several orders of magnitude, an increase in atmospheric radioactivity (related to such radioactive elements as radon, radium, uranium, thorium, and actinium and their decay products), and an increase in the injection of soil aerosols (Alekseev & Alekseeva, 1992; Virk & Singh, 1994; Heincke et al., 1995; Igarashi et al., 1995; Biagi, 2009; Yasuoka et al., 2012).

The ground-based observations which are aimed at searching electromagnetic phenomena related with processes of EQ preparation and evolution, have started in the last tens of years of XX century. The following phenomena were observed with a lot of hopes: ULF magnetic and electric emissions (Fraser-Smith et al., 1990; Molchanov et al., 1992; Kopytenko et al., 1993; Hayakawa et al., 1996a), acoustic emissions (Gorbatikov et al., 2002), amplitude and phase anomalies of subionospheric VLF/LF signals from powerful transmitters (Hayakawa et al., 1996b; Molchanov & Hayakawa, 1998; Rozhnoi et al., 2004), ionosphere perturbations measured by the ionospheric sounding (Pulinets et al., 1994; Liu, 2009), airglow anomalies (Gladychev & Fishkova, 1994) and some others. Uniform and global-size observations of possible ionospheric effects from many EQs can be carried out together with the estimation of the size of seismo-active region.

A joint analysis of observational results led us to conclude that seismic activity stimulated the development of intense processes in the lower atmosphere. Earth's surface seismic waves, chemically active and radioactive substances, and charged aerosols are likely to act simultaneously on the lower atmosphere. There then occur heating of the lower atmosphere, sharp changes in its electrophysical parameters, the generation of acoustic waves, and the formation of external electric currents. The acoustic action also appears on the ionosphere because of the upward propagation of infrasonic waves (Liperovsky et al., 1997). Processes in the lower atmosphere (seismic waves, atmosphere heating, and the injection of gases) result in the generation and upward propagation of internal gravity waves (IGWs), which might perturb the ionosphere (Gokhberg et al., 1996). The formation of ultralow-frequency radiation on the Earth's surface by lithospheric sources is considered in Molchanov and Hayakawa (1995), Molchanov (1999), Surkov and Pilipenko (1999), and Sorokin and Pokhotelov (2010), and the possibility of its penetration into the ionosphere is discussed in Molchanov et al. (1995). Numerous studies of the nature of atmosphere-ionosphere interactions aimed at determining their mechanism were performed. For instance, the physical processes of formation of currents in the lithosphere and propagation of their radiation into the ionosphere were considered in Fitterman (1979) and Pilipenko et al. (1999). Alperovich et al. (1979) discussed acoustic actions resulting in ionospheric perturbations and the generation of geomagnetic pulsations was discussed. Similar works were performed for numerous chains of processes between sources and measured parameters. Another approach to study EQ precursors consists in a joint analysis of a set of possible parameters observed. Such an analysis can be physically based on a model that makes it possible to interpret satisfactorily most of satellite- and ground-based observations as a manifestation of one cause. In this case measured parameters proved to be interrelated by certain regularities. One of the important problems of atmosphere-ionosphere interactions is the search for a chain of processes related to acting factors and identification of a set of observed effects of a common nature. It is considered that principal causes of lithosphere–ionosphere coupling are the generation of both acoustic waves and electric field in the seismic region. Below we discuss only one of these influence factors; namely, the purpose of this paper is to discuss the cause and consequences of electric field occurring at an eve of EQs.

## 2. Basic Properties of DC Electric Fields

The seismic-related DC electric fields in the ionosphere had been, for the first time, revealed by Chmyrev et al. (1989). They analyzed the vertical component of quasi-static (DC) electric field $E_z$, and we show one example. They observed such an enhanced $E_z$ onboard the "Intercosmos-Bulgaria 1300" satellite within a 15-min interval before an EQ occurred on January 12, 1982 at 17.50.26 UT. The quasi-static electric field with amplitude of 7-8 mV/m was observed in two zones: above the EQ focus and in its magnetically conjugate region, and the size of those zones was 1°~1.5° in latitude.

Subsequent investigations of DC electric field in the ionosphere based on direct satellite measurements over seismic regions were carried out by Gousheva et al. (2006, 2008, 2009), who analyzed hundreds of seismic events in order to detect DC electric field enhancement in the ionosphere connected with EQs. Seismic events with different magnitudes in different tectonic structures at different latitudes were observed. They selected the

orbits with distance less than 250 km with respect to the EQ epicenter, not crossing terminator and during low magnetic activity. Let us present one case study of their registration results. The DC electric field 5-10 mV/m was detected in the magnetic conjugate regions 11–13 hours before two EQs (magntitude around 5) occurring on 25.08.1981 at 16:54:39 UT and 17:29:07 UT correspondingly (Gousheva et al., 2008). Statistical analyses of the satellite data by Gousheva et al. (2008, 2009) led them to make a conclusion on the existence of seismic-related quasi-static electric field in the ionosphere. The duration of electric field disturbances with amplitude of the order of 10 mV/m can be up to 15 days, and the electric field disturbances in the daytime and nighttime ionospheres were on the same order.

Direct observations of quasi-static electric field in the ionosphere are confirmed by computational modeling of the ionospheric perturbation occurring at an eve of EQs. Spatial distributions of the total electron content (TEC) obtained by GPS receivers in the seismic region were analyzed (Liu, 2009; Pulinets, 2009a), and those TEC anomalies are tried to be interpreted with the use of global model of the upper atmosphere which describes the thermosphere, ionosphere and plasmasphere as an integrated system. The model is based on integration of the non-stationary three-dimensional equations of continuity, impulse and energy balance of multi-component gas simultaneously with the equation for electric field potential. In the frame of computer simulations there was sought an additional electric field which leads to the TEC perturbation coincident with the one observed in the EQ preparation region. For example, Zolotov et al. (2008) considered an EQ in Peru on 26.09.2005. The characteristics of TEC disturbances were given in Zakharenkova et al. (2008), and the TEC perturbation was observed during six days before the EQ from 21.09.2005 till 26.09.2005. Based on the computer simulations Zolotov et al. (2008) have shown that the observable TEC perturbation is due to an additional electric field with an amplitude of 6 mV/m. It is further suggested by Klimenko et al. (2011, 2012) and Namgaladze et al. (2009) that a possible general cause of TEC perturbation is a vertical plasma drift by the zonal electric field. Computer simulations by Klimenko et al. (2012) have shown that the amplitude of electric field disturbance is required to be 3-9 mV/m.

At the same time, observations of the quasi-static electric field on the Earth's surface in seismic regions were carried out by different workers (Kondo, 1968; Jianguo, 1989; Nikiforova & Michnovski, 1995; Vershinin et al., 1999; Hao et al., 2000; Rulenko, 2000). Analyses of those publications show that the local electric field surges with large amplitude reaching several kV/m are observed during the EQ preparation, but their duration is of the order of ten minutes. However, there are absent visible electric field disturbances with duration of several days observed simultaneously over the horizontal distance of hundreds of kilometers.

The indirect confirmation of electric fields occurring in the atmosphere is the observational results of VHF emissions propagating from the source located in the troposphere over a region of EQ preparation (Vallianatos & Nomicos, 1998; Ruzhin et al., 2000; Hayakawa et al., 2006; Ruzhin & Nomicos, 2007; Yonaiguchi et al., 2007a, b; Yasuda et al., 2009). VHF radiations are found to have occurred for several days before an EQ, and their duration reaches several days. If the VHF electromagnetic radiation propagated over a distance more than a wavelength, then the condition of optical propagation is fulfilled, so that it is possible to receive the signal at distance of the order of 300 km just in the case that its source is located in the atmosphere above Earth's surface. The region of generation of VHF electromagnetic radiation is found to be at the altitudes of the order of several kilometers above EQ epicenters located behind the horizon (Fukumoto et al., 2001; Yasuda et al., 2009).

Consequently, both the direct and indirect data of DC electric field observations in the atmosphere and ionosphere over a seismic region allow us to formulate its basic properties. The basic experimental results are summarized as follows:

- The enhancement of seismic activity produces DC electric field disturbances in the ionosphere of the order of 10 mV/m.
- These disturbances occupy the region with horizontal spatial scale from hundreds to thousands km over the seismic region.
- DC electric field enhancements occur in the ionosphere from hours to 10 days before an EQ.
- DC electric field disturbances in the daytime and nighttime ionospheres have the same order of magnitude.
- DC electric field disturbances can reach the breakdown value during from hours to 10 days in the atmosphere at altitudes 1 to 10 km over the EQ zone.
- The quasi-stationary electric field on the Earth's surface does not exceed its background value simultaneously in the seismic area during several days.

## 3. Penetration of DC Electric Field Into the Ionosphere

Lithospheric activity stimulates the processes which are followed by the electric field generation. The enhancement in number density of charged aerosols by one-two orders and the increase in atmosphere radioactivity level by the injection of radon and other radioactive substances are observed during days and weeks before an EQ (Alekseev & Alekseeva, 1992; Virk & Singh, 1994; Voitov & Dobrovolsky, 1994; Heinke et al., 1995; Pulinets et al., 1997; Yasuoka et al., 2006, 2012; Omori et al., 2007; Biagi, 2009). Data on injection of the soil gases such as radon, helium, hydrogen, carbon dioxide in the surface atmosphere with horizontal spatial scale of 500 km during from several hours to several weeks before an EQ have been reported by King (1986). Igarashi et al. (1995) described the surge in five times of the radon concentration in the soil water, and the data on significant emissions of metallic aerosols Cu, Fe, Ni, Zn, Pb, Co, Cr and radon were given by Boyarchuk (1997). Quasi-static electric field disturbances in the ionosphere are observed at the same time as the injection of active substances in the lower atmosphere.

There are observed the local short-time releases of active substances along with large scale growth of the level of active substances in the lower atmosphere. They can generate the impulses of electric field near the Earth's surface, whose amplitude can reach tens kV/m but its duration does not exceed tens of minutes. A model of the generation of pulses of local electric fields with characteristic time scales of 1-10 min for the atmospheric conditions above fracture regions of EQs was considered by Liperovsky et al. (2005, 2008). They have proposed that aerosols, increased ionization velocity and upstreaming air flows occur at night-time conditions, and that water condensates at the aerosols at night when the temperature in the near-earth air is low and the relative humidity increases above earth-fracture regions. Then, the relatively large aerosol particles are mainly negatively charged, while the charge of smaller particles is overwhelmingly positive. It is assumed that aerosol clouds of small dimensions are suddenly injected into the locally heated surface atmosphere and move with the air up to higher altitudes. The vertical velocity of small particles is much smaller than that of large ones, which is equal to a few cm/s. As a consequence of the shift between the small and large particles, there occur the local pulses of the electric field in the atmosphere. The amplitude of such a field is estimated as $10^3$~$3\times10^3$ V/m, but the relaxation time of a cloud of aerosols is estimated 10 minutes. Anomalous emanation of radons preceding a large EQ was observed by Omori et al. (2007), who have analyzed atmospheric radon concentrations and estimated changes of electrical conditions in the atmosphere due to the preseismic radon anomaly. These authors used the model by Liperovsky et al. (2005), and they have shown that the radon emanation reduces the atmospheric electric field by 40%. Their estimation of field amplitude gives $10^4$~$10^5$ V/m at the observable value of radon emanation, but unfortunately there are no calculations of electric field in the ionosphere in the above-mentioned works. Nevertheless it is assumed that this impulse electric field can be a source of lithosphere–ionosphere coupling. We should note that this impulse field is observed only in local regions. The duration of such impulses is 10 minutes, while ionospheric precursors and DC electric field in the ionosphere exist during a much longer interval of several days. The field occurs inside a dipole layer of charged aerosols cloud and the field slumps outside the dipole layer. Consequently, the local impulse electric field observed on the Earth's surface cannnot be a cause of the ionospheric effects and appearance of DC electric field in the ionosphere; that is, the radon injection in the frame of model does not affect the lithosphere–ionosphere coupling.

A generation mechanism of electric field in the lithosphere based on the result of laboratory experiments has been proposed by Freund et al. (2006, 2009) and Freund (2010). Their experiments show that when stresses are applied to one end of a block of igneous rocks, two currents flow out of the stressed rock volume. One current is carried by electrons and the other current is carried by p-holes. Positive electric potential, ionization of air molecules and corona discharge occur on the rock surface. It is assumed that air ionization is a cause of ionospheric disturbances, glows and IR emissions, but there are no calculations on the possible effect of this source to the ionosphere. This mechanism seems to be used to interpret the impulse phenomena because the source duration is over 10 minutes, but it seems to be an unlikely explanation of the existence of DC electric field over a long period of time.

Below we consider the generation mechanisms for quasi-static electric fields in the ionosphere. Spatial distribution of this field has a horizontal scale (100–1000) km and its duration is from tens hours to tens of days. The field is quasi-static if its temporal variation exceeds considerably the relaxation time ($\tau$) of charges in the surface atmosphere $\tau \sim \varepsilon_0/\sigma \sim 10$~$30$ min ($\varepsilon_0$ is the permittivity of free space, and $\sigma$ is the surface atmosphere conductivity). An equivalent circuit is often used to explain the generation of atmospheric electric field (Goldberg, 1984; Sapkota & Varshneya, 1990; Rycroft et al., 2000). The current flowing in the circuit is excited by a generator which is the resultant action of thunderstorms all over the world. The fair weather current density is of the order of $10^{-12}$ A/m$^2$ in the closed circuit (e.g., MacGorman & Rust, 1998; Rakov & Uman, 2002). It is

assumed that the conductivity of near-Earth atmosphere of $10^{-14}$ S/m yield an electric field value on the Earth's surface of $10^2$ V/m (Rakov & Uman, 2002). The fair weather current gives an electric field of $10^{-3}$ mV/m in the ionosphere with conductivity $10^{-6}$ S/m. Since the background electric field of magnetospheric and ionospheric origin has a value (0.1~1) mV/m in the middle-latitude ionosphere, then the field of atmospheric origin with intensity of $10^{-3}$ mV/m is negligible in the ionosphere. The variation of DC electric field in the ionosphere over a seismic region can be realized in two different ways. First of all one can change the load resistance and in a different way one can include an additional EMF (electro-motive force) in the global circuit.

Let us consider the first way. The processes of EQ preparing impact take place in the lower atmosphere which contributes to over 80% of load resistance of the global circuit. The injection of radioactive and chemical substances and aerosols into the atmosphere, and the variation of aerosols size and atmospheric state result in a change of load resistance. In the final analysis, all of these processes change conductivity of the surface atmosphere. In Figure 1 there is depicted the circuit with selected part of current over a seismic region. The resistance of atmosphere over thunderstorms is denoted by $R_1$, $R_2$ is the resistance of the region of thunderstorms activity, $R_3$ is the resistance of near-Earth atmosphere, $R_i$ is the resistance of ionosphere, and $I$ is the fair weather current (Rycroft et al., 2000). The load resistance $R$ is much smaller than all of these resistances. The disturbed part of circuit (designated by red color in Figure 1) consists of the following resistances: $r_1$ is the resistance of upper troposphere, $r_i$ is the resistance of ionosphere over a seismic region, and $r_2$ is the resistance of surface atmosphere. The disturbance of surface atmospheric conductivity results in the variation of $r_2$ and the electric current in this part of the circuit. For the first time, Sorokin and Yaschenko (2000) and Sorokin et al. (2001) have carried out the calculation of altitude dependence of DC electric field variation in the Earth-ionosphere layer produced by a source of ionization and a growth of conductivity in the lower atmosphere. They performed theoretical investigations of the atmospheric ionization by alpha particles and gamma quantum of the nuclear decay, and they calculated the altitude dependence of ionization rate and conductivity for different levels of radioactivity. It is shown that the electric field can be changed by 1.5-2 times in the ionosphere by the growth of conductivity in the surface atmosphere. Such a variation field does not impact onto the ionosphere because the amplitude of variations is considerably smaller than the background value. That is, the field variation is invisible in the ionosphere. This result is confirmed by Omori et al. (2008), who have shown that the quasi-static electric field is reduced by 1.5 times due to the growth of radioactivity and conductivity during the radon injection. In spite of evident results of continued unsuccessful attempts to explain the appearance of seismic-related quasi-static electric fields in the ionosphere due to the variation of conductivity in the lower atmosphere, for example, Pulinets (2009a) has assumed that an anomalous electric field in the ionosphere over an active fault occurs by the variation of conductivity in the near-earth atmosphere. The conductivity is varied due to the growth of additional radon ionization and the reduction in ions mobility by the generation of large clusters. There are missing both the proof of speculation and the calculation of field value in the ionosphere. According to Omori et al. (2007, 2008), the radon surge with magnitude 10 $Bq/m_3$ leads to an increase in ionization rate up to ($10^6$~$10^7$) $1/m^3s$. As a result, conductivity of the near-earth atmosphere is increased in 1.5 times, and the field is varied approximately by 1.5 times in the ionosphere as well. Since the field of fair weather in the ionosphere is $10^{-3}$ mV/m, then its variation by 1.5 times will be much smaller than the background value (0.1~1) mV/m. Harrison et al. (2010) have shown that an increase of ionization rate by radon in two times leads to a variation of the current flowing from the ionosphere to the Earth in 10%, and then the field is varied on the same quantity in the ionosphere. Thereby, any models based on the assumption that the ionization of lower atmosphere leads to a conclusion that the seismic-related electric field formation in the ionosphere is in apparent contradiction with experimental data that the electric field is up to 10 mV/m in the ionosphere.

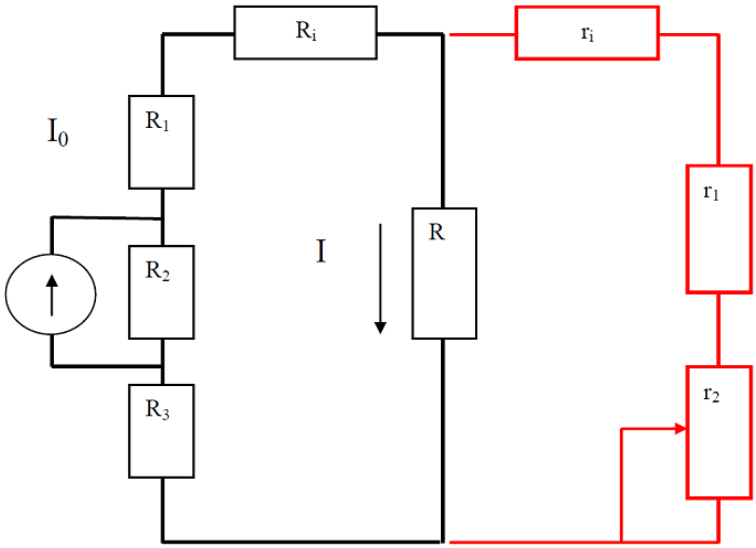



Figure 1. Equivalent electric circuit of DC electric field formation in the ionosphere over a region of conductivity disturbance in the lower atmosphere. Black color denotes the conventional global circuit, and red color indicates the part of circuit over the region of disturbed conductivity

We consider a different way of DC electric field formation in the ionosphere. The electric current and electric field are varied due to the inclusion of a seismic-related EMF in the global circuit. The EMF can be located in the lithosphere, in the atmosphere and in the vicinity of boundary between the lithosphere and atmosphere. The scheme of altitude dependence of total electric current $\mathbf{j} = \sigma\mathbf{E}+\mathbf{j}_e$ ($\mathbf{j}_e$: EMF external current) in these three cases is depicted in Figure 2. The origin of coordinate system is located on the Earth's surface. We consider the case corresponding to the left panel of Figure 2. In the frame of this model it is assumed that the EMF is located in the lithosphere and the field is transferred through the atmospheric layer with specified altitude dependent electric conductivity. The vertical component of electric field disturbance is given on the Earth's surface, and the section of closed global electric circuit is depicted in Figure 3. The uniform Ohm's law for a subcircuit without the EMF is performed in the Earth-ionosphere layer. The nature of electric field source on the Earth surface and its characteristics are not discussed in the papers based on this model. The source of field is expected to create a quasi-static electric current in the circuit for several days. The field in the ionosphere is calculated at given spatial distributions of its vertical component on the Earth's surface (Kim & Hegai, 1999; Pulinets et al., 2000, 2003; Grimalsky et al., 2003; Rapoport et al., 2004; Denisenko et al., 2008; Ampferer et al., 2010). Electric fields in the ionosphere are computed for different boundary conditions, shape and size of field horizontal distribution on the Earth's surface. Kim and Hegai (1999) showed that the field reaches (0.3~0.7) mV/m in the nighttime ionosphere if the field near the Earth has a value of 1000 V/m. Since the field in the seismic region does not exceed approximately 100 V/m (Kondo, 1968; Vershinin et al., 1999), then their calculated value of field in the ionosphere should be reduced to (0.03~0.07) mV/m. Taking into account that the conductivity of daytime ionosphere is larger than that of nighttime ionosphere by one-two order, the field value in the daytime ionosphere is approximately $10^{-3}$ mV/m. Calculations fulfilled in Pulinets et al. (2000, 2003) show that electric field in the nighttime ionosphere can reach (0.1~1) mV/m if it reaches a value ($10^3$~$10^4$) V/m on the Earth's surface in a seismic area with horizontal scale 100 km. This value of field on the ground surface is required to remain during several days, but such a field is unlikely to exist. Calculations performed by Denisenko et al. (2008) confirm this conclusion. It is shown that the field reaches a value $10^{-3}$~$10^{-4}$ mV/m in the ionosphere at the maximal field value $E_0$ = 100 V/m on the Earth surface. So, we can say that there exists, in the ionosphere, no static electric field of lithospheric origin. This can be obtained from a simple consideration of the continuity equation $\nabla\cdot\mathbf{j} = \nabla\cdot\sigma\mathbf{E} = 0$ for the vertical conductivity current $j = \sigma(z)E(z)$ in the conductive atmosphere. The estimate of maximum magnitude can be made simply in 1D (one dimensional) approximation $d\sigma E/dz = 0$. Let $\sigma_0$, $\sigma_1$ be the conductivity in the near-ground of the atmosphere and that in the ionosphere and $E_0$, $E_1$ are the electric fields near the ground and in the ionosphere, then we obtain $E_1 = E_0(\sigma_0/\sigma_1)$. Taking into account that $\sigma_0 \approx 10^{-14}$ S/m; $\sigma_1 \approx 10^{-6}$ S/m; $E_0$ = 100 V/m we find $E_1 \approx 10^{-3}$ mV/m, which is four orders of magnitude lower than the background ionospheric field. Thus the considered model is found to contradict with the well-known experimental data

which indicate that the preparation processes of large magnitude EQs are accompanied by an enhancement of DC electric field in the ionosphere over the epicenter zone up to 10 mV/m.

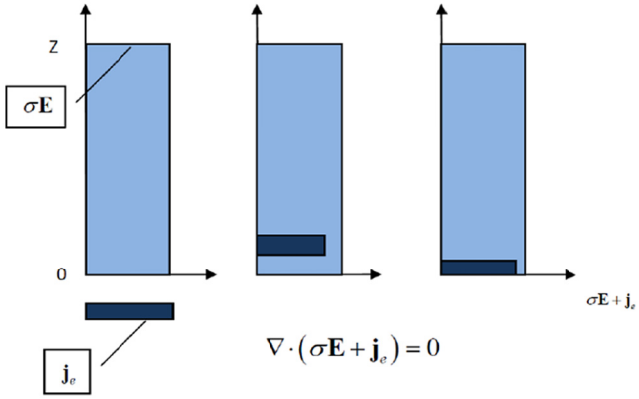



Figure 2. Dependence on altitude of the total electric current in the atmosphere. Blue figures are the conductivity current σE, and dark rectangles are the external current of EMF (electro-motive force) $j_e$

The case of EMF location in the atmosphere is corresponding to the middle panel of Figure 2. The same situation is expected for thunderstorms, and the penetration of the electric field of thunderstorm clouds into the ionosphere had been calculated by Park and Dejnakarintra (1973). This method was used to study the seismic-related electric field penetration in the ionosphere in several works. There is a principal difference between the phenomenon of electric field penetration into the ionosphere during thunderstorm and EQ preparation. Namely, the quasi-static electric field with magnitude up to ($10^3$~$10^4$) V/m is observed under a thunderstorm cloud (e.g., Rakov & Uman, 2002), while the field is not exceeding its background value on the surface of EQ preparing area. Therefore, the use of the above-mentioned method does not allow us to elaborate the mechanism for DC electric field penetration in the ionosphere at EQ preparing, for example, in Molchanov and Hayakawa (1996) and Pulinets et al. (2000).

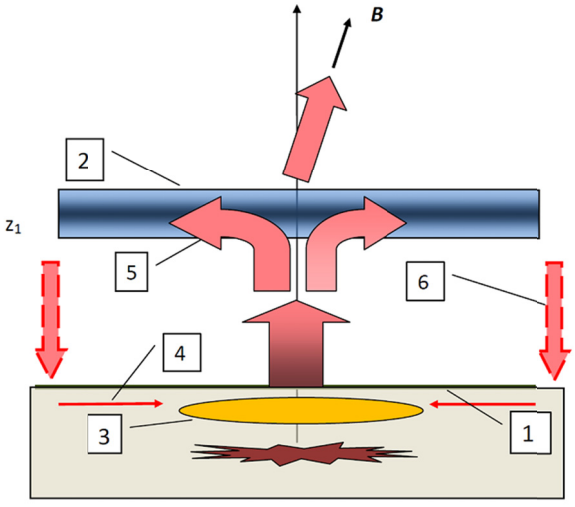



Figure 3. Model for DC electric field penetration from the lithosphere into the ionosphere. 1. Earth surface, 2. Conductive layer of the ionosphere, 3. Lithospheric source of electric field, 4. Electric field in the ground, 5. DC electric field in the ionosphere, and 6. Atmosphere-ionosphere electric circuit

It seems that the only possible way to explain results of DC electric field observation in the ionosphere over a seismic region is illustrated by the right panel of Figure 2. A principally different model which is actively developed now (Sorokin et al., 2001; Sorokin et al., 2005a; Sorokin et al., 2007; Sorokin & Chmyrev, 2010, references therein), is based on the assumption that the current source in the circuit connected with pre-EQ

processes is situated not in the lithosphere or in the atmosphere, but in the near-ground atmospheric layer. The EQ preparation processes modify the atmosphere in this layer and form an EMF in the seismic zone. The additional source of electric current is generated in the global circuit at the stage of EQ preparing. The range of EMF is formed in the near-earth atmosphere and includes the boundary between the lithosphere and atmosphere. In this case the observable electric field on the surface is located inside the EMF range, and the scheme of EMF formation is depicted in Figure 4. Upward transfer of the charged aerosols by atmospheric convection and their gravitational sedimentation result in the EMF formation. Aerosols are injected into the atmosphere by soil gases with an increase in seismic activity. The external current of EMF is reduced with altitude, while the conductivity current increases with altitude, so that the total current in the circuit is constant. The value of conductivity current near the surface can be of the order of that of fair weather current, while the quantity of external current exceeds the fair weather current by four-five orders. Therefore, the conductivity current in the ionosphere is on the order of external current of EMF near the surface. Figure 5 illustrates the circuit with selected parts of current in which we included an EMF over the seismic region. Horizontal component of the electric field $E_1$~10 mV/m corresponds to the conductivity current flowing along the ionosphere $j \sim \sigma_1 E_1 \sim (10^{-8} \sim 10^{-7})$ A/m$^2$. Following Sorokin et al. (2001) the conductivity current can be $10^{-12}$ A/m$^2$ near the surface and the electric field can be 100 V/m. This fact can be understood by a simple estimation. The continuity equation for total current in the atmosphere is expressed by a form $\nabla \cdot (\sigma \mathbf{E} + \mathbf{j}_e) = 0$, where $\mathbf{j}_e$ is the EMF external current density. The simplest field estimate for the ionosphere in 1D case gives $\sigma_0 E_0 + j_{e0} = \sigma_1 E_1$. Using this equation we find $E_1 = E_0(\sigma_0/\sigma_1)(1+j_{e0}/\sigma_0 E_0)$, where $j_{e0}$ is the density of the EMF external current near the Earth's surface. The first term on the right side of this equation corresponds to the above-mentioned model used by Kim and Hegai (1999), Pulinets et al. (2000, 2003), Denisenko et al. (2008), and Ampferer et al. (2010). If we suppose, for example, that the external current is caused by the movement of aerosols with concentration N and charge Ze under the action of vertical atmospheric convection with velocity v, then the current density can be estimated as $j_{e0} \sim ZeNv$. The aerosol charge in the atmosphere lies in the range from 100 to 800. Assuming $Z = 3 \times 10^2$; $N = 8 \times 10^9$ m$^{-3}$; v = 0.3 m/s, we obtain the electric field for the ionosphere: $E_1 \approx 10^{-6}(1+10^4)$ V/m $\approx$ 10 mVm. Thus even this rough estimate suggests that the field penetration model which has been chosen by above-mentioned authors, leads to a loss of five orders of the field magnitude in the ionosphere. The result does not depend on the way of interpolation of the altitude distribution of conductivity, so that the field penetration model used by above-mentioned authors may be wrong.

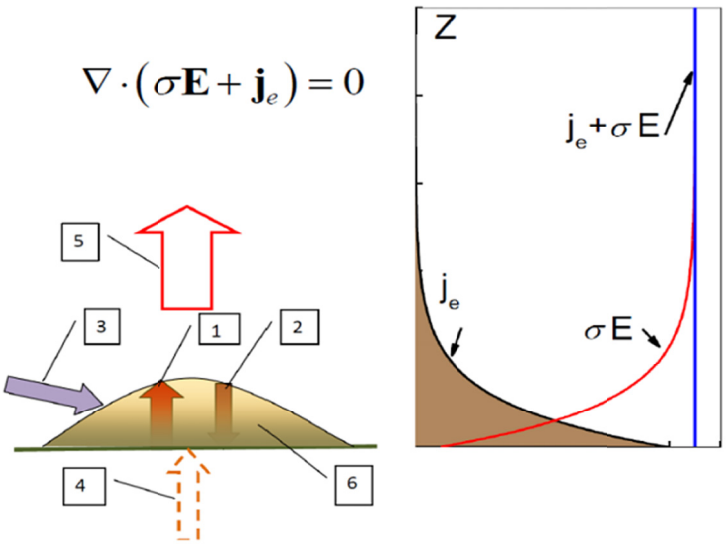



Figure 4. EMF formation in the surface atmosphere. 1. Atmospheric convection and turbulent diffusion, 2. Gravitational sedimentation, 3. Atmospheric radioactivity, 4. Soil gases, 5. Conduction electric current, and 6. External current of EMF

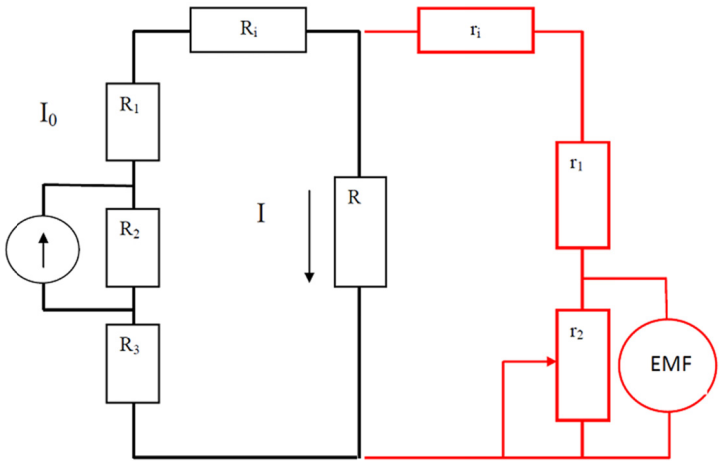



Figure 5. Equivalent circuit of DC electric field formation in the ionosphere over the region of EMF occurred in the surface atmosphere. Black color denotes the global circuit, and red color refers to the part of circuit over the region of EMF occurrence

## 4. Perturbation of DC Electric Field in the Atmosphere–Ionosphere Global Circuit

The formation mechanism of external current of EMF related with the dynamics of charged aerosols in the surface atmosphere has been considered in Sorokin and Yaschenko (2000) and Sorokin et al. (2001, 2007). The EMF occurs with the intensification of injection of charged soil aerosols or with the variation of meteorological conditions. Both the upward turbulent and convective transfer and gravitational sedimentation result in the quasi-static altitude distribution of aerosols density in the atmosphere. The turbulent transfer takes place due to two main reasons. The first is related with the vertical gradient of horizontal wind velocity and the transformation of wind kinetic energy into the energy of turbulent pulsations. The second is caused by the thermal instability of the atmosphere arising when the negative temperature gradient exceeds its adiabatic gradient. Turbulent vortices transfer aerosols from the altitudes where their concentration is high, to the altitudes of lower concentration. An equilibrium is attained when the vertical flux of aerosols is balanced by their gravitational sedimentation. The particle dynamics in the turbulent atmosphere can be described by the stochastic differential equations for the probability distribution function (Sorokin et al., 2001), which is the probability that a particle has the charge Ze at a moment $t$ on the altitude z. Spatial and temporal dependencies of concentration of aerosols, their electric charge and external current densities are expressed as the moments of distribution function. There is obtained an estimation of the value of external current $j_e(0,t)$ near the Earth's surface $j_e(0,t) = (e\sigma_0/\varepsilon_0)(Z_+N_+ - Z_-N_-)H_j$, where e is the elementary electric charge, $\varepsilon_0$ is the permittivity of free space, $\sigma_0$ is the conductivity of surface atmosphere, $Z_\pm$ is the amount of positive and negative charges on aerosol, $N_\pm$ is the concentration of positive and negative charged aerosols, and $H_j$ is the scale height of vertical distribution of external current. If we take $\sigma_0 \approx 2\times10^{-14}$ S/m, $Z_+ = 300$, $N_+ = (1\sim5)\times10^9$ 1/m$^3$, $H_j = (2\sim5)\times10^3$m, one obtains $j_e(0,t) \approx (10^{-6}\sim10^{-7})$ A/m$^2$. This value of external current of EMF shows that the mechanism discussed enables us to obtain the observed conductivity current in the ionosphere $\sim10^{-8}$ A/m$^2$. In many cases the injection of soil aerosols into the atmosphere is realized jointly with radioactive substances, so that the ionization increases in the atmosphere conductivity. The presence of aerosols might lead to a reduction in the conductivity due to the attachment of ions to aerosols. Moreover, the interaction of ions with aerosols changed the charge of aerosols. These effects would lead to the variation of EMF on the surface level of atmosphere. First, these processes have been studied theoretically in details by Sorokin et al. (2007), who have obtained the vertical distribution of ion production rate as a result of absorption in the atmosphere of the gamma radiation and the alpha particles from the decay of radioactive elements being constituents of the atmospheric radioactivity. An example of their computational results on the altitude dependence of ion production rate q = q(z) is depicted in Figure 6a. The parameter A is the index of growth of radioactivity in the near Earth layer. In Figure 6a they have chosen that the ion production rates due to the action of alpha particles and gamma rays are equal to each other, the background ion production rate in the atmosphere near the Earth's surface is $10^7$ 1/m$^3$s and the ion production rate at a maximum in the stratosphere is $4\times10^7$ 1/m$^3$s. As follows from this plot, the vertical distribution of ion formation rate is different from the exponential altitude dependence of atmospheric radioactivity. We notice a significant increase in ion production rates in maximum. Equilibrium values of ion number densities are determined by the

recombination process and the adhesion to aerosols in the atmosphere. The light singly-charged ions and the heavy ions are produced as a result of light ions adhesion to aerosols in the atmosphere near the Earth's surface. We have used the self-consistent system of nonlinear equations for the calculation of spatial distribution of external current, atmosphere conductivity and DC electric field in the Earth-ionosphere circuit at given intensity of aerosols injection and the atmosphere radioactivity. The computational result of atmospheric conductivity is depicted in Figure 6b. Conductivity depends on both the level of atmospheric radioactivity and the number density of aerosols, and we have chosen the value of aerosols concentration in the surface atmosphere equal to 2 $\times 10^9$ 1/m$^3$. Calculations show that the growth of radioactivity level in the surface atmosphere results in an increase in conductivity and the growth of aerosols concentration results in a reduction in conductivity due to the loss of light ions attached to aerosols.

Sorokin et al. (2007) have shown that the external current of EMF is defined on the atmospheric conductivity layer and the electric field is generated by its vertical component on the near-Earth level. As it is noted above, the significant (up to 1 kV/m) pre-EQ vertical electric fields on the Earth's surface have the characteristic temporal scale of the order of tens of minutes. At the same time the atmospheric electric field variations with typical scale exceeding 1 day at the distances within hundreds to thousands km from the EQ center during a seismically active period are characterized by the magnitude not exceeding ~100 V/m. The cause of such a limitation can be explained in terms of the mechanism of feedback between the disturbances of vertical electric field and the causal external currents on the Earth's surface. Such a feedback is caused by the formation of a potential barrier on the ground-atmosphere boundary at the passage of upward moving charged aerosols through this boundary. Their upward movement is performed due to the viscosity of soil gases flowing into the atmosphere. If, for example, a positively charged particle goes from the ground to the atmosphere, the Earth's surface is charged negatively. So, the downward electric field prevents more particles from penetration through the surface. At the same time this field stimulates the going out on the surface of the negatively charged particles. In the presence of such coupling the magnitude of external currents on the Earth's surface depends on the vertical component of the electric field on the surface. The first study of the mechanism of this field limitation (Sorokin et al., 2005a, 2007) yields that the value of vertical component of quasi-static electric field does not exceed a maximal value of the order of 90 V/m at any amount of external current of EMF. The self-consistent equation for the external current and electric field has been derived first by Sorokin et al. (2007). An example of those calculation results of altitude distribution of external current with taking into account the feedback with electric current is depicted in Figure 6c. It is shown that the external current is generally located at the altitudes up to 10 km and its value can be ($10^{-8}$~$10^{-6}$) A/m$^2$.

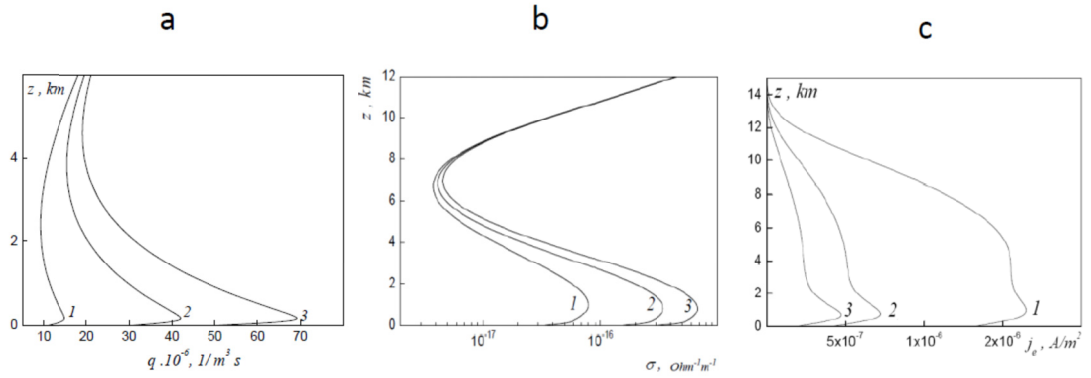



Figure 6. Computational results by a self-consistent system of non-linear equations (Sorokin et al., 2007). a) Vertical distribution of the ion production rate, b) Altitude profiles calculated for the atmosphere conductivity over the center of disturbed area, and c) Altitude dependences of external electric current over the center of disturbed region. Curves 1, 2 and 3 correspond to different levels of atmospheric radioactivity (A is the index of radioactivity growth in the near-Earth layer) 1. A=0, 2. A=2, 3. A=4

The theory of generation of the seismic-related DC electric field conforming to both the direct and indirect observation data of fields in the ionosphere is elaborated first by Sorokin et al. (2001, 2005a, 2007) and Sorokin and Chmyrev (2010). The field is associated with electric current disturbances flowing in the global electric circuit. The source of current disturbances is the EMF included in the global circuit, which is generated

by the injection of charged aerosols into the atmosphere and their upward transfer and gravitational sedimentation. The scheme of the formation of electric current disturbance in the global circuit is presented in Figure 7. The self-consistent system of nonlinear equation to compute the spatial distribution of external current, electric field, atmosphere conductivity, concentration of ions and charged aerosols is included in the theory (Sorokin et al., 2005b). Figure 8 presents an example of their calculation of spatial distribution of electric field in the ionosphere and on the Earth`s surface. Horizontal distribution of external current on the surface is chosen to be ellipse-like with axis directed under an angle to the meridian plane. These two figures illustrate that the horizontal electric field in the ionosphere reaches ~10 mV/m, while the vertical electric field on the Earth's surface is limited by a magnitude ~ 90 V/m over an active fault. Another important result is that DC electric field in the ionosphere has maximal magnitudes at the edges of area of external current. The horizontal scale of vertical electric field enhancement on the ground exceeds the characteristic horizontal scale of external current. Within this area the vertical field practically does not depend on distance. These calculations show that the field component in the meridian plane strongly depends on the magnetic field inclination.

Investigation of the spatial distribution of electric field in the atmosphere connected with disturbances of current in the global circuit based on the above-mentioned theory has been carried out by Sorokin et al. (2011, 2012a, 2012b), who have shown that the DC electric field at the certain conditions can reach the breakdown value in the troposphere. Figure 9 illustrates the spatial distribution of electric field normalized by the breakdown value for the axial symmetric horizontal distribution of external current with horizontal spatial scale 100 km on the surface. It is possible to expect occurrence of the one or two levels with thickness (1~2) km located at different altitudes (5~10) km in which the electric field reaches a breakdown value. Characteristics of these levels depend on the parameters of atmosphere and aerosols.

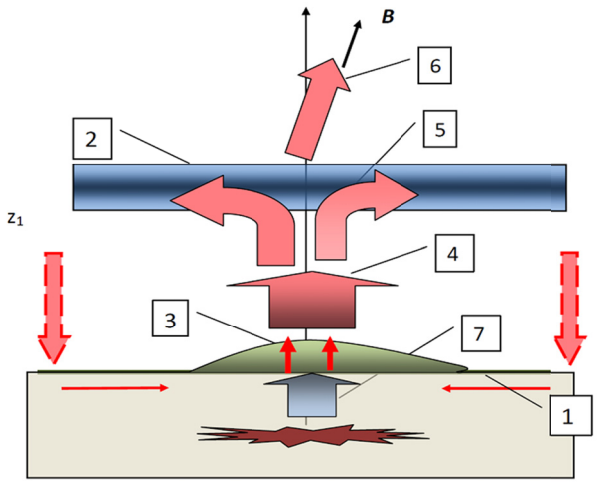



Figure 7. Model of DC electric field generation in the ionosphere by seismic related EMF (electro-motive force) in the lower atmosphere (Sorokin et al., 2005a). 1. Earth surface, 2. Conductive layer of the ionosphere, 3. External electric current of EMF in the surface atmosphere, 4. Conductivity electric current in the atmosphere–ionosphere circuit, 5. DC electric field in the ionosphere, 6. Field-aligned electric current, and 7. Charged aerosols injected into the atmosphere by soil gases

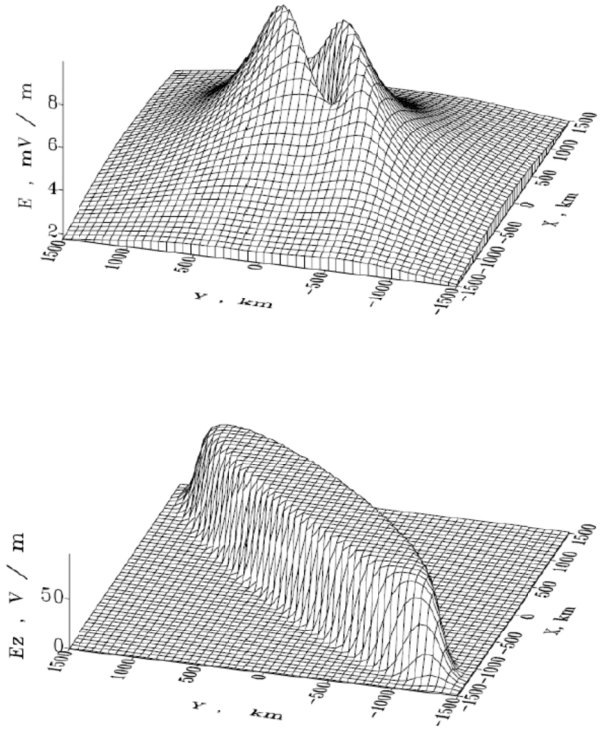



Figure 8. Spatial distributions of horizontal electric field in the ionosphere (upper panel) and vertical electric field near the Earth surface (bottom panel) over the ellipsoidal fault (Sorokin et al., 2006). The angle of fault axis orientation to the meridian plane is $\beta = 45^{\circ}$. The angle of magnetic field inclination is $\alpha = 20^{\circ}$

According to the above-mentioned calculations, the DC electric field can reach 10 mV/m in the ionosphere even if the field magnitude does not exceed 100 V/m on the Earth's surface. The theory of DC electric field amplification and penetration into the ionosphere is based on the following issues.

1) The electric field in the ionosphere and on the Earth's surface is a field of conductivity current flowing in the atmosphere-ionosphere circuit.
2) The source of conductivity current is an external current of EMF included in this circuit.
3) The EMF is formed by convective transport of the charged aerosols and the radioactive elements which are injected along with soil gases from the lithosphere into the lower atmosphere.
4) The field limitation on the Earth surface is caused by a feedback mechanism between the excited electric field and the causal external current. This feedback is produced by the potential barrier for charged particle at its transfer from the ground to the atmosphere.

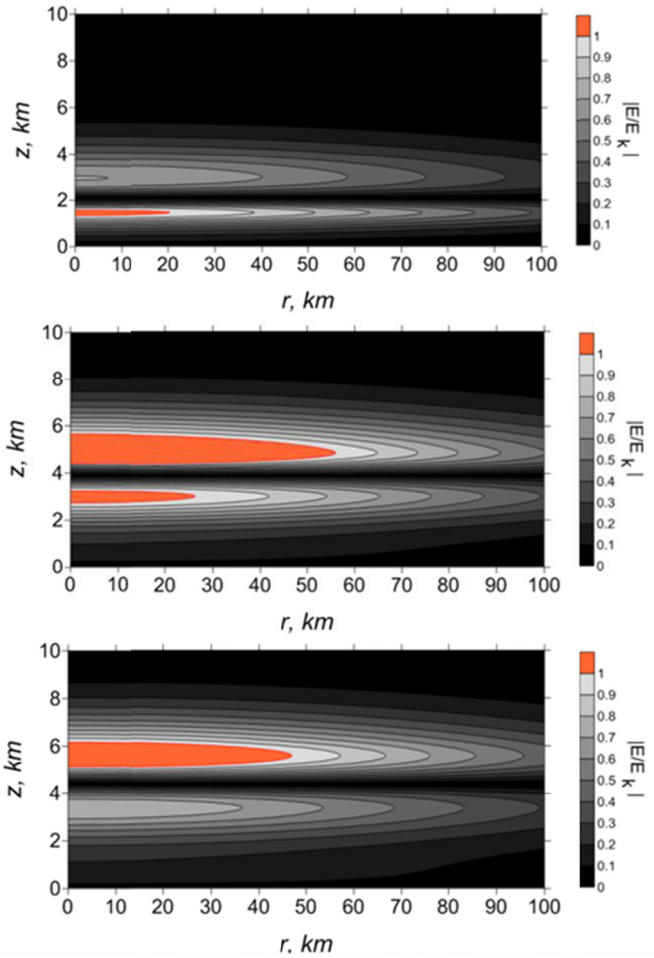



Figure 9. Examples of spatial distribution of DC electric field magnitude in the lower atmosphere normalized to the breakdown electric field (Sorokin et al., 2011)

## 5. LAI Coupling Models

In order to understand the nature of numerous ionospheric and electromagnetic EQ precursors, it is necessary to study the physical processes and then to create the model of seismicity impact to the ionosphere plasma or LAI coupling mechanism. At the present time we consider that this impact is realized generally by the two major hypotheses, (i) internal gravity waves (IGW) or (ii) electric field (see Hayakawa et al., 2004; Molchanov & Hayakawa, 2008; Pulinets & Ouzounov, 2011; Kuo et al., 2011).

Though we do not go into the details of the former channel, we make a brief description on this channel. Gokhberg and Shalimov (2000) gave the analyses of experimental data obtained at the final stage of EQ preparing, who considered the ionospheric perturbations developing as ionospheric irregularities for several days before strong EQs. Those authors believed that the occurrence of these irregularities is connected with IGW propagation through the ionosphere. Their source could be the long wave earth oscillations, local green gas effect or an unsteady injection of lithospheric gases, but the more effective mechanism is the growth of lithospheric gas emanation into the atmosphere. They suppose that the IGW generation should be considered as a mechanism of LAI coupling because the atmosphere stratification favors the wave amplification as it propagates upward. Further they discussed various likely scenarioes of the accompanying processes developing in the ionosphere. Molchanov et al. (2004) and Molchanov (2009) presented recently a general concept of the role of IGW in the LAI coupling. The atmospheric perturbation of temperature and density could follow preseismic hot water/gas releases resulting in the generation of atmospheric gravity waves (AGW) with periods in a range of 6-60 min. Seismo-induced AGW could lead to the modification of ionospheric turbulence (Molchanov, 2009) and to the change of the-over-the horizon VHF radio wave propagation in the atmosphere (Devi et al., 2012), perturbations of VLF/LF waves in the lower ionosphere (Rozhnoi et al., 2004, 2005; Hayakawa et al., 2010) and the depression of ULF emissions on the ground (Schekotov et al., 2006). There are some difficulties in the

interpretation of observational results of EQ precursors based on the model of IGW propagation. One of them is as follows. These waves are propagated angularly to the Earth's surface, and the angle increases depending on wave period. So, IGW reaches the ionosphere at a distance of the order of 1000 km from the EQ epicenter, which seems to be in conflict with the localization of the plasma and electromagnetic disturbances in the vicinity of EQ epicenter. Though there have been recently published a few papers suggesting the important role of AGW channel in the LAI coupling (Korepanov et al., 2009; Hayakawa et al., 2011a), in which you can find a summary of recent findings in favor of this hypothesis.

Below we consider the mechanism of the influence of seismic-related electric fields on the ionosphere. The model of LAI coupling was described in Pulinets et al. (2000), which consists of two stages which are not related with each other. They considered the formation of electric field in the near-ground atmosphere due to the appearance of metallic aerosols and ionization source. The source of ionization produces positive and negative ions, and then heavy ions are formed by adhesion of water molecules to the light ions. They calculated the altitude dependence of electric field caused by diffusion, transfer of ions and aerosols by the electric field, gravitational sedimentation of the heavy particles and upward moving of the light particles by the atmosphere convection. The interaction between ions with different signs and their adhesion to the aerosols are taken into account. Their calculations show that during 50 seconds after the turning on the ionization source there is formed an electrode layer up to 30 cm altitude. The number density of the positive and negative ions is different and the electric field is reduced in 1.5 times in this layer. The value of electric field grows in three times above this layer. As they show that this mechanism could be used to explain the electric field variation at fog occurring in the near-ground level. Further, those authors gave a solution of the problem on the electric field penetration into the ionosphere through the conducting atmosphere with exponentially upward increasing conductivity. By imposing the boundary condition on the horizontal distribution of vertical component of electric field on the Earth's surface, they calculated the horizontal component of electric field at the altitude 90 km based on the spatial scale of horizontal distribution of electric field on the Earth's surface. Obviously, the ionosphere conductivity in night time is less than that in day time, so that the electric field at night will be more enhanced than at day. Their calculation shows that even though the radius of disturbed region is 200 km and the field on the ground is 100 V/m, the magnitude of electric field will be 0.07 mV/m. This field is likely to be much smaller than the background field in the ionosphere and consequently it cannot have any effect on the ionosphere. Further those authors conclude that if the field on the ground will be 1000 V/m, then the effect of this field on the ionosphere will be possible. However, such a field in the seismic region with radius 200 km is considered to be implausible. So the above-mentioned work cannot be a basis of LAI coupling model.

Pulinets (2009b) has then made an attempt to explain the possible ionosphere modification due to the atmospheric ionization during the radon injection in the vicinity of active faults. The process of local modification of global electric circuit and the ionosphere variability for tectonic activity is discussed. He supposes that the occurrence of any additional source of ionization has a double effect on the atmosphere conductivity. The appearance of additional ions increases the atmosphere conductivity, while the generation of heavy cluster ions leads to its reduction. However, there is no estimation on the resultant value of conductivity. Further the author supposes that the anomaly of atmosphere conductivity leads to the variation of electric current in the local part of the global circuit, but no calculation of this field has been performed. One should keep in mind that there are theoretical investigations of the atmosphere conductivity modification during the course of ionization. In application to the seismic effect Sorokin et al. (2007) studied in details the processes of conductivity formation during the course of gamma and alpha decay based on the solution of a system of self-consistent nonlinear equations for electric field, atmosphere conductivity, density of ions and aerosols with taken into account their interaction. The well-known value of fair weather current is $\sim 10^{-12}$ A/m$^2$ and atmosphere conductivity is $\sim 10^{-14}$ mho/m, then the field on the ground has a value of ~100 V/m. The value of ionosphere conductivity is $\sim 10^{-6}$ mho/m, so that the field in the ionosphere for the current with the same density has a value of $\sim 10^{-3}$ mV/m. The variation of conductivity in the near-ground atmosphere due to the ionization in two times results in the variation of current density of the same order in the local part of circuit. So that, this additional electric field is on three-four orders less than the ionospheric field and its effect on the ionosphere and equatorial anomaly is negligible. Therefore, the hypothesis suggested is physically not well grounded and it cannot be a candidate for the creation of LAI coupling mode. On this reason the suggestion by Pulinets (2009b) is in contradiction with results obtained by Klimenko et al. (2012), who show that observed disturbances of TEC occur on the assumption that DC electric field reaches (3~9) mV/m in the ionosphere. After all, they use the work by Pulinets (2009b) to interpret the data in spite of the contradiction with obtained results.

A principally alternative physical idea based on the electrodynamic model of plasma and electromagnetic

disturbances accompanying the processes of EQ preparing was developed in Sorokin et al. (2001, 2007) and Sorokin and Chmyrev (2010). First, this model allowed them to explain the results of observation of quasi-static electric field both in the ionosphere and on the Earth's surface in the seismic region, because other models could not explain the nature of such a field. In the frame of this model they found the mechanism for the enhancement of conducting electric current with altitude and the mechanism for limitation of electric field vertical component on the ground surface. The enhancement mechanism is realized by a decrease with altitude of EMF external current at the condition of conservation of the total current. This current is equal to the sum of conductivity and external currents. The external current of EMF is formed in the near ground atmosphere, as seen in Figure 5. In this case even at the growth in conductivity with altitude the field can reach an amplitude of 10 mV/m in the ionosphere. While the conductivity current is appeared by including an additional EMF in the global circuit. The EMF is formed during the injection of charged aerosols by soil gases in the atmosphere and their transfer in the convective atmosphere. The field is limited by a feedback between the external current of EMF and the electric field generated near Earth's surface. Calculations show that the amplitude of disturbed electric field does not exceed their background value on the Earth's surface. In one sense the above-mentioned model is similar to the model of AGW influence to the ionosphere, because the amplitude of AGW grows with altitude by a decrease in atmosphere density. By analogy, the value of conductivity current grows with altitude by a decrease of external current. This implies that the effects are becoming stronger in the ionosphere, but it is difficult to identify AGWs above the background in the near-ground atmosphere. Similarly it is difficult to select the disturbances of conductivity current because their amplitude on the ground does not exceed the background value which is equal to the fair weather current. Both of these effects have a unified source which are lithospheric gases injected into the atmosphere. One can suppose that both AGW and electric current can affect the ionosphere simultaneously, though the consequences of these effects can be different.

According to the electrodynamic model, the growth of electric field in the ionosphere is caused by the EMF formation and the corresponding variation of electro-physical characteristics of lower atmosphere as a result of injection of soil gases, aerosols and radioactive substances during EQ preparation. In the frame of our model, the theory of generation of quasi-static electric field in the atmosphere–ionosphere system was developed, and the methods for calculation of electric field spatial distribution were elaborated. Sorokin et al. (2001, 2005a, 2006a, 2007) and Sorokin and Chmyrev (2010) carried out the theoretical investigation of mechanisms of EMF formation in the lower atmosphere, who have shown that the quasi-static electric field reaches 10 mV/m in the ionosphere while their value is of the order of 100 V/m on the Earth's surface. Moreover, the field can reach a breakdown value in the layer with thin 1-2 km on the altitudes 5-10 km in the troposphere (Sorokin et al., 2011, 2012a, b). Value of the external current of EMF can be approximately $10^{-8}$~$10^{-6}$ A/m$^2$ near the ground. They have further investigated theoretically plasma and electromagnetic effects accompanying the generation of conducting current in the global circuit, and have shown that the appearance of EMF in the global circuit leads to the stimulation of a set of observed plasma and electromagnetic phenomena. An enhancement of the electric field might result in the instability of AGWs in the ionosphere (Sorokin et al., 1998), but the exponential growth of AGW amplitude by the electric field in the ionosphere is limited by vortex formation (Chmyrev & Sorokin, 2010). As a result, the horizontal irregularities of conductivity with scale of approximately 10 km are expected to take place in the E layer of ionosphere. This process is accompanied by field-aligned currents and plasma irregularities stretched along magnetic field lines in the upper ionosphere (Sorokin et al., 1998; Sorokin et al., 2000). Their appearance leads to ULF oscillations (Sorokin et al., 1998) and spectral broadening of VLF transmitter signals (Chmyrev et al., 2008) registered on satellites. The scattering of background electromagnetic emissions by horizontal irregularities of conductivity in the lower ionosphere results in the enhancement of electromagnetic ELF emissions registered on satellites (Borisov et al., 2001) and generation of gyrotropic waves propagated along E layer of the ionosphere. Their propagation forms line spectra of ULF oscillations (Sorokin et al., 2003; Sorokin & Hayakawa, 2008; Sorokin et al., 2009) and the change of resonance frequency of Schuman resonances (Hayakawa et al., 2005, 2011b). Moreover, the appearance of irregularities in the nighttime ionosphere leads to depressions of ULF pulsations of magnetosphere origin (Sorokin et al., 2004; Schekotov et al., 2006; Hayakawa et al., 2013). The growth of electric field up to the breakdown value is caused by random electric discharges, which might generate VHF radio emission in the troposphere over the EQ epicenter (Sorokin et al., 2011, 2012a, b). The generation of conductivity current in the global circuit is accompanied by the modification of ionosphere. Perturbations in the D region of the ionosphere may be generated by both the transfer of charged particles and electron heating (Laptukhov et al., 2009). The electrons are in the upper part of D layer and negative charged ions are in the bottom part of D layer which occurs by quick adhesion of electrons to the neutral molecules. The layer with much density of electrons is appeared in the D region by the transferring charged particles and changing the type of charge carrier by the electric current flowing. The enhancement of

electric current which flows from the atmosphere to the ionosphere results in the growth of plasma density in the E region and sporadic layer formation (Sorokin et al., 2006b). Modification of F region and growth of light ions density are caused by heating release due to the electric current flow in the E region of ionosphere (Sorokin & Chmyrev, 1999). Theoretical investigations of above-mentioned phenomena are accompanied by the calculation of observed parameters and their comparison with experimental data. The lithosphere, atmosphere and ionosphere are an integrated environment in which physical phenomena are related with each other. According to the above-mentioned model, the intensive processes in the lithosphere and atmosphere are the cause of electrodynamic effect onto the ionospheric plasma. Figure 10 illustrates the scheme of processes and registered parameters consisting of electrodynamic model of atmosphere–ionosphere coupling. The left units denote processes stimulated by local current disturbances of global circuit, and the right units denote parameters observed by both satellite and ground-based methods. The injection of charged aerosols by soil gases in the atmosphere forms an additional EMF in the global circuit. Inclusion of the EMF in this circuit leads to the above-mentioned different phenomena.

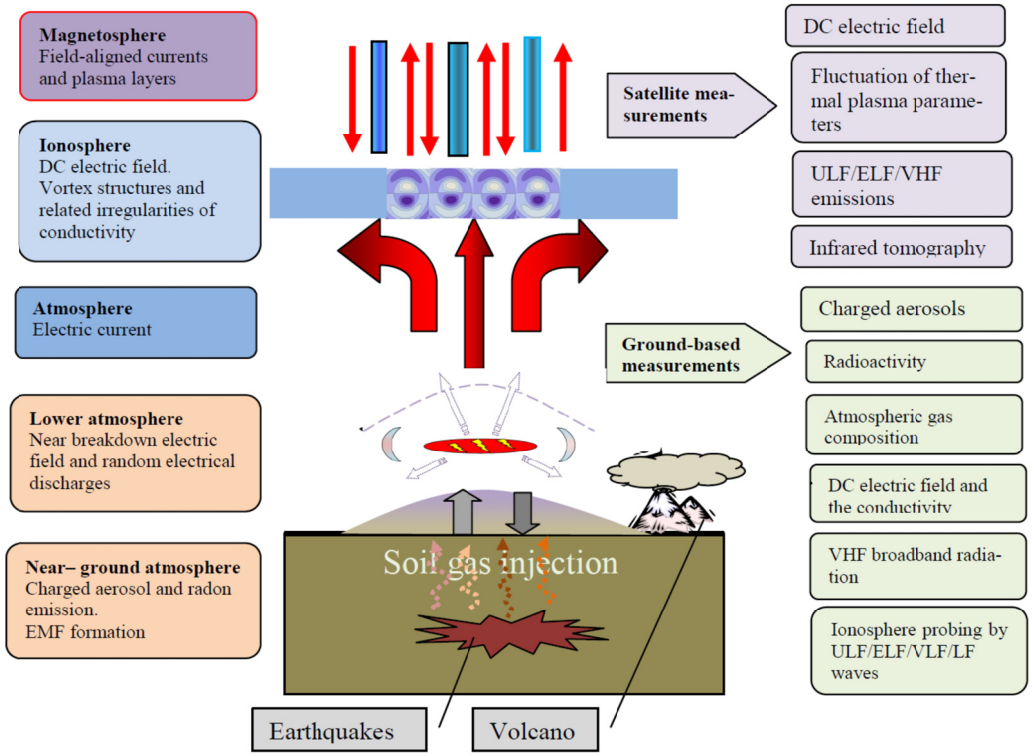



Figure 10. The scheme of electrodynamic model of atmosphere–ionosphere coupling

## 6. Conclusions

Satellite-and ground-based observations of DC electric field over a seismic region show that its amplitude reaches 10 mV/m in the ionosphere, the field reaches a breakdown value in the lower atmosphere, but, at the same time, it does not exceed the background value on the Earth`s surface. These disturbances occupy the region with a horizontal spatial scale from hundreds to thousands km over the seismic region during several days before an EQ. Critical review of existing models has allowed us to rule out the models which are in contradiction with above-mentioned experimental data. It seems that any models based on the assumption that the quasi-static electric field in the ionosphere occurred due to both the variation of atmospheric conductivity by radon injection and transfer of field given on the Earth surface are definitely in contradiction with the experimental data, so that these models cannot be a possible candidate of lithosphere-ionosphere coupling mechanism. Because it is impossible to explain the growth of electric field up to 10 mV/m in the ionosphere area with horizontal scale (100-1000) km and, at the same time, to explain the absence of visible field variation on the Earth's surface. The field penetration models lead to the loss of five orders of the field magnitude in the ionosphere in comparison with experimental data.

This review suggests that there is only one possible way to satisfactorily explain the results of DC electric field

observation in the ionosphere and in the atmosphere over a seismic region and on the ground. A principally different model is based on the assumption that the current source in the circuit connected with pre-EQ processes are situated not in the lithosphere or in the atmosphere, but in the near-ground atmospheric layer. The EQ preparation processes modify the atmosphere in this layer and form an EMF in the seismic zone. The additional source of electric current is generated in the global circuit at the stage of EQ preparing. The range of EMF is formed in the near-earth atmosphere and includes the boundary between the lithosphere and atmosphere. In this case the observable electric field on the surface is located inside the EMF range. Upward transfer of the charged aerosols by atmospheric convection and their gravitational sedimentation result in the EMF formation. Aerosols are injected into the atmosphere by soil gases during an increase in seismic activity. The external current of EMF is reduced with altitude while the current of conductivity increases with altitude, so the total current in the circuit is constant. The value of conductivity current near the surface can be of the order of the value of fair weather current, while the external current exceed the conductivity current by four–five orders. Therefore, conductivity current in the ionosphere is on the order of the external current of EMF near the surface.

The above-mentioned model is similar to the model for AGW influence to the ionosphere, because the amplitude of AGW grows with altitude by decreases of atmosphere density. By analogy, the value of conductivity current increases with altitude by decreases of external current. This implies that the effects become stronger in the ionosphere, but it is difficult to identify any AGW above the background in the near ground atmosphere. So that, it is difficult to identify the disturbances of conductivity current on the ground because their amplitude does not exceed the background value which equals the fair weather current. Both of these effects have a unified source which is injected lithospheric gases in the atmosphere, since one can suppose that both AGW and electric current can affect the ionosphere simultaneously. Lithosphere, atmosphere and ionosphere are an integrated environment in which physical phenomena are related with each other. On the basis of the above-mentioned model the intensive processes in the lithosphere and atmosphere such as EQs, volcanoes, typhoons, thunderstorms are the cause of electrodynamic effect onto the ionospheric plasma. All of these processes are accompanied by numerous electromagnetic and plasma phenomena, which will be discussed elsewhere.

**References**

Afonin, V. V., Molchanov, O. A., Kodama, T., Hayakawa, M., & Akentieva, O. A. (1999). Statistical study of ionospheric plasma response to seismic activity: search for reliable result from satellite observations. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 597-617). Tokyo, Japan: Terra Scientific Publishing Company (TERRAPUB).

Alekseev, V. A., & Alekseeva, N. G. (1992). Investigation of metal transfer in the biosphere during gaseous emission in zones of tectonic activity using methods of nuclear physics. *Nucl. Geophys., 6*, 99-105.

Alperovich, L. S., Gokhberg, M. B., Sorokin, V. M., & Fedorovich, G. V. (1979). On generation of the geomagnetic variation by acoustic oscillations occurring in time of earthquake. *Transaction of USSR Academy of Sciences, Physics of the Earth, 3*, 58-68.

Ampferer, M., Denisenko, V. V., Hausleitner, W., Krauss, S., Stangl, G., Boudjada, M. Y., & Biernat, H. K. (2010). Decrease of the electric field penetration into the ionosphere due to low conductivity at the near ground atmospheric layer. *Annales Geophysicae, 28*, 779-787. http://dx.doi.org/10.5194/angeo-28-779-2010

Biagi, P. F. (2009). Pre and post seismic disturbances revealed on the geochemical data collected at Kamchatka (Russia) during the last 30 years. In M. Hayakawa (Ed.), *Electromagnetic Phenomena Associated with Earthquakes* (pp. 97-117). Trivandrum, India: Transworld Research Network.

Borisov, N., Chmyrev, V., & Rybachek, S. (2001). A new ionospheric mechanism of electromagnetic ELF precursors to earthquakes. *J. Atmos. Solar-Terr. Phys., 63*, 3-10. http://dx.doi.org/10.1016/S1364-6826(00)00153-X

Boskova, J., Smilauer, I., Triska, P., & Kudela, K. (1994). Anomalous behaviour of plasma parameters as observed by the Intercosmos-24 satellite prior to the Iranian earthquake of 20 June 1990. *Studia Geophys. Geodet., 8*, 213-220. http://dx.doi.org/10.1007/BF02295915

Boyarchuk, K. A. (1997). Kinetics of elementary ions in the lower atmosphere acted upon by ionizing radiation. *Proc. Russian Acad. Sci., Atmos. Ocean. Phys., 33*(2), 236-240.

Buchachenko, A. L., Oraevsky, V. N., Pokhotelov, O. A., Strakhov, V. N., Sorokin, V. M., & Chmyrev, V. M. (1996). Ionosphere earthquake precursors. *Usp. Fiz. Nauk, 166*(9), 1023-1029. http://dx.doi.org/10.3367/UFNr.0166.199609g.1023

Chmyrev, V. M., Isaev, N. V., Bilichenko, S. V. & Stanev, G. A. (1989). Observation by space - borne detectors of electric fields and hydromagnetic waves in the ionosphere over on earthquake center. *Phys. Earth Planet. Inter., 57*, 110-114. http://dx.doi.org/10.1016/0031-9201(89)90220-3

Chmyrev, V. M., Isaev, N. V., Serebryakova, O. N., Sorokin, V. M., & Sobolev, Y. P. (1997). Small-scale plasma inhomogeneities and correlated ELF emissions in the ionosphere over an earthquake region. *J. Atmos. Solar-Terrr. Phys., 59*, 967-974.

Chmyrev, V. M., Sorokin, V. M., & Shklyar, D. R. (2008). VLF transmitter signals as a possible tool for detection of seismic effects on the ionosphere. *J. Atmos. Solar - Terr. Phys., 70*, 2053-2060. http://dx.doi.org/10.1016/j.jastp.2008.09.005

Chmyrev, V. M., & Sorokin, V. M. (2010). Generation of internal gravity vortices in the high-latitude ionosphere. *J. Atmos. Solar-Terr. Phys., 72*, 992-996. http://dx.doi.org/10.1016/j.jastp.2010.05.013

Denisenko, V. V., Boudjada, M. Y., Horn, M., Pomozov, E. V., Biernat, H. K., Schwingenschuh, K., … Cristea, E. (2008). Ionospheric conductivity effects on electrostatic field penetration into the ionosphere. *Natural Hazards Earth Syst. Sci., 8*, 1009-1017. http://dx.doi.org/10.5194/nhess-8-1009-2008

Devi, M., Barbara, K., Ruzhin, Y. Y., & Hayakawa, M. (2012). Over-the-horizon anomalous VHF propagation and earthquake precursors. *Survey Geophys., 33*, 1081-1106. http://dx.doi.org/10.1007/s10712-012-9185-z

Fitterman, D. V. (1979). Theory of electrokinetic-magnetic anomalies in a faulted half–space. *J. Geophys. Res., 84*, 6031-6040. http://dx.doi.org/10.1029/JB084iB11p06031

Fraser-Smith, A. C., Bernardi, A., McGill, P. R., Ladd, M. E., Helliwell, R. A., & Villard, O. G. Jr. (1990). Low-frequency magnetic field measurements near the epicenter of the Ms=7.1 Loma Prieta earthquake. *Geophys. Res. Lett., 17*, 1465-1468. http://dx.doi.org/10.1029/GL017i009p01465

Freund, F. (2010). Toward a unified solid state theory for preearthquake signals. *Acta Geophysica, 58*(5), 719-766. http://dx.doi.org/10.2478/s11600-009-0066-x

Freund, F. T., Kulahci, I. G., Cyr, G., Ling, J., Winnick, M., Tregloan-Reed, J., & Freund, M. M. (2009). Air ionization at rock surfaces and pre-earthquake signals. *J. Atmos. Solar-Terr. Phys., 71*, no. 17-18, 1824-1834. http://dx.doi.org/10.1016/j.jastp.2009.07.013

Freund, F. T., Takeuchi, A., & Lau, B. (2006). Electric currents streaming out of stressed igneous rocks-A step towards understanding pre-earthquake low frequency EM emissions. *Phys. Chem. Earth, 31*, 389-396. http://dx.doi.org/10.1016/j.pce.2006.02.027

Fukumoto, Y., Hayakawa, M., & Yasuda, H. (2001). Investigation of over-horizon VHF radio signals associated with earthquakes. *Natural Hazards Earth Syst. Sci., 1*, 107-112. http://dx.doi.org/10.5194/nhess-1-107-2001

Galperin, Yu. I., Gladyshev, V. A., Dzhordzhio, N. V., Larkina, V. I., & Mogilevsky, M. M. (1993). Precipitation of high-energy captured particles in the magnetosphere above the epicenter of an incipient earthquake. *Cosmic Research, 31*, 72-89.

Gladyshev, V., & Fishkova, L. M. (1994). Optical research of seismoactivity. In M. Hayakawa & Y. Fujinawa (Eds.), *Electromagnetic Phenomena Related to Earthquake Prediction* (pp. 35-42). Tokyo, Japan: Terra Scientific Publishing Company, TERRAPUB.

Gokhberg, M. B., Morgunov, V. A., & Pokhotelov, O. A. (1988). *Seismoelectromagnetic Phenomena*. Moscow, Russia: Nauka, p. 173.

Gokhberg, M. B., Nekrasov, A. K., & Shalimov, S. L. (1996). On influence of greenhouse gases instable injection to the ionosphere in seismic active regions. *Transaction of USSR Academy of Sciences, Physics of the Earth, 8*, 52-59.

Gokhberg, M. B., & Shalimov, L. S. (2000). Lithosphere-ionosphere coupling and their modelling. *Russian J. Terrestrial Sciences, 2*(2), 95-108.

Goldberg, R. A. (1984). Middle atmospheric electrodynamics: status and future. *J. Atmos. Terr. Phys., 46*, 1083-1101. http://dx.doi.org/10.1016/0021-9169(84)90010-2

Gorbatikov, A. V., Molchanov, O. A., Hayakawa, M., Uyeda, S., Hattori, K., Nagao, T., … Maltsev, P. (2002). Acoustic emission possibly related to earthquakes, observed at Matsushiro, Japan and its implications. In M. Hayakawa & O. A. Molchanov (Eds.), *Seismo Electromagnetics: Lithosphere-Atmosphere-Ionosphere Coupling* (pp. 1-10). Tokyo, Japan: TERRAPUB.

Gousheva, M., Danov, D., Hristov, P., & Matova, M. (2008). Quasi-static electric fields phenomena in the ionosphere associated with pre- and post earthquake effects. *Natural Hazards Earth Syst. Sci., 8*, 101-107. http://dx.doi.org/10.5194/nhess-8-101-2008

Gousheva, M., Danov, D., Hristov, P., & Matova, M. (2009). Ionospheric quasi-static electric field anomalies during seismic activity August-September 1981. *Natural Hazards Earth Syst. Sci., 9*, 3-15. http://dx.doi.org/10.5194/nhess-9-3-2009

Gousheva, M., Glavcheva, R., Danov, D., Angelov, P., Hristov, P., Kirov, B., & Georgieva, K. (2006). Satellite monitoring of anomalous effects in the ionosphere probably related to strong earthquakes. *Adv. Space Res., 37*, 660-665. http://dx.doi.org/10.1016/j.asr.2004.12.050

Grimalsky, V. V., Hayakawa, M., Ivchenko, V. N., Rapoport, Yu. G., & Zadorozhnii, V. I. (2003). Penetration of an electrostatic field from the lithosphere into the ionosphere and its effect on the D-region before earthquakes. *J. Atmos. Solar-Terr. Phys., 65*, 391-407. http://dx.doi.org/10.1016/S1364-6826(02)00341-3

Hao, J. G. (1989). Near earth surface anomalies of the atmospheric electric field and earthquakes. *Acta Seismol. Sinica, 2*, 289-298.

Hao, J., Tang, T., & Li, D. (2000). Progress in the research of atmospheric electric field anomaly as an index for short-impending prediction of earthquakes. *J. Earthquake Pred. Res., 8*, 241-255.

Harrison, R. G., Aplin, K. L., & Rycroft, M. J. (2010). Atmospheric electricity coupling between earthquake regions and the ionosphere. *J. Atmos. Solar-Terr. Phys., 72*, 376-381. http://dx.doi.org/10.1016/j.jastp.2009.12.004

Hayakawa, M. (Ed.). (2009). *Electromagnetic Phenomena Associated with Earthquakes.* p. 279. Trivandrum, India: Transworld Research Network.

Hayakawa, M. (Ed.). (2012). *The Frontier of Earthquake Prediction Studies*. p. 794. Tokyo, Japan: Nihon-senmontosho-Shuppan.

Hayakawa, M., & Hobara, Y. (2010). Current status of seismo-electromagnetics for short-term earthquake prediction. *Geomatics, Natural Hazards and Risk, 1*(2), 115-155. http://dx.doi.org/10.1080/19475705.2010.486933

Hayakawa, M., Hobara, Y., Ohta, K., Izutsu, J., Nickolaenko, A. P., & Sorokin, V. M. (2011b). Seismogenic effects in the ELF Schumann resonance band. *IEEJ Transactions on Fundamentals and Materials, 131*(9), 684-690. http://dx.doi.org/10.1541/ieejfms.131.684

Hayakawa, M., Kasahara, Y., Nakamura, T., Muto, F., Horie, T., Maekawa, S., …, Molchanov, O. A. (2010). A statistical study on the correlation between lower ionospheric perturbations as seen by subionospheric VLF/LF propagation and earthquakes. *J. Geophys. Res., 115*, A09305. http://dx.doi.org/10.1029/2009JA015143

Hayakawa, M., Kasahara, Y., Nakamura, T., Hobara, Y., Rozhnoi, A., Solovieva, M., …, Korepanov, V. (2011a). Atmospheric gravity waves as a possible candidate for seismo-ionospheric perturbations. *J. Atmos. Electr., 31*(2), 129-140.

Hayakawa, M., Kawate, R., Molchanov, O. A., & Yumoto, K. (1996a). Result of ultra-low-frequency magnetic field measurements during the Guam earthquake of 8 August 1993. *Geophys. Res. Lett., 23*, 241-244. http://dx.doi.org/10.1029/95GL02863

Hayakawa, M., & Molchanov, O. A. (Eds.) (2002). *Seismo-Electromagnetics: Lithosphere-Atmosphere-Ionosphere Coupling.* p. 477. Tokyo, Japan: TERRAPUB.

Hayakawa, M., Molchanov, O. A., & NASDA/UEC team. (2004). Summary report of NASDA's earthquake remote sensing frontier project. *Phys. Chem. Earth, 29*, 617-625. http://dx.doi.org/10.1016/j.pce.2003.08.062

Hayakawa, M, Molchanov, O. A., Ondoh, T., & Kawai, E. (1996b). Anomalies in the sub-ionospheric VLF signals for the 1995 Hyogo-ken Nanbu earthquake. *J. Phys. Earth, 44*, 413-418. http://dx.doi.org/10.4294/jpe1952.44.413

Hayakawa, M., Ohta, K., Nickolaenko, A. P., & Ando, Y. (2005). Anomalous effect in Schumann resonance phenomena observed in Japan, possibly associated with the Chi-chi earthquake in Taiwan. *Ann. Geophysicae, 23*, 1335-1346. http://dx.doi.org/10.5194/angeo-23-1335-2005

Hayakawa, M., Ohta, K., Maekawa, S., Yamauchi, T., Ida, Y., Gotoh, T., …, Nakamura, T. (2006). Electromagnetic precursors to the 2004 Mid Niigata Prefecture earthquake. *Phys. Chem. Earth, 31*, 356-364. http://dx.doi.org/10.1016/j.pce.2006.02.023

Hayakawa, M., Rozhnoi, A., Solovieva, M., Hobara, Y., Ohta, K., Schekotov, A., & Fedorov E. (2013). The lower-ionospheric perturbation as a precursor to the 2011 March 11 Japan earthquake. *Geomatics, Natural Hazards and Risk*, 1-13.

Heincke, J., Koch, U., & Martinelli, G. (1995). $CO_2$ and radon measurements in the Vogtland area (Germany) - a contribution to earthquake prediction research. *Geophys. Res. Lett., 22*, 774-779.

Igarashi, G., Saeki, T., Takahata, N., Sano, Y., Sumikawa, K., Tasaka, S., Sasaki, Y., & Takahashi, M. (1995). Groundwater radon anomaly before the Kobe earthquake. *Science, 269*, 60-61. http://dx.doi.org/10.1126/science.269.5220.60

Kim, V. P., & Hegai, V. V. (1999). A possible presage of strong earthquakes in the night - time mid - latitude F2 region ionosphere. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 619-627). Tokyo, Japan: TERRAPUB.

King, C. Y. (1986). Gas geochemistry applied to earthquake prediction: an overview. *J. Geophys. Res., 91*(B12), 12269-12281. http://dx.doi.org/10.1029/JB091iB12p12269

Klimenko, M. V., Klimenko, V. V., Zakharenkova, I. E., & Pulinets, S. A. (2012). Variations of equatorial electrojet as possible seismo-ionospheric precursor at the occurrence of TEC anomalies before strong earthquake. *Adv. Space Res., 49*, 509-517. http://dx.doi.org/10.1016/j.asr.2011.10.017

Klimenko, M. V., Klimenko, V. V., Zakharenkova, I. E., Pulinets, S. A., Zhao, B., & Tsidilina, M. N. (2011). Formation mechanism of great positive TEC disturbances prior to Wenchuan earthquake on May 12, 2008. *Adv. Space Res., 48*(3), 488-499. http://dx.doi.org/10.1016/j.asr.2011.03.040

Kondo, G. (1968). The variation of the atmospheric electric field at the time of earthquake. *Memoirs of the Kakioka Magnetic Observatory, 13*(1), 11-23.

Kopytenko, Yu. A., Matiashvili, T. G., Voronov, P. M., Kopytenko, E. A., & Molchanov, O. A. (1993). Detection of ultra-low-frequency emissions connected with the Spitak earthquake and its aftershock activity, based on geomagnetic pulsations data at Dusheti and Vardzia observatories. *Phys. Earth Planet. Inter., 77*, 85-95. http://dx.doi.org/10.1016/0031-9201(93)90035-8

Korepanov, V., Hayakawa, M., Yampolski, Y., & Lizunov, G. (2009). AGW as a seismo-ionospheric coupling responsible agent. *Phys. Chem. Earth, 34*(6-7). In M. Hayakawa, J. Y. Liu, K. Hattori, & L. Telesca (Eds.), Special issue, Electromagnetic Phenomena Associated with Earthquakes and Volcanoes, 485-495.

Kuo, C. L., Huba, J. D., Joyce, G., & Lee, L. C. (2011). Ionosphere plasma bubbles and density variations induced by pre-earthquake rock currents and associated surface charges. *J. Geophys. Res., 116*, A10317. http://dx.doi.org/10.1029/2011JA016628

Laptukhov, A. I., Sorokin, V. M., & Yaschenko, A. K. (2009). Disturbance of the ionosphere D region by electric current in the atmosphere–ionosphere electric circuit. *Geomagn. Aeron., 49*, 805-811. http://dx.doi.org/10.1134/S0016793209060103

Larkina, V. I., Migulin, V. V., Molchanov, O. A., Kharkov, I. P., Inchin A. S. & Schvetsova, V. B. (1989). Some statistical results on very low frequency radio wave emissions in the upper ionosphere over earthquake zones. *Phys. Earth Planet. Inter., 57*, 100-109. http://dx.doi.org/10.1016/0031-9201(89)90219-7

Liperovsky, V. A., Meister, C. V., Liperovskaya, E. V., & Bogdanov, V. V. (2008). On the generation of electric field and infrared radiation in aerosol clouds due to radon emanation in the atmosphere before earthquakes. *Natural Hazards Earth Syst. Sci., 8*, 1199-1205. http://dx.doi.org/10.5194/nhess-8-1199-2008

Liperovsky, V. A., Meister, C. V., Liperovskaya, E. V., Davidov, V. F., & Bogdanov, V. V. (2005). On the possible influence of radon and aerosol injection on the atmosphere and ionosphere before earthquakes. *Natural Hazards Earth Syst. Sci., 5*, 783-789. http://dx.doi.org/10.5194/nhess-5-783-2005

Liperovsky, V. A., Meister, C. V., Schlegel, K., & Haldoupis, Ch. (1997). Currents and turbulence in and near mid-latitude sporadic E-layers caused by strong acoustic impulses. *Ann. Geophys., 15*, 767-773.

Liperovsky, V. A., Pokhotelov, O. A., & Shalimov, S. L. (1992). *Ionospheric Earthquake Precursors*. p. 304. Moscow, Russia: Nauka.

Liu, J. Y. (2009). Earthquake precursors observed in the ionospheric F-region, In M. Hayakawa (Ed.), *Electromagnetic Phenomena Associated with Earthquakes* (pp. 187-204). Trivandrum, India: Transworld Research Network.

McGorman, D. R., & Rust, W. D. (1998). *The Electrical Nature of Storms*. Oxford, UK: Oxford Univ. Press.

Molchanov, O. A. (1993). Wave and plasma phenomena inside the ionosphere and the magnetosphere associated with earthquakes. In W. R. Stone (Ed.), *Review of Radio Science 1990-1992* (pp. 591-600). New York, USA: Oxford University Press.

Molchanov, O. A. (1999). Fracturing as an underlying mechanism of seismo-electric signals. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 349-356). Tokyo, Japan: TERRAPUB.

Molchanov, O. A. (2009). Lithosphere-atmosphere-ionosphere coupling due to seismicity. In M. Hayakawa (Ed.), *Electromagnetic Phenomena Associated with Earthquakes* (pp. 255-279). Trivandrum, India: Transworld Research Network.

Molchanov, O. A., & Hayakawa, M. (1995). Generation of ULF electromagnetic emissions by microfracturing. *Geophys. Res. Lett., 22*, 3091-3094. http://dx.doi.org/10.1029/95GL00781

Molchanov, O. A., & Hayakawa, M. (1996). VLF transmitter earthquake precursors influenced by a change in atmospheric electric field. *Proc. 10th International Conference on Atmospheric Electricity*, June 10-14, Osaka, Japan, 428-431.

Molchanov, O. A., & Hayakawa, M. (1998). Subionospheric VLF signal perturbations possibly related to earthquakes. *J. Geophys. Res., 103*, 17489-17504. http://dx.doi.org/10.1029/98JA00999

Molchanov, O. A., & Hayakawa, M. (2008). *Seismo Electromagnetics and Related Phenomena: History and latest results*. p. 189. Tokyo, Japan: TERRAPUB.

Molchanov, O. A., Kopytenko, Yu. A., Voronov, P. M., Kopytenko, E. A., Matiashvili, T. G., Fraser-Smith, A. C., & Bernardy, A. (1992). Results of ULF magnetic field measurements near the epicenters of the Spitak (Ms=6.9) and Loma Prieta (Ms=7.1) earthquakes: Comparative analysis. *Geophys. Res. Lett., 19*, 1495-1498. http://dx.doi.org/10.1029/92GL01152

Molchanov, O. A., Mazhaeva, O. A., Golyavin, A. N., & Hayakawa, M. (1993). Observation by Intercosmos-24 satellite of ELF-VLF electromagnetic emissions associated with earthquakes. *Ann. Geophys., 11*, 431-440.

Molchanov, O. A., Hayakawa, M., & Rafalsky, V. A. (1995). Penetration characteristics of electromagnetic emissions from an uderground seismic source into the atmosphere, ionosphere, and magnetosphere. *J. Geophys. Res., 100*, 1691-1712. http://dx.doi.org/10.1029/94JA02524

Molchanov, O. A., Fedorov, E., Schekotov, A., Gordeev, E., Chebrov, V., Surkov, V., …, Biagi P. F. (2004). Lithosphere-atmosphere-ionosphere coupling as governing mechanism for preseismic short-term events in atmosphere and ionosphere. *Natural Hazards Earth Syst. Sci., 4*, 757-767. http://dx.doi.org/10.5194/nhess-4-757-2004

Namgaladze, A. A., Klimenko, M. V., Klimenko, V. V., & Zakharenkova, I. E., (2009). Physical mechanism and mathematical simulation of ionosphere earthquakeprecursors observed in total electron content. *Geomagn. Aeron., 49*, 252-262. http://dx.doi.org/10.1134/S0016793209020169

Nikiforova, N. N., & Michnowski, S. (1995). Atmospheric electric field anomalies analysis during great Carpatian Earthquake at Polish Observatory Swider. *IUGG XXI General Assembly Abstr.*, Boulder, Colo.: VA11D-16.

Omori, Y., Nagahama, H., Kawada, Y., Yasuoka, Y., Ishikawa, T., Tokonami, S., & Shinogi M. (2008). Preseismic alteration of atmospheric electrical conditions due to anomalous radon emanation. *Phys. Chem. Earth, 33*, 276-284.

Omori, Y., Yasuoka, Y., Nagahama, H., Kawada, Y., Ishikawa, T., Tokonami, S., & Shinagi, M. (2007). Anomalous radon emanation linked to preseismic electromagnetic phenomena. *Natural Hzards Earth Syst. Sci., 7*, 629-635. http://dx.doi.org/10.5194/nhess-7-629-2007

Ouzounov, D., Pulinets, S., Hattori, K., Kafatos, M., & Taylor, P. (2012). Atmospheric signals associated with major earthquakes: A multi-sensor approach. In M. Hayakawa (Ed.), *The Frontier of Earthquake Prediction Studies* (pp. 510-531). Tokyo, Japan: Nihon-senmontosho-Shuppan.

Park, C. G., & Dejnakarintra, M. (1973). Penetration of thundercloud electric fields into the ionosphere and magnetosphere. 1. Middle and auroral latitudes. *J. Geophys. Res., 84*, 960-964. http://dx.doi.org/10.1029/JA084iA03p00960

Parrot, M. (1994). Statistical study of ELF/VLF emissions recorded by low-latitude satellite during seismic events. *J. Geophys.Res., 99*(23), 339-347.

Parrot, M. (2009). Anomalous seismic phenomena: View from space, In M. Hayakawa (Ed.), *Electromagnetic Phenomena Associated with Earthquakes* (pp. 205-233). Trivandrum, India: Transworld Research Network.

Parrot, M. (2011). Statistical analysis of the ion density measured by the satellite DEMETER in relation with the seismic activity. *Earthquake Sci., 24*, 513-548. http://dx.doi.org/10.1007/s11589-011-0813-3

Parrot, M., & Lefeuvre, F. (1985). Correlation between GEOS VLF emissions and earthquakes. *Ann. Geophysicae, 3*, 737-748.

Pilipenko, V. A., Fedorov, E. N., Yagova, N. V., & Yumoto, K. (1999). Attempt to detect ULF electro - magnetic activity preceding earthquake. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 203-214). Tokyo, Japan: Terra Scientific Publishing Company (TERRAPUB).

Pulinets, S. A. (2009a). Lithosphere-atmosphere-ionosphere coupling (LAIC) model, In M. Hayakawa (Ed.), *Electromagnetic Phenomena Associated with Earthquakes* (pp. 235-253). Trivandrum, India: Transworld Research Network.

Pulinets, S. A. (2009b). Physical mechanism of the vertical electric field generation over active tectonic faults. *Adv. Space Res., 44*, 767-773. http://dx.doi.org/10.1016/j.asr.2009.04.038

Pulinets, S. A., Alekseev, V. A., Legenka, A. D., & Khegai, V. V. (1997). Radon and metallic aerosols emanation before strong earthquakes and their role in atmosphere and ionosphere modification. *Adv. Space Res., 20*, 2173-2176. http://dx.doi.org/10.1016/S0273-1177(97)00666-2

Pulinets, S. A., & Boyarchuk, K. A. (2004). *Ionospheric Precursors of Earthquakes.* p. 315. Berlin, Germany: Springer.

Pulinets, S. A., Boyarchuk, K. A., Hegai, V. V., Kim, V. P., & Lomonosov, A. M. (2000). Quasielectrostatical model of atmosphere-thermosphere-ionosphere coupling. *Adv. Space Res., 26*, 1209-1218. http://dx.doi.org/10.1016/S0273-1177(99)01223-5

Pulinets, S. A., Legen'ka, A. D., & Alekseev, V. A. (1994). Pre-earthquake ionospheric effects and their possible mechanisms. In H. Kikuchi (Ed.), *Dusty and Dirty Plasmas, Noise, and Chaos in Space and in the Laboratory* (pp. 545-557). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4615-1829-7_46

Pulinets, S. A., Legen'ka, A. D., Gaivoronskaya, T. V., & Depuev, V. Kh. (2003). Main phenomenological features of ionospheric precursors of strong earthquakes. *J. Atmos. Solar-Terr. Phys*., 65, 1337-1347. http://dx.doi.org/10.1016/j.jastp.2003.07.011

Pulinets, S. A., & Ouzounov, D. (2011). Lithosphere-Atmosphere-Ionosphere Coupling (LAIC) model-An unified concept for earthquake precursors validation. *J. Asian Earth Sci., 41*, 371-382. http://dx.doi.org/10.1016/j.jseaes.2010.03.005

Qiang, Z. J., Dian, C. G., & Li, L. Z. (1999). Satellite thermal infrared precursors of two moderate - strong earthquakes in Japan and impending earthquake prediction. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 747-745). Tokyo, Japan: Terra Scientific Publishing Company (TERRAPUB).

Rakov, V. A., & Uman, M. A. (2002). *Lightning: Physics and Effects.* Cambridge, UK: Cambrdige Univ. Press.

Rapoport, Y., Grimalsky, V., Hayakawa, M., Ivchenko, V., Juarez-R, D., Koshevaya, S., & Gotynyan, O. (2004). Change of ionospheric plasma parameters under the influence of electric field which has lithospheric origin and due to radon emanation. *Phys. Chem. Earth, 29*, 579-587. http://dx.doi.org/10.1016/j.pce.2003.09.018

Rozhnoi, A., Solovieva, M., Molchanov, O. A., & Hayakawa, M. (2004). Middle latitude LF(40kHz) phase variations associated with earthquakes for quiet and disturbed geomagnetic conditions. *Phys. Chem. Earth, 29*, 589-598. http://dx.doi.org/10.1016/j.pce.2003.08.061

Rozhnoi, A. A., Solovieva, M. S., Molchanov, O. A., Hayakawa, M., Maekawa, S., & Biagi, P. F. (2005). Anomalies of LF signal during seismic activity in November–December 2004. *Natural Hazards Earth Syst. Sci., 5*, 657-660. http://dx.doi.org/10.5194/nhess-5-657-2005

Rulenko, O. P. (2000). Operative precursors of earthquakes in the near-ground atmosphere electricity. *Volcanology and Seismology, 4*, 57-68.

Ruzhin, Y., Nomicos, C., & Vallianatos, F. (2000). High frequency seismoprecursor emissions. *Proc. 15th Wroclaw EMC Symposium*, 512-517.

Ruzhin, Yu., & Nomicos, C. (2007). Radio VHF precursors of earthquakes. *Natural Hazards, 40*, 573-583. http://dx.doi.org/10.1007/s11069-006-9021-1

Rycroft, M. J., Israelsson, S., & Price, C. (2000). The global atmospheric electric circuit, solar activity and climate change. *J. Atmos. Solar-Terr. Phys., 62*, 1563-1576. http://dx.doi.org/10.1016/S1364-6826(00)00112-7

Sapkota, B. K., & Varshneya, N. C. (1990). On the global atmospheric electrical circuit. *J. Atmos. Terr. Phys., 52*, l -20. http://dx.doi.org/10.1016/0021-9169(90)90110-9

Schekotov, A., Molchanov, O. A., Hattori, K., Fedorov, E., Gladyshev, V. A., Belyaev, G. G., …, Hayakawa, M. (2006). Seismo-ionospheric depression of the ULF geomagnetic fluctuations at Kamchatka and Japan. *Phys. Chem. Earth, 31*, 313-318. http://dx.doi.org/10.1016/j.pce.2006.02.043

Sorokin, V. M. (2007). Plasma and Electromagnetic Effects in the Ionosphere Related to the Dynamics of Charged Aerosols in the Lower Atmosphere. *Russian J. Physical Chemistry B, 1*(2), 138-170.

Sorokin, V. M., & Chmyrev, V. M. (1999). Modification of the ionosphere by seismic related electric field. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 805 - 818). Tokyo, Japan: Terra Scientific Publishing Company (TERRAPUB).

Sorokin, V. M., & Chmyrev, V. M. (2010). Atmosphere-ionosphere electrodynamic coupling. In V. L. Bychkov et al. (Eds.), *The Atmosphere and Ionosphere: Dynamics, Processes and Monitoring* (pp. 97-146). Berlin, Germany: Springer. http://dx.doi.org/10.1007/978-90-481-3212-6_3

Sorokin, V. M., Chmyrev, V. M., & Hayakawa, M. (2000). The Formation of ionosphere - magnetosphere ducts over the seismic zone. *Planet. Space Sci., 48*, 175-180. http://dx.doi.org/10.1016/S0032-0633(99)00096-3

Sorokin, V. M., Chmyrev, V. M., & Isaev, N. V. (1998). A generation model of mall-scale geomagnetic field-aligned plasma inhomogeneities in the ionosphere. *J. Atmos. Solar-Terr. Phys., 60*, 1331-342. http://dx.doi.org/10.1016/S1364-6826(98)00078-9

Sorokin, V. M., Chmyrev, V. M., & Yaschenko, A. K. (2001). Electrodynamic model of the lower atmosphere and the ionosphere coupling. *J. Atmos. Solar-Terr. Phys., 63*, 1681-1691. http://dx.doi.org/10.1016/S1364-6826(01)00047-5

Sorokin, V. M., Chmyrev, V. M., & Yaschenko, A. K. (2003). Ionospheric generation mechanism of geomagnetic pulsations observed on the Earth's surface before earthquake. *J. Atmos. Solar-Terr. Phys., 64*, 21-29. http://dx.doi.org/10.1016/S1364-6826(02)00082-2

Sorokin, V. M., Chmyrev, V. M., & Yaschenko, A. K. (2005a). Theoretical model of DC electric field formation in the ionosphere stimulated by seismic activity. *J. Atmos. Solar-Terr. Phys*., 67, 1259-1268. http://dx.doi.org/10.1016/j.jastp.2005.07.013

Sorokin, V. M., Chmyrev, V. M., & Yaschenko, A. K. (2006a). Possible DC electric field in the ionosphere related to seismicity. *Adv. Space Res., 37*, 666-670. http://dx.doi.org/10.1016/j.asr.2005.05.066

Sorokin, V. M., Fedorov, E. N., Schekotov, A. Yu., Molchanov, O. A., & Hayakawa, M. (2004). Depression of the ULF geomagnetic pulsation related to ionospheric irregularities. *Ann. Geophysics, 47*, 191-198.

Sorokin, V. M., & Hayakawa, M. (2008). On the generation of narrow-banded ULF/ELF pulsations in the lower ionospheric conducting layer. *J. Geophys. Res., 113*, a06306. http://dx.doi.org/10.1029/2008JA013094

Sorokin, V. M., & Pokhotelov, O. A. (2010). Generation of ULF geomagnetic pulsations during early stage of earthquake preparation. *J. Atmos. Solar -Terr. Phys., 72*, 763-766. http://dx.doi.org/10.1016/j.jastp.2010.03.021

Sorokin, V. M., Ruzhin, Yu. Ya., Kuznetsov, V. D., & Yaschenko, A. K. (2012a). Model of electric discharges formation in the lower atmosphere over a seismic region. *Geomatics, Natural Hazards and Risk, 3*, 225-238. http://dx.doi.org/10.1080/19475705.2011.604799

Sorokin, V. M., Ruzhin, Yu. Ya., Yaschenko, A. K., & Hayakawa, M. (2011a). Generation of VHF radio emissions by electric discharges in the lower atmosphere over a seismic region. *J. Atmos. Solar-Terr. Phys., 73*, 664-670. http://dx.doi.org/10.1016/j.jastp.2011.01.016

Sorokin, V. M., Ruzhin, Yu. Ya., Yaschenko, A. K., & Hayakawa, M. (2012b). Seismic - related electric discharges in the lower atmosphere. In M. Hayakawa (Ed.), *The Frontier of Earthquake Prediction Studies* (pp. 592-611). Tokyo, Japan: Nihon-senmontosho-Shuppan.

Sorokin, V. M., Sergeev, I. Yu., & Pokhotelov, O. A. (2009). Low latitude gyrotropic waves in a finite thickness ionospheric conducting layer. *J. Atmos. Solar-Terr. Phys., 71*, 175-179. http://dx.doi.org/10.1016/j.jastp.2008.10.001

Sorokin, V. M., Yaschenko, A. K., Chmyrev, V. M., & Hayakawa, M. (2005b). DC electric field amplification in the mid-latitude ionosphere over seismically active faults. *Natural Hazards Earth Syst. Sci., 5*, 661-666. http://dx.doi.org/10.5194/nhess-5-661-2005

Sorokin, V. M., Yaschenko, A. K., & Hayakawa, M. (2006b). Formation mechanism of the lower ionosphere disturbances by the atmosphere electric current over a seismic region. *J. Atmos. Solar-Terr. Phys., 68*(11), 1260-1268. http://dx.doi.org/10.1016/j.jastp.2006.03.005

Sorokin, V. M., Yaschenko, A. K., & Hayakawa, M. (2007). A perturbation of DC electric field caused by light ion adhesion to aerosols during the growth in seismic-related atmospheric radioactivity. *Natural Hazards Earth Syst. Sci., 7*, 155-163. http://dx.doi.org/10.5194/nhess-7-155-2007

Sorokin, V. M., & Yaschenko, A. K. (2000). Electric field disturbance in the earth-ionosphere layer. *Adv. Space Res., 26*(8), 1219-1223. http://dx.doi.org/10.1016/S0273-1177(99)01221-1

Surkov, V., & Pilipenko, V. (1999). The physics of pre-seismic electromagnetic ULF signals. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 357-370). Tokyo, Japan: Terra Scientific Publishing Company (TERRAPUB).

Tronin, A. A. (1999). Satellite thermal survey application for earthquake prediction. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 717-746). Tokyo, Japan: Terra Scientific Publishing Company (TERRAPUB).

Tronin, A. A. (2006). Remote sensing and earthquakes: A review. *Phys. Chem. Earth, 31*, 138-142. http://dx.doi.org/10.1016/j.pce.2006.02.024

Tronin, A. A., Hayakawa, M., & Molchanov, O. A. (2002). Thermal IR satellite data application for earthquake research in Japan and China. *J. Geodynamics, 33*, 519-534. http://dx.doi.org/10.1016/S0264-3707(02)00013-3

Uyeda, S., Nagao, T., & Kamogawa, M. (2009). Short-term earthquake prediction: current state of seismo-electromagnetics, *Tectonophys., 470*, 205-213. http://dx.doi.org/10.1016/j.tecto.2008.07.019

Vallianatos, F., & Nomicos, K. (1998). Seismogenic radioemissions as earthquake precursors in Greece. *Phys. Chem. Earth, 23*(9-10), 953-957. http://dx.doi.org/10.1016/S0079-1946(98)00126-8

Varotsos, P. (2001). A review and analysis of electromagnetic precursory phenomena. *Acta Geophysica Polonica, 49*(1), 1-42.

Vershinin, E. F., Buzevich, A. V., Yumoto, K., Saita, K., & Tanaka, Y. (1999). Correlations of seismic activity with electromagnetic emissions and variations in Kamchatka region. In M. Hayakawa (Ed.), *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (pp. 513-517). Tokyo, Japan: Terra Scientific Publishing Company (TERRAPUB).

Virk, H. S., & Singh, B. (1994). Radon recording of Uttarkashi earthquake. *Geophys. Res. Lett., 21*, 737-741. http://dx.doi.org/10.1029/94GL00310

Voitov, G. I., & Dobrovolsky, I. P. (1994). Chemical and isotope - carbonic instability of the soil gases in the seismic regions. *Izvestiya AN SSSR, Fizika Zemli, 3*, 20-27.

Yasuda, Y., Ida, Y., Goto, T., & Hayakawa, M. (2009). Interferometric direction finding of over-horizon VHF transmitter signals and natural VHF radio emissions possibly associated with earthquakes. *Radio Sci., 44*, RS2009. http://dx.doi.org/10.1029/2008RS003884

Yasuoka, Y., Igarashi, G., Ishikawa, T., Tokonami, S., & Shinogi, M. (2006). Evidence of precursor phenomena in the Kobe earthquake obtained from atmospheric radon concentration. *Appl. Geochem., 21*, 1064-1072. http://dx.doi.org/10.1016/j.apgeochem.2006.02.019

Yasuoka, Y., Ishikawa, T., Nagahama, H., Kawada, Y., Omori, Y., Tokonami, S., & Shinogi, M. (2012). Radon anomalies prior to earthquakes. In M. Hayakawa (Ed.), *The Frontier of Earthquake Prediction Studies* (pp. 410-427). Tokyo, Japan: Nihon-senmontosho-Shuppan.

Yonaiguchi, N., Ida, Y., Hayakawa, M., & Masuda, S. (2007a). A comparison of different fractal analyses for VHF electromagnetic emissions and their self-organization for the off-sea Miyagi-prefecture earthquake. *Natural Hazards Earth System Sci., 7*, 485-493. http://dx.doi.org/10.5194/nhess-7-485-2007

Yonaiguchi, N., Ida, Y., Hayakawa, M., & Masuda, S. (2007b). Fractal analysis for VHF electromagnetic noises and the identification of preseismic signature of an earthquake. *J. Atmos. Solar-terr. Phys., 69*, 1825-1832. http://dx.doi.org/10.1016/j.jastp.2007.08.002

Zakharenkova, I. E., Shagimuratov, I. I., Tepenitzina, N. Yu., & Krankowski, A. (2008). Anomalous modification of the ionospheric Total Electron Content prior to the 26 September 2005 Peru earthquake. *J. Atmos. Solar-Terr. Phys., 70*, 1919-1928. http://dx.doi.org/10.1016/j.jastp.2008.06.003

Zolotov, O. V., Namgaladze, A. A., Zakharenkova, I. E., Shagimuratov, I. I., & Martynenko, O. V. (2008). Simulations of the equatorial ionosphere to the seismic electric field sources. *Proc. 7th International Conference "Problems of Geocosmos",* St. Petersburg, Russia: 26-30 May 2008, 492-496.

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# Nature of Solar Radiation as Encouraged to Produce an Increment of Dissolved Oxygen and Hydrogen Peroxide in Oxidation Ponds for Community Wastewater Treatment at H.M.The King's LERD Project Site in Phetchaburi Province, Thailand

Thanit Pattamapitoon[1], Pramote Sirirote[2], Pannee Pakkong[3] & Kasem Chunkao[1]

[1] Department of Environmental Science, Faculty of Environment, Kasetsart University, Bangkok, Thailand

[2] Department of Microbiology, Faculty of Science, Kasetsart University, Bangkok, Thailand

[3] Department of Applied Radiation and Isotopes, Faculty of Science, Kasetsart University, Bangkok, Thailand

Correspondence: Thanit Pattamapitoon, Department of Environmental Science, Faculty of Environment, Kasetsart University, Bangkok 10900, Thailand. E-mail: thanit_lerd@hotmail.com

Received: March 14, 2013 Accepted: April 23, 2013 Online Published: May 13, 2013

doi:10.5539/mas.v7n6p26 URL: http://dx.doi.org/10.5539/mas.v7n6p26

**Abstract**

H.M.The King's initiative nature by nature process for community wastewater treatment has been conducted since 1990 at Laem Phak Bia sub district, Ban Laem district, Phetchaburi province, Thailand on UTM 1442240 to 1443480 N and 0619271 to 0619271 E. The fresh food markets and households were the point sources of Phetchaburi municipal wastewater that flowing newly construction sewage system in order to receive them to four sub stations on both sides of Phetchaburi river before pumping to Klongyang collection pond. The storage wastewater has been pumped about 3,600 cu.m./day through the 18.5 km. HPDE pipe and putting into sedimentation pond 1 and another four consecutive ponds (oxidation ponds 2, 3, 4, and 5) before the effluent flowing into the mangrove forest. However, the results of experiment found the values of BOD that showing drastic decreasing from Klongyang collection pond through the tip of 18.5 km. HPDE pipe to the sedimentation pond 1 then after it was gradually decreased from one oxidation pond to the next one. Highlight was placed on the obtained effluent after treated wastewater flowed over weir crest that finding BOD under standard and also the decreasing of total coliform bacteria and fecal coliform bacteria, particularly the pathogenic bacteria decreasing down to almost zero MPN/100mL. The question came up how total coliform bacteria and fecal coliform bacteria disappeared after the treated wastewater flowing over weir crest of the oxidation pond 3. The hypothesis was set on the effect of solar radiation to DO and $H_2O_2$ which were employed for bacterial organic digestion process in wastewater treatment ponds by taking the measurement randomly on 19 May 2011 that solar radiation showing the solar energy between 0-750 W/m$^2$ all together with UV-A, UV-B, spectrum and net radiation.

In order to accomplish the objectives of study, the wastewater quality indicators were found BOD in sedimentation pond 1 67.5 mg/L and oxidation pond 3 16.6 mg/L for 13 hours measurement. In the same manner, the values of DO and $H_2O_2$ for 13 hours measurement found in oxidation pond 3 greater than sedimentation pond 1 in respect to 8.9 and 7.6 mg/L for DO and 2.27 and 0.31 µg/L for $H_2O_2$. Moreover, the polynomial correlation was studied in order to determine the relationship between net radiation and DO and found very high correlation coefficients in both sedimentation pond 1 and oxidation pond 3. For the quantity of $H_2O_2$, it was obtained the relationship between net radiation and $H_2O_2$ in higher correlation coefficient in sedimentation pond 1 than oxidation pond 3 because of smaller amount of dissolved organic and inorganic matters in oxidation pond 3. It would be the reason that oxygen which occurred from phytoplankton and algae photosynthesis were not employed by organic digestion and respiration of bacteria. Therefore, it had to exist DO and accumulation only in water instead of producing $H_2O_2$ in oxidation pond 3 during daylight time. This is why $H_2O_2$ in oxidation pond 3 was not varied directly to net radiation, and also causing an indication on the relationship between net radiation and $H_2O_2$ in oxidation pond 3 lower than sedimentation pond 1. Summarily speaking, the nature of solar radiation plays vital role in encouraging producing DO and $H_2O_2$ which are the important factors in bacterial organic digestion for community wastewater treatment with high efficiency.

**Keywords:** solar radiation, dissolved oxygen, hydrogen peroxide, oxidation pond, wastewater

## 1. Introduction

Accordance to H.M.The King has been royally initiative the nature by nature process for community wastewater treatment by utilizing oxidation ponds (as similar as lagoon and stabilization ponds) for obtaining free oxygen to sustain the bacterial organic digestion processing under supporting from three nature processes, they are photosynthesis, thermo siphon and thermo osmosis processes. Solar radiation plays the significant role in fulfillment of such three nature processes in order to produce DO by diffusing oxygen in vertical direction from the air to company with cooler wastewater on the surface because of using heat for evaporation (latent heat of evaporation approximately 583 gm-cal/gm-water) due to the thermo siphon process, while phytoplankton produces free dissolved oxygen in wastewater due to photosynthesis of phytoplankton and another floating aquatic plants and also due to thermo osmosis process by means on how oxygen in spongy cells (aerenchyma cells) of very young leaves during photosynthesis of phytoplankton and surrounding aquatic plants to produce oxygen with higher pressure enough to be forced passing through softer membrane to plant tissues and then to the root system (rhizosphere), finally contributing by diffusion process to bacteria in wastewater and soils as the energy sources for organic digestion. Consequently, the obtainable oxygen becomes to supply to bacteria for organic digestion in the said wastewater in order to receive better water quality after driving away of organic matters by oxidation pond wastewater treatment system. It's process includes five stages: firstly, wastewater product from fresh food markets, households, shopping areas, dessert factories, schools, government offices and cultural places; secondly, screening bigger size of organic wastes as well as skimming of oil and grease and also grit chamber for sand and gravel; thirdly, treatment of wastewater by wastewater treatment pond system (five ponds: sedimentation pond 1, oxidation pond 2, 3, 4 and 5) under the processes of bacterial digesting organic matters that are converted to become the inorganic materials as nutrients for the growth of phytoplankton and another types of aquatic plants in wastewater and its surroundings, following by keeping their photosynthesis processes in turn with producing DO as energy supply to bacteria for maintaining the endless organic digesting processes; fourthly, sedimentation of sludge during all time processing and fifth, effluent flow out to the public water sources by gravitational forces. In reality, effluent can be at more or less standards of water surface but it is very often to contaminate some diseases, only if they have to be killed by solar radiation, especially ultraviolet which is the most powerful energy. The phenomena like this normally occurs in the tropical zone with availability of solar energy in which Thailand is characterized for serving the above statement. With previous statement, H.M.The King has pushed an effort to initiate the project on how to treat the community wastewater under nature by nature processes in order to eliminate the contaminated organic wastes, toxic chemicals, oil and grease, color, bad smell gases and diseases which are the most dangerous microorganisms in community wastewater. Actually, the previous research did not find out the *Escherichia coli* (*E.coli*) in treated wastewater which was supposed to be killed in the effluent after flowing over the weir crest of third pond with retention time of 21 days, but it is still in doubt. However, it has been hypothesized that ultraviolet should be the most possible threads rather than something else.

Fundamentally, it is understood among scientists that solar radiation as the bio-conversion is presumable to stimulant plant in producing the $H_2O_2$ by blue green algae (Cyanobacteria) under the photoautotrophic conditions should be special condition of its involvement for decreasing some bacteria species during bacterial organic digesting process one way and another. Naturally, incoming solar radiation to the earth surface is comprised of shortwave (wavelength less than 4 μm) and longwave (wavelength longer 4 μm) which are quantified as about 45% and 55%, respectively but they are depend on the sky conditions of the days and seasons of the years as the same as the location on the earth. Theoretically speaking, higher frequency is provided more heat due to more number of wavelength, that is why higher heat can be obtainable from x-rays and gramma rays (wavelength less 0.1 μm) and being gradual decreased consecutively ultraviolet (wavelength 0.1 to 0.38 μm), visible light (0.38 to 0.68 μm), near infrared (0.68 to 1.0 μm), far infrared (1.0 to 4.0 μm) and longwaves down to wavelengths of TV and radio as shown in Figure 1. Among those wavelengths, the visible light is normally called as shortwave down radiation which is about 45% of total radiation and composing of violet, indigo, blue, green, yellow, orange and red (VIBGAYOR) while ultraviolet (UV) is absorbed by ozone gas in the sky about 85% and the other 15% shortwave down radiation that functioning to control diseases not only contamination in air but also in wastewater for human beings.

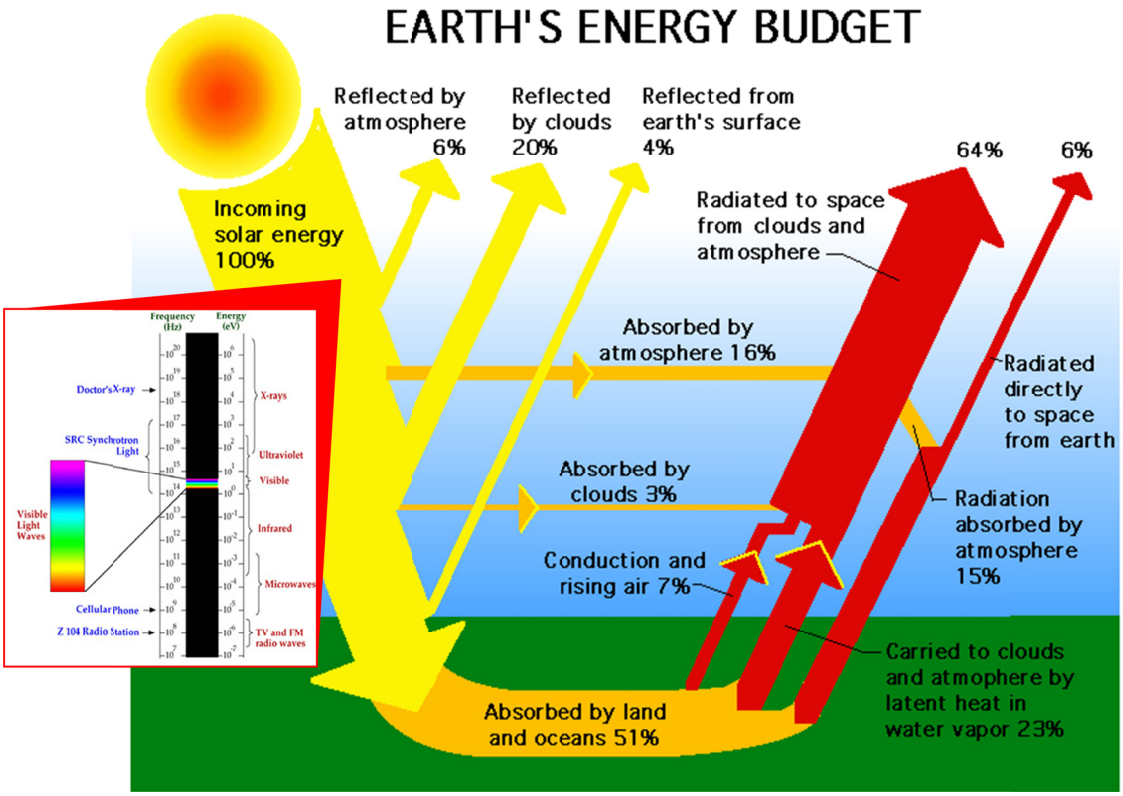



Figure 1. Hypothetical solar energy on the earth surface including UV, visible light and incoming and outgoing radiation (Source: Image courtesy NASA's ERBE (Earth Radiation Budget Experiment) program)

According to the oxidation pond system for wastewater treatment is relied on natural treating that wastewater begins to flow into the system in turn with screening of discarded food, skimming and grit chamber for settling sand and gravel by gravitational forces. Then, after the the treatment is started up by biological process with the support of solar radiation as the main energy source for stimulating on biochemical processes of living cells in order to cause an effect on aerobic process of organic digestion. In consequence, it makes the lowering of organic matters and dissolved oxygen (William et al., 1852; Tapas, 2001; Macos, 2005). However, when wastewater was accomplished the treatment process and flew to the polishing pond, the treated wastewater indicators chemical oxygen demand (COD), biological oxygen demand (BOD), total dissolved solid (TDS) and total coliform bacteria (TCB) were lowered down to the standard of effluent as set by the Ministry of Natural Resource and Environment. In contrary, the value of DO was additionally increased because of the photosynthesis processes of algae and phytoplankton together with thermo osmosis and thermo siphon processes and also wind blow (Richard et al., 1995; Donat et al., 1997; Leena et al., 2011; Chunkao et al., 2012).

Lowering and adding of DO will be influenced to free radical occurrence in wastewater (Volodymyr, 2011) which causes the break down of oxygen molecules by solar radiation at the wavelength of visible light, UV-A and UV-B which are the catalyst of chemical reaction (John et al., 1998; Gerringa et al., 2004; Steven et al., 2010). In consequence, the outcome obtained so many free radical forms such as hydrogen peroxide ($H_2O_2$), hydroxyl radicals ($OH^*$) and superoxide anion ($O_2^*$) but they were found out more potentials and severely specific oxidizing property and named as $H_2O_2$ (Robert & Gambini, 1990) by means of $H_2O_2$ that being oxidized in both the organic and inorganic substances which are contaminated in wastewater will be transformed into COD, BOD, TDS and bad smell (Alenka et al., 2006; Mohamed, 2006; Roberto et al., 2009; Zulfiqar et al., 2011). It is the highlight of this transformation of organic and inorganic substances that results the decrease of humic substances in wastewater which are the major organic constituents to be able to make the water in black color (Gen et al., 2001). Besides, $H_2O_2$ is competent to kill Cyanobacteria and to devastate microcystin structure which is the toxic substance and produced by Cyanobacteria (Rui et al., 2005; Dani et al., 2012), just like *E.coli*, which is pathogenic bacteria to have severe diarrhea and important indicator for community wastewater quality, has to be killed one way or another (Davies et al., 1997; Asad et al., 1998) by running into cell in order to damage to mitochondria and DNA of *E.coli*

causing death in consequence (Yoel et al., 1987; Richard et al., 2003). Moreover, the concentration of $H_2O_2$ between 0.2-0.3 mM. can be incompetent to suppress the disease occurrence of *E.coli* (Hegde et al., 2008) but $H_2O_2$ cannot do any damage to zooplankton, small and big marine fishes (Dani et al., 2012; Hans et al., 2012). It could be presumed that the above statement pointed out the shortwave length of visible light, particularly UV, is the catalyst to stimulate the chemical reaction which might produce some quantity of DO and $H_2O_2$ to decrease not only organic and inorganic substances but they are also competent to kill pathogenic bacteria and *E.coli* in Phetchaburi municipal wastewater. Therefore, it is very necessary to study on the role of solar radiation in relation to produce the quantity of DO and $H_2O_2$ in H.M.The King's initiative nature by nature process for Phetchaburi municipal wastewater treatment system in order to know the involved factors and their mechanisms to decrease COD, BOD, TDS and pathogenic bacteria to be under standard of treated wastewater before draining to the public water sources.

## 2. Location of the Royal LERD Project Site

H.M.The King's initiative nature by nature process on Laem Phak Bia Environmental Research and Development project (Royal LERD Project) at Laem Phak Bia sub detract, Ban Laem district, Phetchaburi province, Thailand between latitude 130* 02'40" to 130* 03'20'N and longitude 1000* 05'10" to 1000* 05'10"E, or UTM at 1442240 N to 1443480 N and 0617780 to 0619271 E, approximately 122 km. from Bangkok to the south (Figure 2). Actually, the project site is far from the city of Phetchaburi (Phetchaburi municipal) about 12 km., which is composed of local people 40,000 persons plus tourists and illegal workers about 10,000 persons/day. The coverage area is about 260 hectares which localizes inside the natural mangrove forest as laid down from Phetchaburi province to Bangkok (Phetchaburi-Bangkok mangrove forest) with the length of about 140 km. and the width (seashore to inland side of mangrove forest about 2 km.). Geographical location, the east side adjoins to the Gulf of Thailand, the north to Phetchaburi-Bangkok mangrove forest plus mud beach, the south to the mangrove forest patch nearby well known Hat Chao Sumran sand beach, and the west close to Wat Samutkodom (Buddhist temple) as belonged to Ban Phanern village.

In fact, the study areas are composed of not only the Royal LERD project site at Laem Phak Bia but also including the whole area of Phetchaburi municipal as point sources of community wastewater producing approximately 7,000 cu.m./day. However, the modification of sewage drainage system in town was taken care for holding back wastewater instead of direct flowing to Phetchaburi river but turning back by lifting up the pipe ends to the collection culverts to the four sub pumping stations then pumping them to the main collection pumping station at Klongyang village. For implementation, Phetchaburi municipal wastewater at Klongyang collection wastewater station has to pump continuously through 18.5 km HPDE pipe to the Royal LERD project site about 3,600 cu.m./day to treat at the first pond (sedimentation pond 1), to the second pond (oxidation pond 2), the third pond (oxidation pond 3), the fourth pond (oxidation pond 4) and the fifth pond (oxidation pond 5) for community wastewater treatment as illustrated in Figure 2. In addition, all 5 ponds wastewater treatment system is laid down in the open air without interruption of sunshine and blow in wind in order to promote photosynthesis of algae and pond evaporation for producing an effect of both the thermo siphon and thermo osmosis processes to add up free oxygen in wastewater for bacterial organic digestion process

## 3. Methods and Procedure

### *3.1 Solar Radiation Measurement*

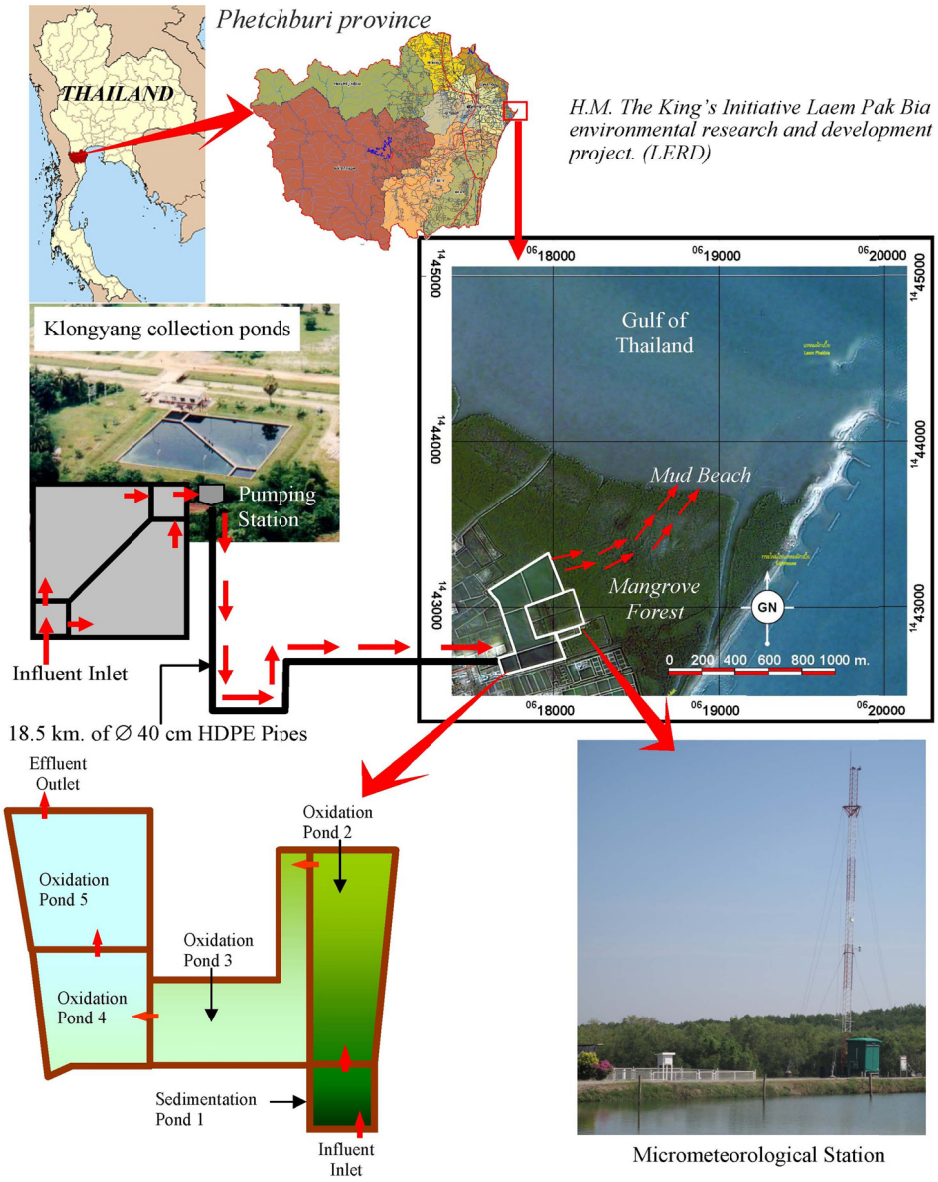
The 40 m height of micrometeorological tower has been settled down at the east conner of the Royal LERD project site (Figure 2) for measuring the incoming shortwave radiation (Rsd), outgoing short wave radiation (Rsu), incoming long wave radiation (Rld), outgoing long wave radiation (Rlu), net radiation (Rn), ultraviolet, spectrum (VIBGYOR), near infrared and also taking the Royal LERD climatic station to measure the daily rainfall, air temperature, relative humidity, pan evaporation, wind speed and direction and sunshine period. Data collection will be analyzed in daily basis in order to serve needs of any experiment that concerning with H.M.The King's initiative nature by nature process on community wastewater treatment and garbage disposal.

### *3.2 Wastewater Quality Sampling*

According to investigate the influences of net radiation to add up the DO and $H_2O_2$ in wastewater, therefore the blue sky day of 19 May 2011 will be selected as the representative for searching the relationship between total radiation to the occurrences of DO and $H_2O_2$ in wastewater. In the same manner, the sampling points have been fixed at the middle of sedimentation pond 1 and oxidation pond 3. Water samples were taken at 30 cm. depth by PE bottles every one hour during 06:00 to 19:00 o'clock.

### *3.3 Wastewater Quality Analysis*

An application of APHA AWWA and WPCF (1995) was conducted to determine pH, temperature, DO, BOD and total bacteria count. In addition, $H_2O_2$ concentration was determined by titration methods (Solvay, 2004). The followings are details of chemical analysis: thermometer for water temperature; pH meter for pH; azide modification for DO; azide modification 5 day 20 °C for BOD and dilution plate count for total bacteria.




| | Sedimentation Pond 1 | Oxidation Pond 2 | Oxidation Pond 3 | Oxidation Pond 4 | Oxidation Pond 5 | Total |
|---|---|---|---|---|---|---|
| Dept. (m.) | 2.3 | 2 | 1.9 | 1.8 | 1.7 | - |
| Area. (sq. m.) | 10,217.00 | 20,408.00 | 24,898.00 | 35,424.00 | 43,131.50 | 154,178.50 |
| Capacity (cu.m.) | 23,499.10 | 60,816.00 | 66,306.20 | 63,763.20 | 73,323.50 | 287,708.00 |

Figure 2. Location of the Royal LERD project site at Laem Phak Bia sub district, Ban Laem district, Phetchaburi province, including Phetchaburi municipal wastewater, Klongyang collection pond, pumping station, 18.5 km. HPDE pipe, five pond for wastewater treatment, micrometeorological tower for solar radiation measurement and climatic station

## 4. Results and Discussion

Due to the objective is to study on nature of solar radiation as the stimulant for producing the increment of DO and $H_2O_2$ in community (Phetchaburi municipal) wastewater on which they are referred to kill some pathogenic bacteria during sunshine period of the day. The analyzed results will be presented as follows:

### *4.1 Quantitative Phetchaburi Municipal Wastewater*

Before the year of 1990, the sewage was directly drained into the Phetchaburi river that causing stream water in unpleasant color and smell. Water in Phetchaburi river was constraint to use for day to day lifestyle, waterworks, aquatic lives and maintaining the local cultural activities. Therefore, the Phetchaburi river water was really needed to take in conservation measures in order to recover for serving any purposes as it used to be. As stated in the previous sections, the Phetchaburi municipal has been composed of fully growth of population approximately 40,000 persons plus another 10,000 tourists, which produces community wastewater about 7,000 cu.m./day. This amount of community wastewater has been forced by redesigning sewage drainage system for turning back to main culverts that being parallel along the riverbanks before pumping to Klongyang collection pond. Only 3,600 cu.m./day can be pumped from Klongyang collection pond under the anaerobic digestion processes flowing through the 18.5 km. HPDE pipe in order to transferring to the designing treatment capacity of 10,000 cu.m./day by starting up on the sedimentation pond 1 and another 4 consecutive ponds as called oxidation pond system. In order to maintain the aquatic ecological balancing in wastewater treatment pond system, the three herbivore fishes/sq.m. have been allowed into every treatment pond for controlling the blooming of phytoplankton and another algae species due to higher concentration of nitrogen and phosphorus as the products of bacterial organic digestion processes in wastewater treatment pond system.

### *4.2 Wastewater Quality in Treatment Pond System*

The previous investigation of the Royal LERD project found that fresh food market and households were the main point sources of Phetchaburi municipal wastewater flowing directly to the river of Phetchaburi and producing organic wastes in form of BOD more than 500 mg/L (COD exceeding 1,200 mg/L) and rapidly decreased down more or less 200 mg/L (ranging between 150 to 300 mg/L, depending on season of the year) in the drainage system, and next to the collection pond at Klongyang pumping station, at the sedimentation pond 1 after flowing through about 80 mg/L (ranging between 50 to 120 mg/L), and finally effluent at the oxidation pond 5 less 10 mg/L (some occasion almost attaching to the effluent standard 20 mg/L). In addition, the Royal LERD research result was shown that the wastewater in sedimentation pond 1 and oxidation ponds were found high density of the water born diseases in group of *E.coli* but very low number. The total perspective of decreasing organic matters according to aerobic and anaerobic processes from fresh food markets and households as the main point sources of Phetchaburi municipal through culverts plus pipes sewage systems, Klongyang collection pond, 18.5 km. pipes, 5 consecutive wastewater treatment ponds, natural mangrove forest as the second wastewater treatment units and Bangkok-Phetchaburi mud beach in part of Phetchaburi province has been illustrated in Table 1.

The drastic decrease of organic wastes in wastewater in terms of BOD and some other indicators are shown in Table 1 due to the fact that the longer degradation period of dissolved organic matters (mostly DOC, DOP, DON and DOS) were completely separable from the solid wastes under both aerobic and anaerobic processes in hot and humid tropical climate of Phetchaburi municipal, particularly from point sources at fresh food markets and households from 546.6 mg/L to 162.7 mg/L (e.g. BOD 70% decreasing) through long and 90 percentage closed sewage drainage system the prior wastewater receiver of Klongyang collection pond from 162.7 mg/L to 79.5 mg/L (e.g. BOD 50% decreasing) and then after pumping through 18.5 km. HPDE pipe (e.g. BOD 46% decreasing), and causing BOD as influent to sedimentation pond 1 from 79.5 mg/L to 43.7 mg/L (e.g. BOD 45% BOD decreasing). After moving through the 5 treatment ponds, the obtained BOD from influent 43.7 mg/L to the effluent 23.0 mg/L (e.g. BOD 47% decreasing), and finally the measured values of 2.5, 2.2 and 2.5 mg/L in natural mangrove forest, mud beach and seashore, respectively (Table 1). It is remarkable to point out that the bacterial digestion processes (both aerobic and anaerobic processes) as the path way of Phetchaburi municipal wastewater treatment system were shown an effectiveness of H.M.The King's initiative nature by nature process without any doubt for Maintainability of sustainable marine water quality. However, the research results also found TDS, TSS, TKN, phosphate, nitrate and total coliform bacteria but opposite direction indicating on COD and pH because of the influences of marine water.

Intensive consideration is placed on an increase of DO in Phetchaburi municipal wastewater which found 0.9 mg/L at the fresh food markets and households 1.0 mg/L at sewage drainage system, 1.0 mg/L at Klongyang collection pond, 1.4 mg/L at the end of 18.5 km. HPDE pipe, effluent DO 7.6 mg/L at the outlet of 5 treatment ponds, 6.0, 5.9, and 6.0 mg/L in the mangrove forest, mud beach, and seashore, respectively. It would be assured the values of DO

in the sea water for supplying to bacteria organic digestion processing that causing the drastic decreasing of BOD, TDS, TSS, TKN, Phosphate and Nitrate. Surprisingly, the total coliform bacteria found very high numbers of $7.9\times10^6$ MPN/100mL in Klongyang collection pond and $3.9\times10^6$ MPN/100mL but the drastic decreasing by indicating the measured value only $1.7\times10^2$ MPN/100mL after passing through 5 treatment ponds of the Royal LERD project. In addition, the existence of total coliform bacteria was accepted to carry in mangrove forest for 9.0 MPN/100mL, in mud beach 11.0 MPN/100mL and in the seashore less 1.8 MPN/100mL, respectively (Table 1).

Table 1. Representatives of wastewater quality indicators as measured at Phetchaburi municipal point sources on fresh market and households, sewage system, Klongyang collection pond, 18.5 pipes, 5 consecutive community wastewater treatment ponds, mangrove forest, mud beach and seashore as conducted during August to October 2010 for about 80 hr. traveling time

| Measuring point | Wastewater Quality Indicators | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | pH (-) | TDS (mg/L) | TSS (mg/L) | DO (mg/L) | BOD (mg/L) | COD (mg/L) | TKN (mg/L) | $NO_3^-$ (mg/L) | $PO_4^{3-}$ (mg/L) | Coliform Bacteria (MPN/100mL) |
| Fresh Food Markets and Households | 6.8 | 2,761.90 | 1,127.00 | 0.9 | 546.6 | 1,156.00 | 14.6 | 0.7 | 25 | - |
| Sewage Drainage System | 7.8 | 462 | 100.2 | 1 | 162.7 | 242.9 | 15.3 | 0.2 | 4.1 | - |
| Klongyang Collection Pond | 6.2 | 410 | 27 | 1 | 79.5 | 96 | 14.6 | 0.6 | 1.9 | $7.9\times10^6$ |
| 18.5 km. Pipe | 7.1 | 480 | 35.9 | 1.4 | 43.7 | 95.2 | 14.1 | 0.6 | 1.6 | $3.9\times10^6$ |
| 5-Treatment Pond (WTP) | 8.8 | 674 | 105 | 7.6 | 23 | 96 | 3.6 | 0.7 | 0.2 | $1.7\times10^2$ |
| Mangrove Forest | 8 | 32,032.00 | 138 | 6 | 2.5 | 922 | 3.8 | 0.1 | 0.1 | 9 |
| Mud Beach | 8 | 31,424.00 | 165 | 5.9 | 2.2 | 984 | 4 | 0.1 | 0.2 | 11 |
| Seashore | 8 | 32,784.00 | 196 | 6.2 | 2.5 | 1,055.00 | 3.8 | 0.1 | 0.22 | ‹1.8 |

### *4.3 Appropriate Numbering of Oxidation Treatment Ponds*

In practical point of view, the Royal LERD project has constructed 5 treatment ponds for conducting Phetchaburi municipal wastewater treatment as transported from Phetchaburi municipal about 18.5 km by HPDE pipes as mentioned the previous sections. In terms of economic purpose, if possible, it should take only 3 rather than 5 treatment ponds. However, an observation on 20 year study indicated that the flowing wastewater over weir crest of oxidation pond 3 found the effluent under surface water quality standards as shown in Figure 3, on which the most probable numbering has been placed on the first pond (sedimentation pond 1), the second pond (oxidation pond 2), and the third pond (oxidation pond 3). It is shown the effectiveness of the drastic decreasing on effluent of treated wastewater quality on BOD, TKN, and coliform bacteria, while DO has the increasing tendency which are followed the bacterial organic digestion principles and going hand in hand with availability of marine water to act both the dilution and adding some chemical compounds to such treated wastewater effluent.

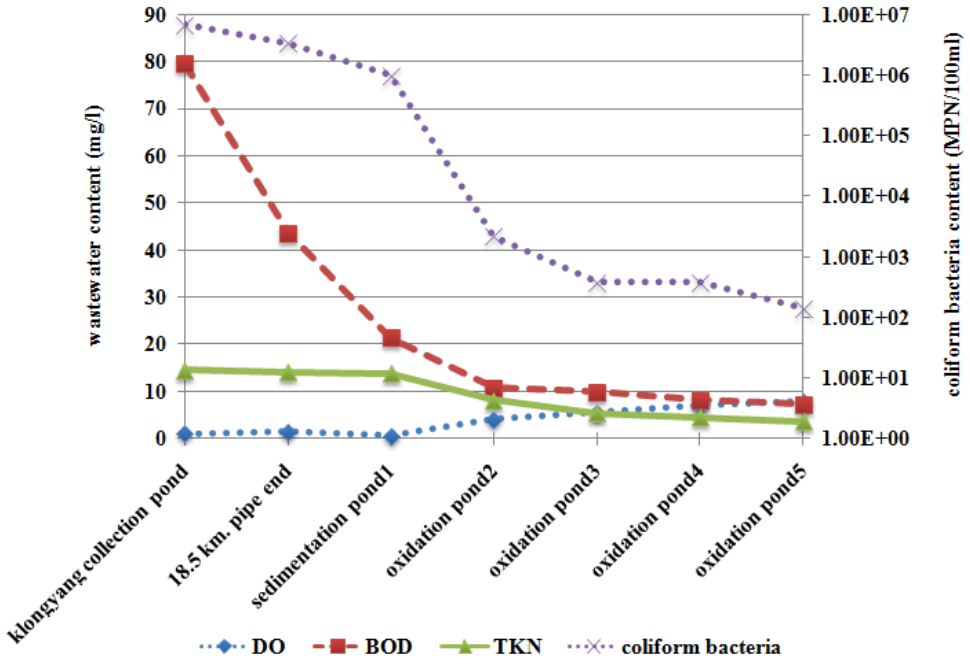



Figure 3. Effluent quality indicators of Phetchaburi municipal wastewater transporting from tip of 18.5 km. HPDE pipe through sedimentation pond 1, oxidation ponds 2, 3, 4 and to final oxidation pond 5 as measured continuously 24 times of sampling (except coliform bacteria only 6 times of sampling) for one year period by beginning on 28 December 1999 to 28 August 2000 as the phenomena representatives

It would be expressed that the treatment efficiency of 3 ponds (sedimentation pond 1, oxidation ponds 2 and oxidation pond 3) as belonged to the Royal LERD project have shown only the decreasing of organic wastes and also to eradicate the total coliform bacteria, Moreover, there is no finding the *E.coli* and another pathogenic bacteria after the wastewater body flowing over the weir crest of the oxidation pond 3 which was investigated by the Department of Medical Science, Ministry of Public Health, Thailand. This mean that whenever the community wastewater treatment has been taken by the 3 ponds system with HRT approximately 21 days, the obtained treated wastewater should be surely clean enough not only decreasing organic matters of both the solid and liquid forms, but also to eradicate pathogenic bacteria in Phetchaburi municipal wastewater without doubtfulness. Seasonal variation of bacteria count was considered to show the drastic change among outlets of treating units as illustrated in Table 2 and Figure 4. Results indicated drastically decreasing in the summer time from the Klongyang collection pond through the 18.5 km. at sedimentation pond 1 for total coliform bacteria and oxidation pond 3 for fecal coliform bacteria. The reason could be pointed out that both facultative bacteria (total coliform bacteria and fecal coliform bacteria) needed oxygen for survival but it was very less in the sewage system as the same as in higher concentration of organic content of Klongyang collection pond and also no oxygen in the 18.5 km pipe. When both the total coliform bacteria and fecal coliform bacteria were flown out to contact at sedimentation pond 1 with higher concentration of organic content but less dissolved oxygen that recovered oxygen causing survival of total coliform bacteria but still decreasing fecal coliform bacteria due to not being enough oxygen for their living. It would be remarkable that the drastic decreasing in rainy and winter season showed the same trends as in summer time on both the total coliform bacteria and fecal coliform bacteria. Due to high concentration of organic matters in Phetchaburi community wastewater, the drastic decreases were indicated at the oxidation pond 3 for both the total coliform bacteria and fecal coliform bacteria.

Table 2. Seasonal variation of bacteria count effluent (treated wastewater) from Klongyang collection pond, through tip of 18.5 km. HDPE pipe and 5 treatment ponds

| Date | Type of Coliform Bacteria | Bacteria Count (MPN/100mL) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Klongyang Collection pond | 18.5 km HDPE Pipe | Pond 1 | Pond 2 | Pond 3 | Pond 4 | Pond 5 |
| 10 May 2007 | Total | 200,000 | 16,000 | 9,200 | 130 | 22 | 33 | 23 |
| (Summer) | Fecal | 20,000 | 2,700 | 140 | 11 | 7 | 11 | 2 |
| 15 Aug 2007 | Total | 120,000 | 14,000 | 280 | 130 | 1,000 | 140 | 350 |
| (Rainy) | Fecal | 90,000 | 1,300 | 180 | 79 | 350 | 79 | 130 |
| 12 Dec 2007 | Total | 2,300,000 | 500,000 | 8,000 | 16,000 | 9,200 | 3,500 | 28,000 |
| (Winter) | Fecal | 800,000 | 200,000 | 5,000 | 170 | 54 | 170 | 120 |

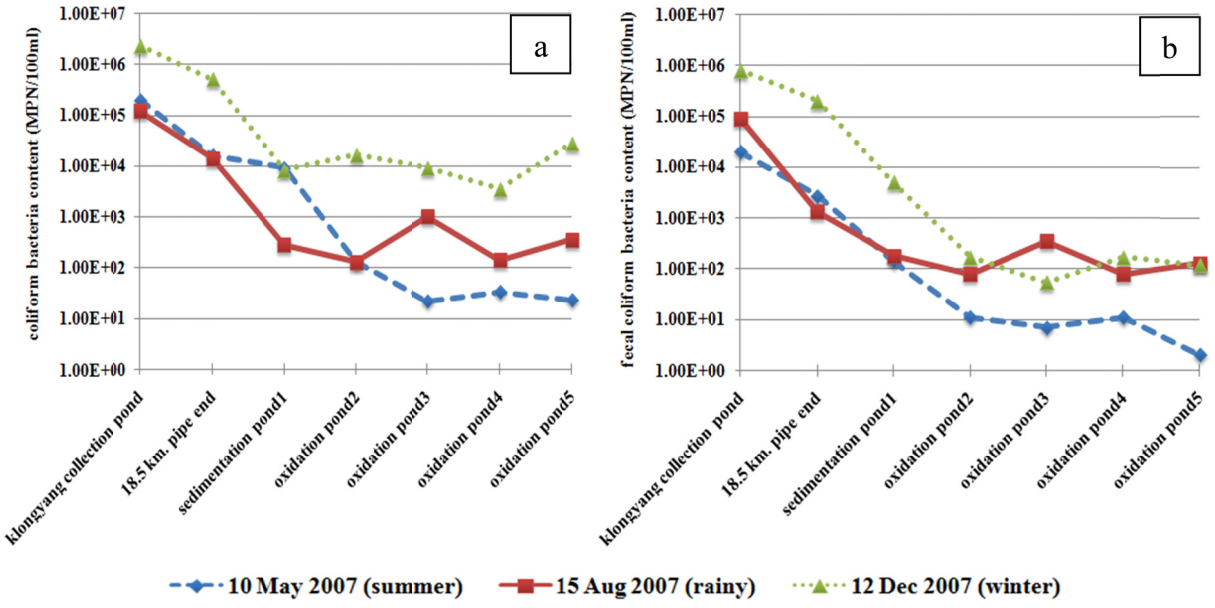



Figure 4. Seasonal variation of total coliform bacteria and fecal coliform bacteria as measured in summer season, rainy season and winter; (a) total coliform bacteria and (b) fecal coliform bacteria

### *4.4 Quantitative Solar Energy Measurement*

Owning to the disappearance of *E.coli* was discovered in treated wastewater after overflowing the weir crest of oxidation pond 3 which took 21 days for hydraulic retention time (HRT) since the first drop going in the sedimentation pond 1 as mentioned in previous section. The cause of *E.coli* death was hypothesized on the ultraviolet radiation (UVR) and all wavelength of spectrum, therefore the measurement of solar energy was conducted at the Royal LERD experimental area. Unfortunately, the instrument can measure only UV-A and UV-B, not for UV-C measuring but primary evaluation can be taken in relation to the disappearance of *E.coli* after overflowing the oxidation pond 3 weir crest before moving into the oxidation pond 4. However, the measuring solar energy on the spectrum (violet, indigo, blue, green, yellow, orange, and red), UV-B and UV-A were shown in Figure 5a, while the energy balance also measured on net radiation (Rn), incoming shortwave radiation (Rsd) and incoming longwave radiation (Rld) as seen Figure 5b. The relationship can be expressed as Rn = Rsd-Rsu+Rld-Rlu. Natural point of view, Rn is supposed to use for latent heat flux (LE), sensible heat flux (H), soil heat flux (G), photosynthesis (Ph), metabolism (M), and chemical reaction (C) which can be written as Rn = LE+H+G+Ph+M+C. However, all above statements play a significant role in the death of *E.coli* after wastewater flowing over the weir crest of sedimentation pond 1 oxidation pond 2 and oxidation pond 3. It is a matter of fact, the Royal LERD project

area is located on the open uniform site, therefore the effects of blockage due to surrounding trees and buildings can be ignored for consideration on solar energy.

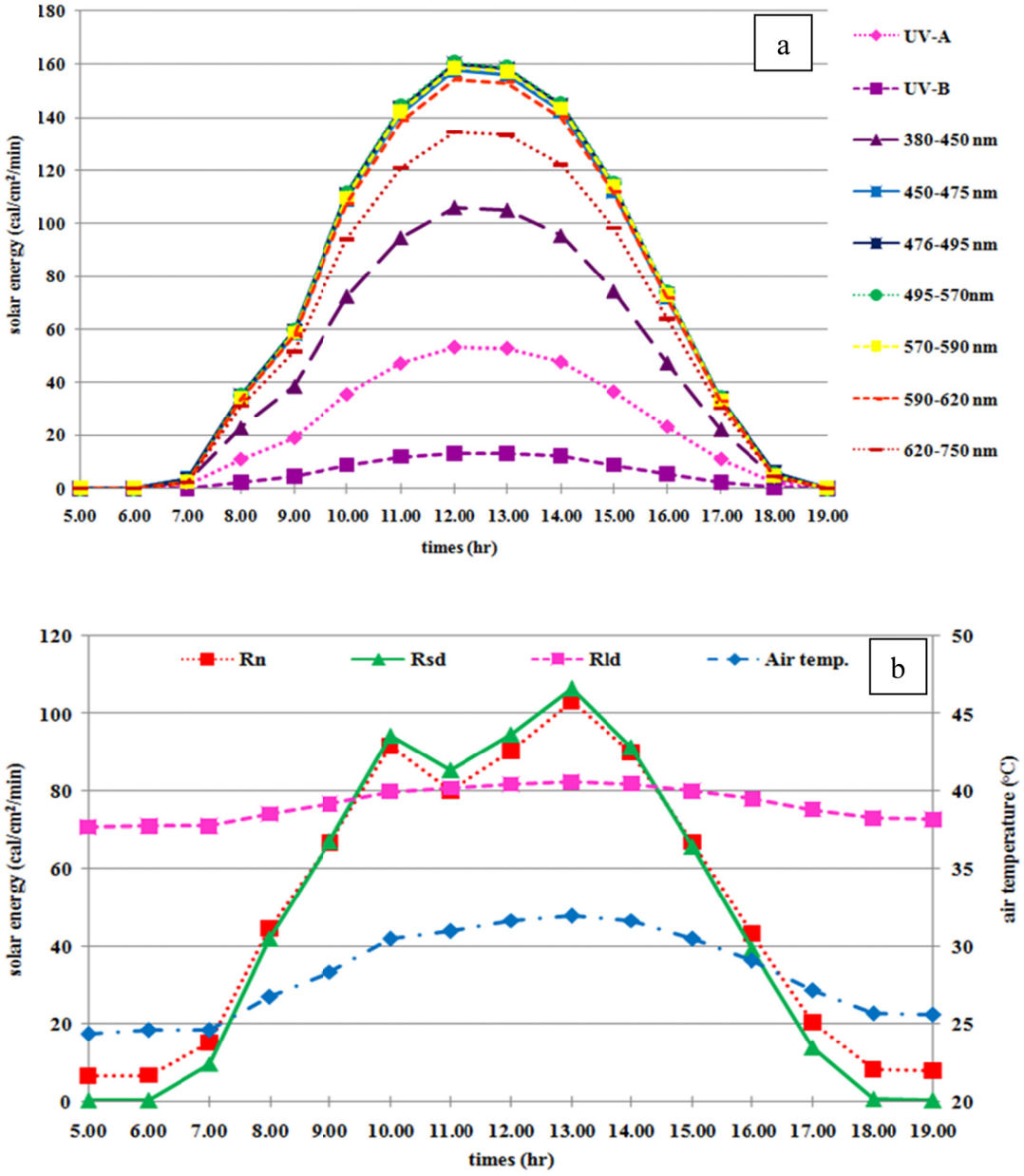



Figure 5. UV-A, UV-B and Spectrum (a) and energy balance (b) of the Royal project site in Phetchaburi province

Results of measuring UV-A, UV-B and spectrum (violet, indigo, blue, green, yellow, orange, and red) and also energy balance were shown in Figures 5a and 5b in which the UV-A, UV-B and spectrum begin to shine after 06:00 a.m. and maximizing about noon. In the same behavior, the heat energy is gradually increased to the peak, and then decreasing down to about 07:00 p.m., in turn to the increasing in heat enough to affect on the increment of DO and $H_2O_2$ for encouraging the bacterial organic digestion processes.

*4.5 Water Quality in Wastewater Treatment System*

First of all, the measurement of the concerned water quality indicators was randomly conducted on the whole day of 19 May 2011, the results as shown in Table 3. It would be noted that this experiment was taken in total bacteria as the representative of *E.coli* which occurs directly from human.

Table 3. Some concerned water quality indicators as measured on the 19 May 2011 at the sedimentation pond 1 and oxidation pond 3

| Times | Sedimentation Pond 1 | | | | | Oxidation pond 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Temp (°C) | pH (-) | DO (mg/L) | BOD (mg/L) | Total Bacteria (CFU/mL) | Temp (°C) | pH (-) | DO (mg/L) | BOD (mg/L) | Total Bacteria (CFU/mL) |
| 06.00 | 29.2 | 6.9 | 1.2 | 116.0 | $1.8\times10^{8}$ | 29.8 | 6.8 | 2.5 | 5.3 | $1.0\times10^{6}$ |
| 07.00 | 29.4 | 6.9 | 1.4 | 118.0 | $5.0\times10^{6}$ | 29.9 | 7.0 | 2.5 | 1.3 | $1.0\times10^{4}$ |
| 08.00 | 30.4 | 6.9 | 2.2 | 120.0 | $7.3\times10^{7}$ | 30.3 | 7.0 | 5.0 | 3.6 | $2.0\times10^{5}$ |
| 09.00 | 30.6 | 7.0 | 2.1 | 130.0 | $3.4\times10^{8}$ | 30.4 | 7.1 | 4.7 | 6.0 | $4.0\times10^{6}$ |
| 10.00 | 31.6 | 7.2 | 4.1 | 136.0 | $1.7\times10^{8}$ | 31.5 | 7.5 | 6.9 | 20.6 | $8.0\times10^{4}$ |
| 11.00 | 33.1 | 7.4 | 7.7 | 36.5 | $4.7\times10^{8}$ | 32.4 | 7.6 | 9.3 | 19.6 | $8.0\times10^{4}$ |
| 12.00 | 34.3 | 7.6 | 11.9 | 35.5 | $1.4\times10^{8}$ | 32.9 | 7.8 | 10.7 | 21.0 | $1.3\times10^{2}$ |
| 13.00 | 34.3 | 7.8 | 13.1 | 36.5 | $3.2\times10^{8}$ | 32.3 | 8.0 | 10.9 | 23.6 | $1.0\times10^{4}$ |
| 14.00 | 33.8 | 7.8 | 13.6 | 39.5 | $2.5\times10^{8}$ | 32.9 | 8.0 | 14.1 | 23.0 | $2.0\times10^{6}$ |
| 15.00 | 34.0 | 8.0 | 15.3 | 39.0 | $1.4\times10^{8}$ | 32.4 | 8.0 | 13.9 | 24.6 | $2.0\times10^{4}$ |
| 16.00 | 33.7 | 8.0 | 15.1 | 30.0 | $6.2\times10^{8}$ | 32.1 | 7.9 | 12.6 | 22.6 | $2.0\times10^{6}$ |
| 17.00 | 32.7 | 7.7 | 9.5 | 33.5 | $4.2\times10^{8}$ | 32.1 | 7.6 | 12.3 | 21.3 | $1.7\times10^{2}$ |
| 18.00 | 32.1 | 7.9 | 8.4 | 37.0 | $3.3\times10^{8}$ | 31.7 | 8.0 | 11.2 | 19.3 | $3.0\times10^{4}$ |
| 19.00 | 31.0 | 8.0 | 1.3 | 37.5 | $4.1\times10^{8}$ | 31.7 | 8.2 | 9.3 | 20.6 | $1.0\times10^{4}$ |
| Average | 31.7 | 7.5 | 7.6 | 67.5 | $2.7\times10^{8}$ | 31.3 | 7.6 | 8.9 | 16.6 | $6.7\times10^{5}$ |

The values of BOD in sedimentation pond 1 was ranged between 116.0-136.0 mg/L during 06:00 a.m. to 10:00 a.m. because of heavy cooler water surface sinking downward to the bottom of the pond that forcing the sediment going up to the surface (under the thermo siphon process), then the aerobic bacteria with the high amount of $10^6$ to $10^8$ CFU/mL which functioned as electron acceptors to digest the organic matters (Cloete et al., 1983). Until 11:00 a.m., the full sunshine was appeared and caused solar energy functioning for photosynthesis of phytoplankton to produce more DO and meeting the maximum at 03:00 p.m., and resulting DO increasing to 15.3 mg/L which was enough for bacteria to employ as the electron acceptor in the organic digestion process that decreasing BOD down between 30.0-39.5 mg/L.Water quality in oxidation pond 3 found temperature 29.8-32.9 °C, pH 6.8-8.0, DO 2.5-14.1 mg/L and BOD 1.3-24.6 mg/L, while the DO was low value of 2.5-5.0mg/L during 06:00 a.m. to 09:00 a.m. but BOD became low only 1.3-6.0 mg/L because of DO was not employed as electron acceptors by aerobes in organic digestion processes in which it would be harmonized with the study on the efficiency of aerobic bacteria organic digestion that employing oxygen as electron acceptor. However, it was discovered that the appropriate quantity of DO for aerobic bacteria organic digestion with using oxygen as electron acceptor could be occurred during DO ranging between 4.0-8.0 mg/L (Taylor et al., 2009). Therefore, time between 10:00 a.m. to 07:00 p.m. of each day should be the period of increased DO values between 6.9-14.1 mg/L which was the appropriate quantity for aerobic bacteria that employing oxygen as electron acceptor to enhance the activities of organic digestion processing, affecting an increasing of organic content in form of BOD.

*4.6 Quantity of Dissolved Oxygen in Wastewater Treatment System*

The measurement of water quality in sedimentation pond 1 resulted DO 1.2 mg/L on 06:00 a.m. and trending to increase in maximum values of 15.3 mg/L on 03:00 p.m. (averaged producing rate 1.5 mg/L/hr) but the DO quantity was rapidly decreased from 15.3 mg/L down to 1.3 mg/L at the period between after 03:00 p.m. to 07:00 p.m. (averaged decreasing rate 3.5 mg/L/hr). In contrary, DO in oxidation pond 3 was found rapid increase from 2.5 mg/L on 06:00 a.m. to 14.1 mg/L on 02:00 p.m. (averaged producing 1.4 mg/L/hr), after 02:00 p.m. DO decreasing from 14.1 mg/L down to 9.3 mg/L on 07:00 p.m. (averaged decreasing rate 0.9 mg/L/hr) as shown in Figure 6.

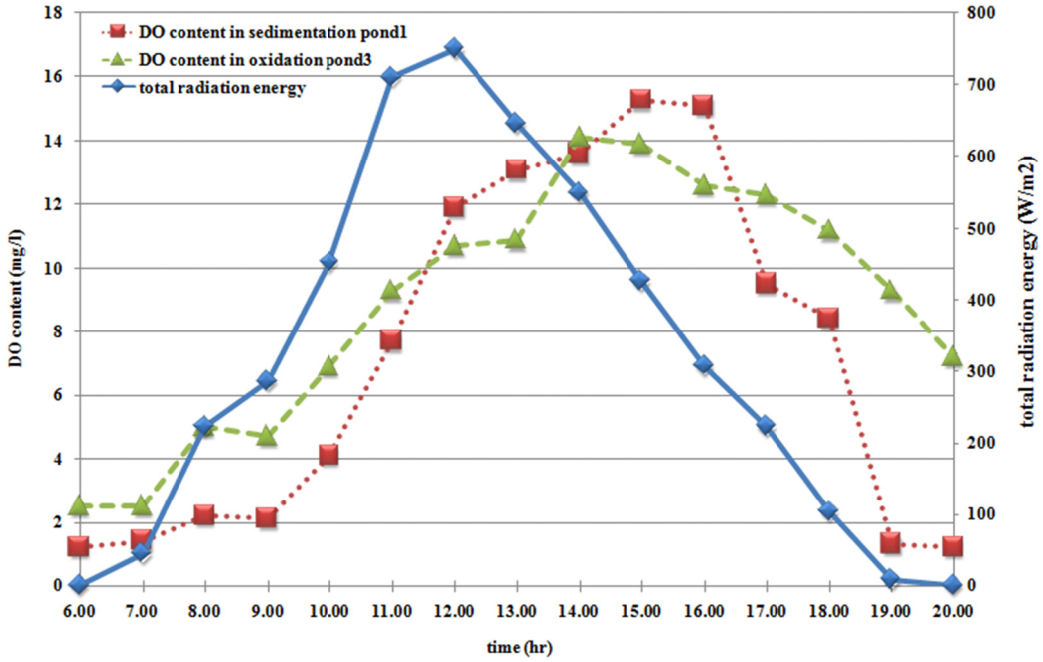
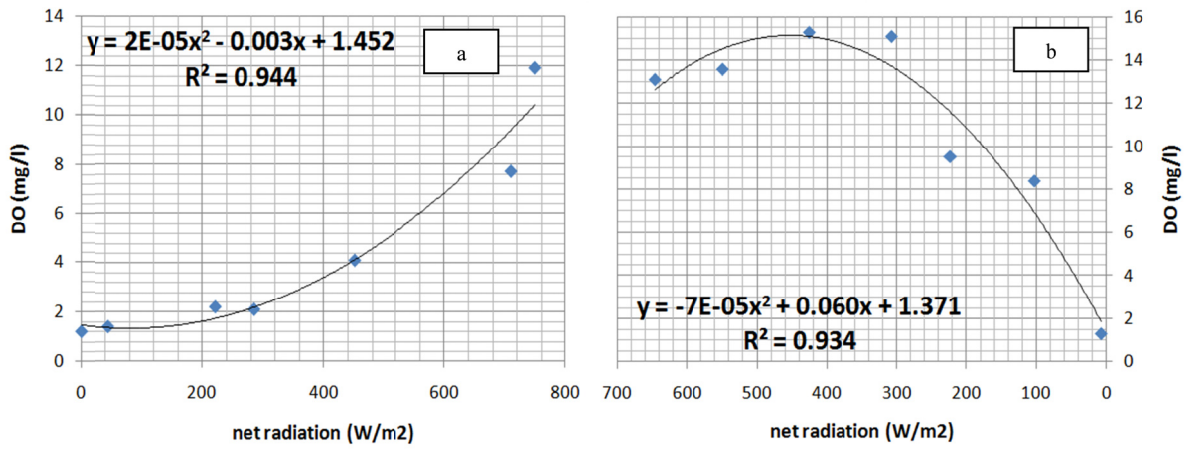



Figure 6. Distribution of DO in wastewater treatment ponds for Phetchaburi municipal wastewater treatment at the Royal LERD project site at Laem Phak Bia sub district, Ban Laem district, Phetchaburi province, Thailand

From Figure 6, quantity of DO in sedimentation pond 1 and oxidation pond 3 were consisted of very close producing rates, i.e., 1.5 mg/L and 1.4 mg/L, respectively. So that, they were resulted from photosynthesis of abundant phytoplankton of both ponds. It is evident since after 12:00 a.m. that DO in sedimentation pond 1 and oxidation pond 3 were decreased due to be employed for electron acceptors in the respiration process (Cloete et al., 1983), but the decreasing rate of DO in sedimentation pond 1 was faster rate than oxidation pond 3, i.e. 3.5 mg/L/hr and 0.9 mg/L/hr, respectively because of more bacteria. In order to insist this statement, the water samples were analyzed the total bacteria in sedimentation pond1 found out ranging between $10^6$ to $10^8$ CFU/mL, while the oxidation pond 3 found the total bacteria ranging between $10^2$ to $10^6$ CFU/mL as shown in Table 3. For making clear understanding, the polynomial correlation was studied on the relationship between solar energy and DO of sedimentation pond 1 and oxidation pond 3 for two periods in the morning (06:00-12:00 a.m.) and in the afternoon (01:00-07:00 p.m.). The results of the correlation coefficient ($R^2$) of sedimentation pond 1 in the morning found 0.944 and in the afternoon 0.934, and also for oxidation pond 3 findings of 0.983 in the morning period and 0.8252 in the afternoon, on which they were shown very high correlation coefficients as illustrated in Figures 7a, b, c and d.



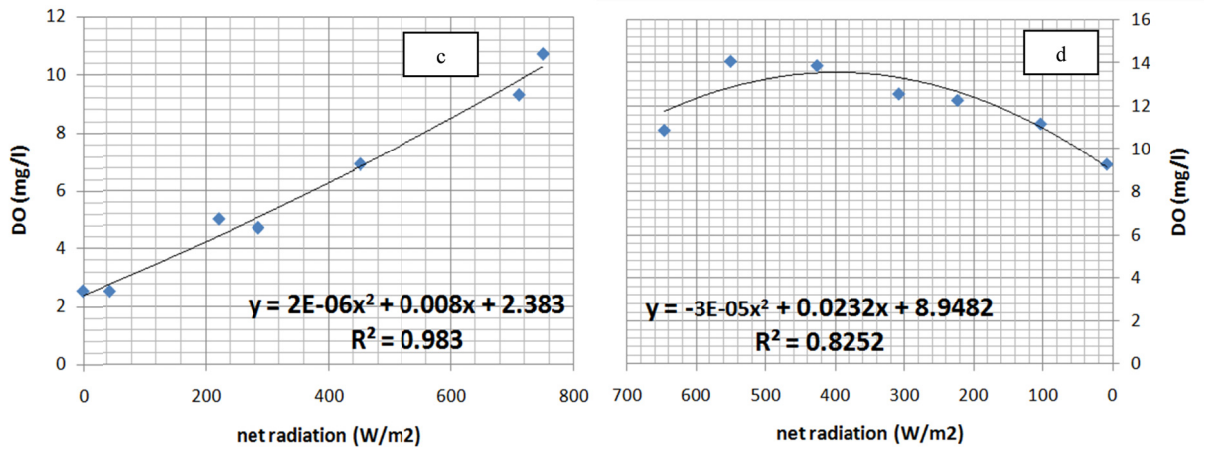



Figure 7. Graphical illustration for the relationship between solar radiation and DO in wastewater treatment ponds of the Royal LERD project site; (a) sedimentation pond 1 in the morning, (b) sedimentation pond 1at noon, (c) oxidation pond 3 in the morning and (d) oxidation pond 3 at noon

*4.7 Quantity of Hydrogen Peroxide in Wastewater Treatment System*

The measurement of $H_2O_2$ in sedimentation pond 1 was found out 0.34 µg/L in the morning at 10:00 a.m. and met the maximum value of 1.18 µg/L at 01:00 pm (producing rate 0.28 µg/L/hr). After that, the quantity of $H_2O_2$ was gradually decreased from 0.76 µg/L at 02:00 p.m. to 0.08 µg/L at 05:00 p.m. (decreasing rate 0.27 µg/L/hr). For oxidation pond 3, the amount of $H_2O_2$ was found 2.04 µg/L at 06:00 a.m. and climbing up to the maximum increasing of 2.89 µg/L at 04:00 p.m. (producing rate 0.11 µg/L/hr), then it decreased down to 2.38 µg/L at 07:00 p.m. (decreasing rate 0.17 µg/L/hr) as shown in Table 4.

Table 4. Quantity of $H_2O_2$ in Phetchaburi municipal wastewater treatment system of sedimentation pond 1 and oxidation pond 3 at the Royal LERD project site

| Time (hr.) | Net radiation ($W/m^2$) | $H_2O_2$ (µg/L) | |
|---|---|---|---|
| | | Sedimentation pond 1 | Oxidation pond 3 |
| 06.00 | 0 | 0.0 | 2.04 |
| 07.00 | 45 | 0.0 | 1.62 |
| 08.00 | 222 | 0.0 | 1.36 |
| 09.00 | 286 | 0.0 | 2.04 |
| 10.00 | 452 | 0.34 | 2.30 |
| 11.00 | 711 | 0.51 | 2.46 |
| 12.00 | 750 | 0.68 | 2.46 |
| 13.00 | 646 | 1.18 | 1.87 |
| 14.00 | 550 | 0.76 | 2.72 |
| 15.00 | 426 | 0.51 | 2.64 |
| 16.00 | 308 | 0.34 | 2.89 |
| 17.00 | 223 | 0.08 | 2.64 |
| 18.00 | 103 | 0.0 | 2.38 |
| 19.00 | 8 | 0.0 | 2.38 |

From Table 4, $H_2O_2$ in sedimentation pond 1 and oxidation pond 3 were varied directly to total solar energy (William & Zika, 1983) and maximizing in the afternoon period, that the same trend as southern California which found maximum accumulative $H_2O_2$ in the sea in the afternoon (Catherine et al., 2010) according that solar energy encouraging molecule of DO and water ($H_2O$) being separated into free radical, i.e., OH*. After that, the OH* will

take an action to gain the product of $H_2O_2$ (Gerringa et al., 2004; Steven et al., 2010) which varied directly to increase and decrease of solar radiation but the decreasing of $H_2O_2$ in sedimentation pond 1 was not similar to oxidation pond 3. The decreasing in sedimentation pond 1 of $H_2O_2$ could be attached at 0.00 µg/L while oxidation pond 3 was still found with continuous delectability because of the rapid decreasing of DO in sedimentation pond 1 during the afternoon period. Since, the molecules of oxygen was affected to the decrease of primary object of $H_2O_2$ occurring, which could not be found out $H_2O_2$ in sedimentation pond 1 after 06:00 p.m.

For clearly understanding, the polynomial correlation was studied on the relationship between solar energy and $H_2O_2$ and found out the correlation coefficient in sedimentation pond 1 0.944 and 0.986 in the morning and afternoon, respectively (Figure 8a and b). For oxidation pond 3, the correlation coefficients of $H_2O_2$ were detected as 0.606 and 0.700 in the morning and in the afternoon, respectively (Figures 8c and d).

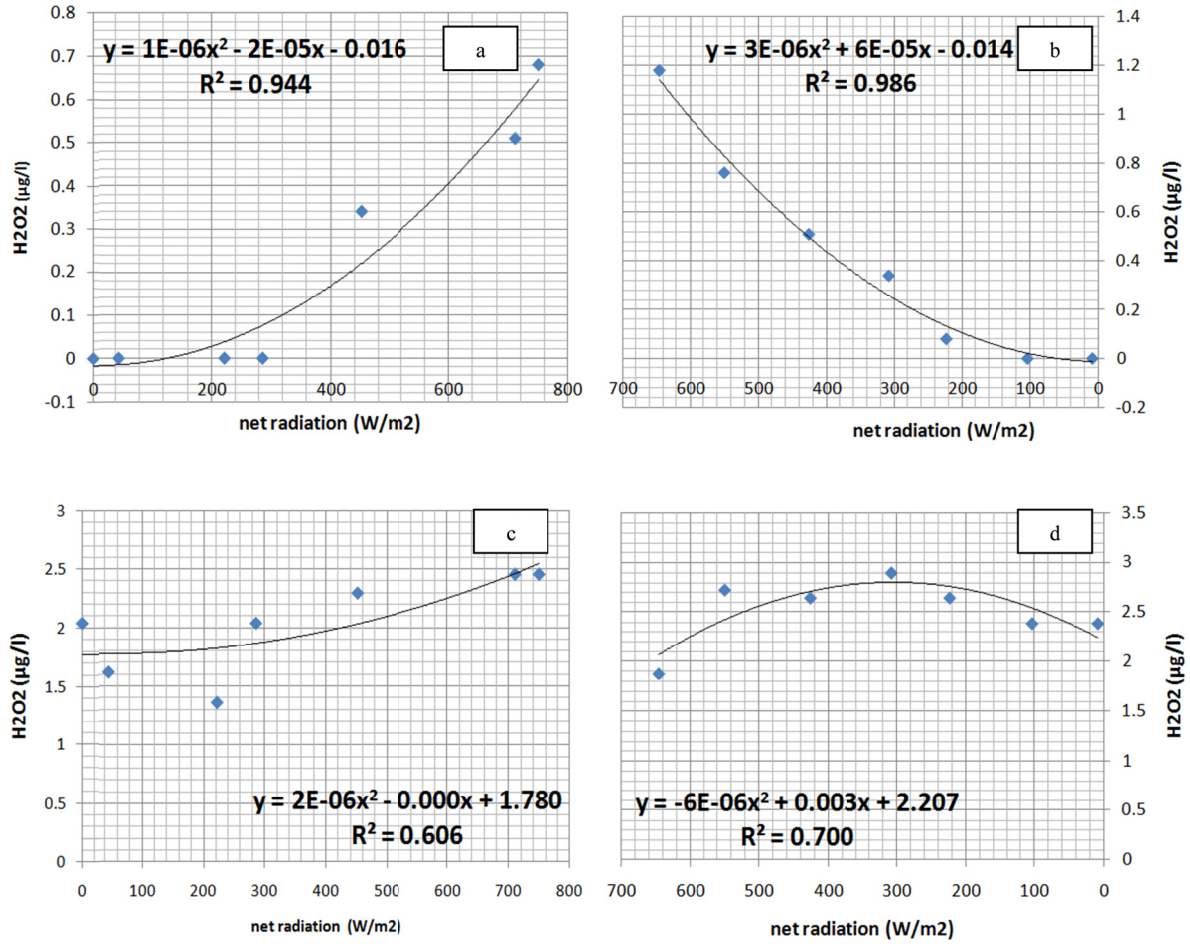



Figure 8. Graphical analysis of the relationship between solar radiation and $H_2O_2$ in Phetchaburi municipal wastewater treatment system in sedimentation pond 1 and oxidation pond 3 at the Royal LERD project site; (a) sedimentation pond 1 in the morning, (b) sedimentation pond 1 at noon, (c) oxidation pond 3 in the morning and (d) oxidation pond 3 at noon

## 5. Conclusion

Due to the previous studies of community wastewater treatment under H.M.The King's initiative nature by nature process on Laem Phak Bia environmental research and development project at Laem Phak Bia sub district, Ban Laem district, Phetchaburi province in Thailand have been implemented since the year of 1990 up to the present time and found the drastic decreasing of total coliform and fecal coliform after draining out of wastewater from the tip of 18.5 km. HPDE pipe to sedimentation pond 1 through another four consecutive oxidation ponds (ponds 2, 3, 4 and 5), especially after overflowing weir crest of oxidation pond 3 that found very less down to zero MPN/ml. In depth study on the effect of solar energy on DO and $H_2O_2$ in wastewater treatment ponds found that the solar radiation as measured at the Royal LERD project site resulted 0-750 Watts/$m^2$. This amount of solar energy can

make the separation of oxygen in wastewater to encourage the $H_2O_2$. Besides, the solar energy was still shown the relation to DO and $H_2O_2$ in sedimentation pond 1 rather than the oxidation pond 3, due to the lower quantity of dissolved organic and inorganic matters in oxidation pond 3 than sedimentation pond 1. It was affected to aerobic bacterial organic digestion processes not employing less oxygen as electron acceptor. In other words, the oxygen as occurred from phytoplankton photosynthesis could not be employed for processing of organic digestion and respiration in very small amount on night time, and let it be accumulated oxygen in water. Therefore, it makes sure that the amount of DO and $H_2O_2$ in oxidation pond 3 did not vary directly to total solar energy.

**Acknowledgment**

This research received financial supports from the H.M.The King'S LERD Project.

**References**

Alenka, T., Marija, V., Željko, P., & Dinka, M. (2006). The hydrogen peroxide as a potentially useful slurry disinfectant. *Livestock Science, 102*, 243-247. http://dx.doi.org/10.1016/j.livsci.2006.03.022

APHA, AWWA, & WPCF. (1995). *Standard Method for the Examination of Water and Wastewater* (19th ed.). Washington, DC: America Public Health Association.

Asad, N. R., Asad, L. M., Silva, A. B., Felzenszwalb, I., & Leitao, A. C. (1998). Hydrogen peroxide effect in *Escherichia coli* cell. *Acta Biochim Pol., 45*, 677-690.

Catherine, D. C., Bruyn, W. J., Hirsch, C. M., & Jakubowski, S. D. (2010). Hydrogen peroxide measurements in recreational marine bathing water in southern California, USA. *Water Research, 44*, 2203-2210. http://dx.doi.org/10.1016/j.watres.2009.12.044

Cloete, T. E., Roux, J. D. L., Toerien, D. F., & Pieterse, A. J. H. (1983). Oxygen Dynamics and Heterotrophic Aspects of a Pond Treatment System for Cattle Feedlot Effluent. *Agricultural Wastes, 7*, 147-174. http://dx.doi.org/10.1016/0141-4607(83)90050-1

Chunkao, K., Nimpee, C., & Duangmal, K. (2012). The King's initiatives using water hyacinth to remove heavy metals and plant nutrients from wastewater through Bueng Makkasan in Bangkok, Thailand. *Ecological Engineering, 39*, 40-52. http://dx.doi.org/10.1016/j.ecoleng.2011.09.006

Dani, J. B., Reichwaldt, E. S., & Ghadouani, A. (2012). The use of hydrogen peroxide to remove cyanobacteria and microcystins from waste stabilization ponds and hypereutrophic systems. *Ecological Engineering* (In press).

Davies-Colley, R. J., Donnison, A. M., & Speed, D. J. (1997). Sunlight wavelengths Inactivating faecal indicator microorganisms in waste stabilization ponds. *Water Science and Technology, 35*, 219-225. http://dx.doi.org/10.1016/S0273-1223(97)00262-X

Donat, P. H., Lebert, M., Moya, A. F., Jimenez, C., Mercado, J., Salles, S., Aguilera, J., & Figueroa, F. L. (1997). Effect of solar radiation on the photosynthetic activity of the red alga *Corallina elongate* Eillis et Soland. *Journal of Photochemistry and Photobiology B: Biology, 37*, 196-202. http://dx.doi.org/10.1016/S1011-1344(96)07402-7

Gen, S. W., Liao, C. H., & Wu, F. J. (2001). Photodegradation of humic acids in the presence of hydrogen peroxide. *Chemosphere, 42*, 379-387. http://dx.doi.org/10.1016/S0045-6535(00)00153-3

Gerringa, L. J. A., Rijkenberg, M. J. A., Timmermans, K. R., & Bumma, A. G. J. (2004). The influence of solar ultraviolet radiation on the photochemical product of $H_2O_2$ In the equatorial Atlantic Ocean. *Journal of Sea Research, 51*, 3-10. http://dx.doi.org/10.1016/j.seares.2003.03.002

Hans, C. P. M., Visser, P. M., Reeze, B., Meeuse, J., Slot, P. C., Wijn, G., Talens, R., & Huisman, J. (2012). Selective suppression of harmful cyanobacteria in an entire lake with hydrogen peroxide. *Water Research, 46*, 1460-1472. http://dx.doi.org/10.1016/j.watres.2011.11.016

Hegde, A., Bhat, G. K., & Mallya, S. (2008). Effect of exposure to hydrogen peroxide on the virulence of *Escherichia coli*. *India Journal of Medical Microbiology*, 26, 25-28. http://dx.doi.org/10.4103/0255-0857.38853

John, H. J., & Bukata, R. P. (1998). Impact of Stratospheric Ozone Depletion on Photoproduction of Hydrogen Peroxide in Lake Ontario. *Journal of Great Lakes Research, 24*, 929-935. http://dx.doi.org/10.1016/S0380-1330(98)70873-1

Leena, S., Rousseau, D. P. L., Hooijmans, C. M., & Lens, P. N. L. (2011). 3D model for a secondary facultative pond. *Ecological Modeling, 222*, 1592-1603. http://dx.doi.org/10.1016/j.ecolmodel.2011.02.021

Macos, V. S. (2005). Modeling of coliform removal in 186 facultative and maturation ponds around the world. *Water Research, 39*, 5261-5273. http://dx.doi.org/10.1016/j.watres.2005.10.016

Mohamed, K. (2006). Chemical Oxidation with Hydrogen Peroxide for Domestic Wastewater Treatment. *Chemical Engineering Journal, 119*, 161-165. http://dx.doi.org/10.1016/j.cej.2006.03.022

Richard, J. W., Washington, D., Howsawkeng, J., Loge, F. J., & Teel, A. L. (2003). Comparative toxicity of hydrogen peroxide, hydroxyl radicals, and superoxide anion to *Escherichia coli*. *Advances in Environmental Research, 7*, 961-968. http://dx.doi.org/10.1016/S1093-0191(02)00100-4

Richard, Z. G., Callaghan, T. V., & David, V. J. (1995). Effect of increased solar ultraviolet radiation on biogeochemical cycles. *Ambio, 24*, 181-187.

Robert, G., & Gambini, D. J. (1990). *Applied radiobiology and radiation protection*. England: Ellis Horwood Limited.

Roberto, R., Rodríguez, A., Antonio, J., Melón, P., Petre, A., & García-Calvo, E. (2009). Oxidation of dissolved organic matter in the effluent of a sewage treatment plant using ozone combined with hydrogen peroxide ($O_3/H_2O_2$). *Chemical Engineering Journal, 149*, 311-318. http://dx.doi.org/10.1016/j.cej.2008.11.019

Rui, P. Q., Li, N., Qi, X. H., Wang, Q. S., & Zhuang, Y. Y. (2005). Degradation of microcystin-RR by UV radiation in the presence of hydrogen peroxide. *Toxicon, 45*, 745-752. http://dx.doi.org/10.1016/j.toxicon.2005.01.012

SOLVAY. (2004). *Determination of Hydrogen Peroxide Concentration (0.1%-5%) Technical Data Sheet*. Solvay Chemicals, Inc.

Steven, A. R., Richard, L. E., Peake, B. M., Cooper, W. J., & Bodeker, G. E. (2010). The influence of solar radiation on hydrogen peroxide concentrations in freshwater. *Marine and Freshwater Research, 61*, 1147-1153. http://dx.doi.org/10.1071/MF10001

Tapas, K. D. (2001). Ultraviolet disinfection application to a wastewater treatment plant. *Clean Prod Processes, 3*, 69-80. http://dx.doi.org/10.1007/s100980100108

Taylor, S. M., Yiliang, H., Bin, Z., & Jue, H. (2009). Heterotrophic ammonium removal characteristics of an aerobic heterotrophic nitrifying-denitrifying bacterium, *Providencia rettgeri* YL. *Journal of Environmental Sciences, 21*, 1336-1341. http://dx.doi.org/10.1016/S1001-0742(08)62423-7

Volodymyr, I. L. (2011). Environmentally induced oxidative stress in aquatic animals. *Aquatic Toxicology, 10*(1), 13-30. http://dx.doi.org/10.1016 /j.aquatox.2010.10.006

William, J. C., & Zika, R. G. (1983). Photochemical formation of hydrogen peroxide in surface and ground waters exposed to sunlight. *Science, 220*, 711-712. http://dx.doi.org/10.1016/S1001-0742(08)62423-7

William, J. O., Asce, A. M., Harold, G. B., & Asce, M. (1852). Photosynthesis in sewage treatment. *American Society of Civil Engineers, 2849*, 73-80.

Yoel, K., Godinger, D., & Aronovith, J. (1987). Temporary exposure to hydrogen peroxide increase intracellular protein degradation in *E. coli*. *FEMS Microbiology Letters, 44*, 277-282. http://dx.doi.org/10.1016/S1001-0742(08)62423-7

Zulfiqar, A. B., Mahmood, Q., Raja, I. A., Malik, A. H., Rashid, N., & Wub, D. (2011). Integrated chemical treatment of municipal wastewater using waste hydrogen peroxide and ultraviolet light. *Physics and Chemistry on the Earth, 36*, 459-464. http://dx.doi.org/10.1016/j.pce.2010.03.024

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# A New Ranking of Environmental Performance Index Using Weighted Correlation Coefficient in Intuitionistic Fuzzy Sets: A Case of ASEAN Countries

Lazim Abdullah[1] & Wan Khadijah Wan Ismail[1]

[1] Faculty of Science and Technology, University Malaysia Terengganu, Kuala Terengganu, Malaysia

Correspondence: Lazim Abdullah, Faculty of Science and Technology, University Malaysia Terengganu, Kuala Terengganu 21030, Terengganu, Malaysia. Tel: 60-9-668-3335. E-mail: lazim_m@umt.edu.my

Received: March 21, 2013 Accepted: April 5, 2013 Online Published: May 14, 2013

doi:10.5539/mas.v7n6p42 URL: http://dx.doi.org/10.5539/mas.v7n6p42

**Abstract**

Growth in number of population and development nowadays indicate a good sign for nation's development. However, the development sometimes might neglects the preservation and conservation of nature and can reflects in environment performance. Concerning on this matters, Environmental Performance Index (EPI) has been introduced since 2006 to depict the environment performance for most of the countries in the world. The index considers ten policy categories associated with environmental public health and ecosystem sustainability. The main mathematics operation in establishing EPI is arithmetic mean of all ten policy categories. One of the weaknesses in the arithmetic mean is the operation might neglects some extreme values in data. Recently, Wan Ismail and Abdullah introduced the EPI using analytic hierarchy process (AHP) but the weight of policy category was not considered. This paper proposes a new ranking of EPI using a decision making tool of weighted correlation coefficient based on intuitionistic fuzzy sets (IFS). An original data of policy categories were converted into IFSs which benefiting in considering two-sided of membership and non membership. Criteria weights for alternatives in fuzzy correlation coefficient were utilized to set new EPI for nine ASEAN countries. A new ranking EPI among ASEAN countries show that Thailand is the highest EPI followed by Malaysia. The new ranking may offer an alternative measure in evaluating environmental performance particularly for ASEAN countries.

**Keywords:** environmental sustainability, environmental performance index, intuitionistic fuzzy sets, weighted correlation coefficient

## 1. Introduction

Sustainable development can be defined generally as the situation when development and preservation on environment get balance. However, other issues like economy sustainability and socio-political sustainability are not neglected. There are many types of sustainable development such as environmental sustainability, economical sustainability, socio-political sustainability, ecological sustainability and cultural sustainability. Environmental sustainability is a process to make sure that the daily life activities and any usage of environment is friendly environmental and preserved environment. An unsustainable environment is the situation where the usage and development does not preserve the environment and the nature's source had been used is more than the replenished. The widely method used to assess the environmental sustainability are Emergy Evaluation (EME) and Ecological Footprint Analysis (EFA). The outcomes of the assessment are focused on resources depletion, consumption patterns, waste production and absorption (Marchetinni et al., 2007). Environmental impact is measured by the emergy investment ratio defined as the ratio of the emergy purchased from the economy divided by the emergy from the local environment (Odum, 1998). Ecological footprint analysis compares human demand on nature with the biosphere's ability to regenerate resources and provide services. It is done by assessing the biologically productive land and marine area required to produce the resources a population consumes and absorb the corresponding waste using prevailing technology (Eco Greenwares, 2009). Per capita ecological footprint (EF) is comparing consumption and lifestyles and checking this against nature's ability to provide for this consumption (Cui & Yu, 2009).

Performance in handling environmental policy categories is another perspective in environmental assessment.

Environmental Performance Index (EPI) which ranks 132 countries on ten policy categories covering both environmental public health and ecosystem vitality is one of the popular measures in assessing environmental performance of a country. This index had been conducted by The Yale Center for Environmental Law and Policy (YCELP) and the Center for Earth Information Science Information Network (CIESIN) at Columbia University. These indices provide a gauge at a national government scale on how close countries are to established environmental policy goals (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2012). Each policy categories is made up of one or more environmental indicators. For each country and indicator, a proximity-to-target value is calculated based on the gap between a country's current result and the policy target. The generic formula for the proximity-to-target in the context of the global EPI is calculated using the following distance to target formula.

((international range) – (distance to target)) / (international range) x 100

The EPI is based on a proximity-to-target methodology whereby each country's performance on any given indicator is measured based on its position within a range established by the lowest performing country, equivalent to 0 on a 0-100 scale and the target, equivalent to 100. The illustration of methodology is shown in Figure 1. The words "Better" and "Worse" are relative terms and refer to the distance to the target.

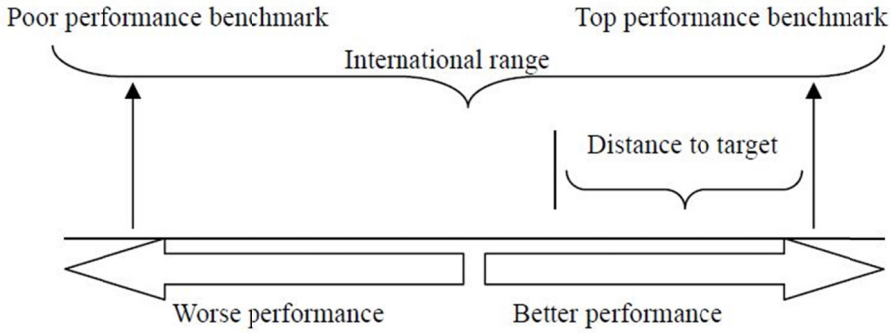



Figure 1. Framework of proximity-to-target methodology (Emerson et al., 2012)

All values of proximity-to-target indicators were summed and averaged. Data selection was made from official statistics reported by governments, spatial data, observing from monitoring stations and from modelled data (Emerson et al., 2012). It can be seen that EPI use a simple average calculation thereby may neglect some extreme values in the data.

With the advent of computing based technology, Liqian and Jianming (2012) conducted an eco-environmental quality assessment of Xining city based on geographic information system (GIS) and analytic hierarchy process (AHP). The assessment result marks with colours using GIS, which show the qualities; red for stop and green for pass. Also, the assessment conducted ranked the factors influenced the eco-environment in Xining area using weight obtained in AHP. Social cycle is a dominant factors eco-environment in Xining area followed by hydrosphere, atmosphere, lithosphere and biosphere. Assessment of environmental issues was also gained attention by intelligent methods based on fuzzy sets theory. Silvert (2000), for example, found that fuzzy logic can be used to classify and quantify environmental effects of a subjective nature, such as bad odours and it even provides formalism for dealing with missing data. The fuzzy memberships not only can be used as environmental indices but it is also possible to obtain a more traditional type of index through defuzzification. The fuzzy methodology also used to evaluate of the effects of finfish mariculture on coastal zone water quality. In line with the development of fuzzy sets theory, the dual assessment concept intuitionistic fuzzy sets were hybridized with AHP. This integrated approach which later known as IF-AHP was meant to handle both vagueness and ambiguity related uncertainties in the environment decision-making process. Sadiq and Tesfamariam (2009) used IF-AHP methodology to select best drilling fluid (mud) for drilling operations under multiple environmental-based criteria like air emissions, spills, water column, bioaccumulation, benthic effects, air emission and ground water contamination.

In another attempt to further proliferate the AHP in environmental assessment, Wan-Ismail and Abdullah (2012) proposed a new environmental index. The method used pair wise comparison scale in analytic hierarchy process to set a new EPI for ASEAN countries. A comparison scale was given using a pair-wise comparison scale for AHP preference introduced by Saaty and Windy (1980). A weight of each policy category was considered as an important element prior to proposing overall index. Despite this success, the AHP has its own weaknesses. While

using AHP, the decision problem is decomposed into a number of subsystems in which substantial number of pair wise comparisons need to be completed. Number of pair wise comparisons to be made may become very large depending on the size of matrix. The relation (n(n−1)/2) where n is size of decision matrix clearly lead to a lengthy task (Macharis et al., 2004). Another disadvantage of the AHP method is the artificial limitation of the use of the 9-point scale. Sometimes the decision maker might find difficult to distinguish among them. Also, the AHP method cannot cope with the fact that an alternative is 25 times more important than another alternative (Belton & Gear, 1983; Belton, 1986).

As an effort to overcome these weaknesses, a new approach in calculating environmental performance index is proposed. The proposed method for measuring environmental performance is taken into account weight of each policy category. The method was originally proposed by Ye (2010) as a method in fuzzy decision-making method based on the weighted correlation coefficient. To the best of authors' knowledge, this dual assessment method has not been tested to environmental performance. The weight of each policy category and two-sided memberships of intuitionistic fuzzy sets (IFSs) is considered as an important characteristic in this calculation. The aim of this paper is to propose weights for policy categories and subsequently propose a new rank of EPI. Nine ASEAN countries are tested to the weighted correlation coefficient method as an alternative method in calculating new EPI. The paper unfolds as follows. The next section briefly introduces some definitions related to the method. Section 3 describes the weighted correlation coefficient in IFSs proposed by Ye (2010). Section 4 presents weights for policy categories and a new ranking of EPI. Discussions section presents a new ranking of EPI. A comparison between original EPI and related study is made in this section. Conclusions appear in the last section.

## 2. Preliminaries

This section introduces some definitions that self-contained to the paper.

**Definition 1** Fuzzy set theory (Zadeh, 1965).

A fuzzy set theory $A$ in the universe of discourse $X = \{x_1, x_2, ..., x_n\}$ is defined as:

$$A = \{< x, \mu_A(x) > | \, x \in X\}, \tag{1}$$

which is characterized by membership function $\mu_A(x) : X \to [0,1]$, where $\mu_A(x)$ indicates the membership degree of the element $x$ to the set $A$.

**Definition 2** Intuitionistic fuzzy sets (Atanassov, 1986).

An IFS in $X$ is an expression $A$ is defined by

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\}, \tag{2}$$

where $\mu_A(x) : X \to [0,1]$ and $\nu_A(x) : X \to [0,1]$, with the condition $0 \le \mu_A(x) + \nu_A(x) \le 1$. The numbers $\mu_A(x)$ and $\nu_A(x)$ represent respectively the membership degree and non-membership degree of the element $x$ to the set $A$. For each IFS in $X$:

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x), \tag{3}$$

for all $x \in X$. Then $\pi_A(x)$ is called the intuitionistic index or hesitancy degree of the element $x$ in the set $A$. It can be seen that $0 \le \pi_A(x) \le 1, x \in X$.

For two IFS $A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\}$ and $B = \{< x, \mu_B(x), \nu_B(x) > | x \in X\}$ the two relations are follows:

(1) $A \subseteq B$ if and only if $\mu_A(x) \le \mu_B(x)$ and $\nu_A(x) \le \nu_B(x)$ for any $x \in X$;

(2) $A = B$ if and only if $\mu_A(x) = \mu_B(x)$ and $\nu_A(x) = \nu_B(x)$ for any $x \in X$.

**Definition 3** Correlation coefficient of IFSs (Gerstenkorn & Manko, 1991)

Let $A$ and $B$ be two IFSs in the universe of discourse $X = \{x_1, x_2, ..., x_n\}$. The correlation coefficient of $A$ and $B$ is given by

$$k(A,B) = \frac{C(A,B)}{\sqrt{T(A).T(B)}}, \tag{4}$$

where the correlation of two IFSs $A$ and $B$ is given by $C(A,B) = \sum_{i=1}^{n} (\mu_A(x_i)\mu_B(x_i) + \nu_A(x_i)\nu_B(x_i))$ and the informational intuitionistic energies of

two IFSs $A$ and $B$ are given by $T(A)=\sum_{i=1}^{n}(\mu_A^2(x_i)+\nu_A^2(x_i))$ and $T(B)=\sum_{i=1}^{n}(\mu_B^2(x_i)+\nu_B^2(x_i))$, respectively. The correlation coefficient of two IFSs $A$ and $B$ satisfies the following properties:

(1) $0\le k(A,B)\le 1$,

(2) $k(A,B)=k(B,A)$,

(3) $k(A,B)=1$ if $A=B$.

**Definition 4** Intuitionistic entropy of IFSs (Burillo & Bustince, 1996)

Let $A$ be an IFS in the universe of discourse $X=\{x_1,x_2,...,x_n\}$. The intuitionistic fuzzy entropy of an IFS $A$ is defined as follows:

$$E(A)=\sum_{i=1}^{n}(1-\mu_A(x_i)-\nu_A(x_i))=\sum_{i=1}^{n}\pi_A(x_i). \tag{5}$$

## 3. Method and Material

### *3.1 Weighted Correlation Coefficient (Ye, 2010)*

Based on the correlation coefficient between IFSs proposed by Gernstenkorn and Manko (1991), the correlation coefficient between an alternative $A_i$ and the ideal alternative $A^*$ with entropy weights for criteria can be measured by the weighted correlation coefficient $W_i(i=1,2,...,m)$

$$W_i(A^*,A_i)=\frac{\sum_{j=1}^{n}w_j\mu_A(C_j)}{\sqrt{\sum_{j=1}^{n}w_j(\mu_{A_i}^2(C_j)+\nu_{A_i}^2(C_j))}}. \tag{6}$$

If the information about weight $w_j$ of the criterion $C_j(j=1,2,....,n)$ is completely unknown, then the model of entropy weights is applied to determine the criteria weight:

$$w_j=\frac{1-H_j}{n-\sum_{j=1}^{n}H_j}, \tag{7}$$

where $w_j\in[0,1],\sum_{j=1}^{n}w_j=1,H_j=\frac{1}{m}E(C_j)=\frac{1}{m}\sum_{i=1}^{m}(1-\mu_{A_i}(C_j)-\nu_{A_i}(C_j))=\frac{1}{m}\sum_{i-1}^{m}\pi_A(C_j)$ and $0\le H_j\le 1$ for $(j=1,2,....,n)$.

The larger the value of weighted correlation coefficient $W_i$, the better alternative $A_i$, as the alternative $A_i$, is closer to the alternative $A^*$. Therefore, all the alternatives can be ranked according to the weighted correlation coefficients so that the best alternatives can be selected. The weighted correlation coefficient $W_i(i=1,2,...,m)$ of $A_i$ and $A^*$ have the following properties:

(1) $0\le W_i(A^*,A_i)\le 1$,

(2) $W_i(A^*,A_i)=W_i(A_i,A^*)$,

(3) $W_i(A^*,A_i)=1$ if $A^*=A_i$.

### *3.2 Environmental Performance Index 2012*

Environmental Performance Index 2012 and ranking issued by Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network (2012) among nine ASEAN countries are given in Table 1. Nine ASEAN countries are indicated as Ai (i=1, 2, 3, 4, 5, 6, 7, 8, 9).

Table 1. Environmental performance index (EPI) 2012

| Ranking | Countries ($A_i$) | EPI Score |
|---|---|---|
| 1 | Malaysia ($A_1$) | 62.5 |
| 2 | Brunei ($A_2$) | 62.5 |
| 3 | Thailand ($A_3$) | 60.0 |
| 4 | Philippines ($A_4$) | 57.4 |
| 5 | Singapore ($A_5$) | 56.4 |
| 6 | Cambodia ($A_6$) | 55.3 |
| 7 | Myanmar ($A_7$) | 52.7 |
| 8 | Indonesia ($A_8$) | 52.3 |
| 9 | Vietnam ($A_9$) | 50.6 |

Data of the ten policy categories, $c_i$ (i=1, 2, 3, 4, 5, 6, 7, 8, 9, 10) were retrieved as secondary data in proposing a new ranking of EPI. Environmental burden of disease, air pollution (impact on humans), water (impact on humans), air pollution (impact on ecosystem), water (impact on ecosystem), biodiversity, forestry, fisheries, agriculture and climate change were the nine policy categories. Table 2(a) and 2(b) shows the data for ASEAN countries extracted from Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network (2012).

Table 2(a). Index each policy categories

| ASEAN Countries | Environmental Burden of Disease, $c_1$ | Air Pollution (impact on humans), $c_2$ | Water Pollution (impact on humans), $c_3$ | Air Pollution (impact on ecosystem), $c_4$ | Water Pollution (impact on ecosystem), $c_5$ |
|---|---|---|---|---|---|
| Malaysia | 80.6 | 97.3 | 82.6 | 41.5 | 48.4 |
| Brunei | 86.4 | 100.0 | 38.2 | 37.1 | 99.6 |
| Thailand | 87.6 | 40.3 | 70.0 | 42.9 | 18.2 |
| Philippines | 58.0 | 55.4 | 38.9 | 39.1 | 36.4 |
| Singapore | 100.0 | 100.0 | 100.0 | 31.2 | 14.5 |
| Cambodia | 35.7 | 42.0 | 11.6 | 64.4 | 45.3 |
| Myanmar | 40.7 | 33.8 | 28.7 | 70.2 | 50.9 |
| Indonesia | 57.7 | 54.3 | 23.1 | 38.9 | 46.7 |
| Vietnam | 42.5 | 31.0 | 42.5 | 43.8 | 37.8 |

Table 2(b). Index each policy categories (continuation from Table 2(a))

| ASEAN Countries | Biodiversity, $c_6$ | Forestry, $c_7$ | Fisheries, $c_8$ | Agriculture, $c_9$ | Climate Change, $c_{10}$ |
|---|---|---|---|---|---|
| Malaysia | 90.1 | 17.4 | 31.0 | 95.5 | 28.0 |
| Brunei | 90.7 | 66.7 | 67.6 | 44.2 | 5.2 |
| Thailand | 78.9 | 87.0 | 34.2 | 93.9 | 39.2 |
| Philippines | 66.0 | 90.1 | 25.8 | 92.4 | 64.7 |
| Singapore | 34.1 | 79.4 | 18.4 | 98.5 | 28.3 |
| Cambodia | 94.8 | 28.3 | 21.6 | 66.7 | 73.9 |
| Myanmar | 53.6 | 26.3 | 33.3 | 84.8 | 77.3 |
| Indonesia | 75.3 | 54.7 | 38.1 | 54.6 | 48.9 |
| Vietnam | 54.1 | 81.4 | 19.4 | 47.8 | 56.5 |

The index for each policy category shall be converted into intuitionistic fuzzy sets (IFSs) notation to fit with the weighted correlation coefficient formula.

*3.3 Computational Method*

The summary of computations procedure is given in Figure 2.

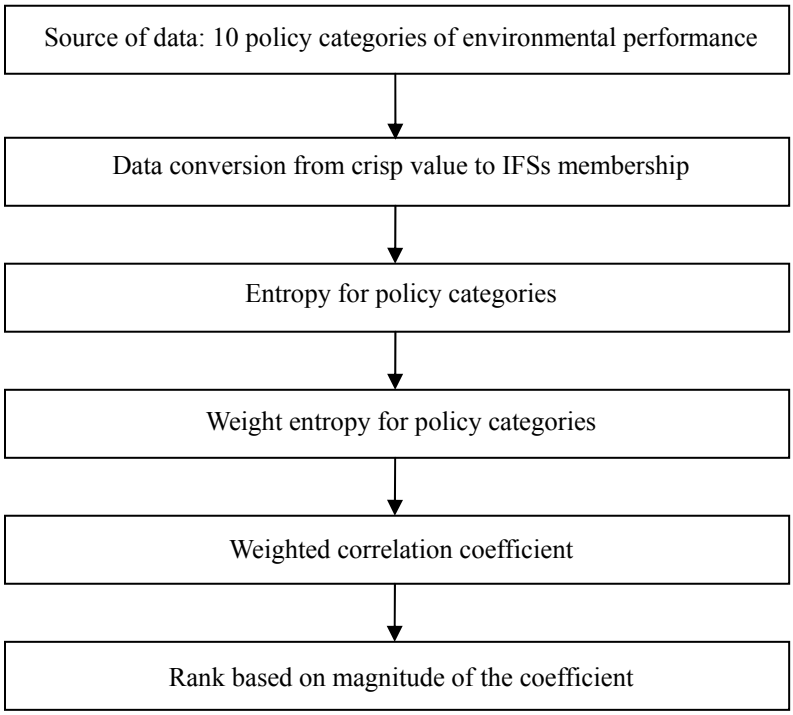



Figure 2. Flows of the computation

The computational framework is implemented with the data of policy categories of environmental performance.

**4. Implementation and Results**

The process of calculating the weighted correlation coefficient for ASEAN countries can be divided into two phases. The first phase is converting the entire index for policy categories into IFSs form. Then, in the second phase, weighted correlation coefficients of IFSs in ASEAN countries are calculated. A new ranking based on environmental performance can be obtained using the following steps:

Step 1 Convert to interpretation score, $\mu''(x)$.

Let's take data of Malaysia as an example in this calculation. The original data of environmental burden of disease policy category are converted to IFSs. Since the maximum value of is 100%, then the interpretation score is 80.6% (see Table 2(a)).

Step 2 Determine the value of hesitation, $\pi(x)$.

Levels of consistency and membership grades in Table 3 are used to determine value of hesitation.

Table 3. Conversion of consistency expressions to membership grades (Hersh, 2006)

| Consistency | $\mu''(x)$ | $\pi(x)$ |
|---|---|---|
| No or very low consistency | 0.0-0.2 | 0.8-1.0 |
| Low consistency | 0.2-0.4 | 0.6-0.8 |
| Moderate consistency | 0.4-0.6 | 0.4-0.6 |
| High consistency | 0.6-0.8 | 0.2-0.4 |
| Very high or total consistency | 0.8-1.0 | 0.0-0.2 |

If the value of $\mu''(x) = 1$ then $\pi(x) = 0$ and if $\mu''(x) = 0.9$ then $\pi(x) = 0.1$. From Step 1, $\mu''(x) = 0.806$, then hesitation value is 0.2, in the level of very high consistency.

Step 3 Calculate the value of membership, $\mu(x)$.

The value of membership, $\mu(x)$ is calculated using the equation $\mu(x) = \mu''(x)[1-\pi(x)] = 0.806[1-0.2] = 0.6448$.

Step 4 Calculate the value of non membership, $\nu(x)$.

The value of non membership, $\nu(x)$ calculate using Equation (3), $\nu(x) = 1-\mu(x)-\pi(x) = 1-0.6448-0.2 = 0.1552$.

Step 5 Arrange the memberships in the IFSs notation.

$$C_1 = (\mu(x), \nu(x), \pi(x))$$

$$= (0.6448, 0.1552, 0.2)$$

Memberships for other policy categories of ASEAN countries are calculated with the similar fashion. It is listed in Table 4(a), Table 4(b) and Table 4(c).

Table 4(a). IFS of policies for ASEAN countries

| Countries | Environmental Burden of Disease (DALYs), $c_1$ | Air Pollution (impact on humans), $c_2$ | Water (impact on humans), $c_3$ | Air Pollution (impact on ecosystem), $c_4$ |
|---|---|---|---|---|
| Malaysia | (0.6448, 0.1552, 0.2) | (0.9730, 0.0270, 0.0) | (0.6608, 0.1392, 0.2) | (0.1660, 0.2340, 0.6) |
| Brunei | (0.7776, 0.1224, 0.1) | (1.0000, 0.0000, 0.0) | (0.1528, 0.2472, 0.6) | (0.1484, 0.2516, 0.6) |
| Thailand | (0.7884, 0.1116, 0.1) | (0.1612, 0.2388, 0.6) | (0.4900, 0.2100, 0.3) | (0.1716, 0.2284, 0.6) |
| Philippines | (0.3480, 0.2520, 0.4) | (0.3324, 0.2676, 0.4) | (0.1556, 0.2444, 0.6) | (0.1564, 0.2436, 0.6) |
| Singapore | (1.0000, 0.0000, 0.0) | (1.0000, 0.0000, 0.0) | (1.0000, 0.0000, 0.0) | (0.0936, 0.2964, 0.7) |
| Cambodia | (0.1428, 0.2572, 0.6) | (0.1680, 0.2320, 0.6) | (0.0116, 0.0884, 0.9) | (0.3864, 0.2136, 0.4) |
| Myanmar | (0.1628, 0.2372, 0.6) | (0.1014, 0.1986, 0.7) | (0.0861, 0.2139, 0.7) | (0.4914, 0.2086, 0.3) |
| Indonesia | (0.3462, 0.2538, 0.4) | (0.2715, 0.2285, 0.5) | (0.0462, 0.1538, 0.8) | (0.1556, 0.2444, 0.6) |
| Vietnam | (0.1700, 0.2300, 0.6) | (0.0930, 0.2070, 0.7) | (0.1700, 0.2300, 0.6) | (0.1752, 0.2248, 0.6) |

Table 4(b). IFS of policies for ASEAN countries (continuation from Table 4(a))

| Countries | Water (impact on ecosystem), $c_5$ | Biodiversity, $c_6$ | Forestry, $c_7$ |
|---|---|---|---|
| Malaysia | (0.2420, 0.2580, 0.5) | (0.8109, 0.0891, 0.1) | (0.0348, 0.1652, 0.8) |
| Brunei | (0.9960, 0.0040, 0.0) | (0.8163, 0.0837, 0.1) | (0.4669, 0.2331, 0.3) |
| Thailand | (0.0364, 0.1636, 0.8) | (0.6312, 0.1688, 0.2) | (0.7830, 0.1170, 0.1) |
| Philippines | (0.1456, 0.2544, 0.6) | (0.4620, 0.2380, 0.3) | (0.8109, 0.0891, 0.1) |
| Singapore | (0.0145, 0.0855, 0.9) | (0.1023, 0.1977, 0.7) | (0.6352, 0.1684, 0.2) |
| Cambodia | (0.2265, 0.2735, 0.5) | (0.8532, 0.0468, 0.1) | (0.0849, 0.2151, 0.7) |
| Myanmar | (0.2545, 0.2455, 0.5) | (0.2680, 0.2320, 0.5) | (0.0789, 0.2211, 0.7) |
| Indonesia | (0.2335, 0.2665, 0.5) | (0.6024, 0.1976, 0.2) | (0.2735, 0.2265, 0.5) |
| Vietnam | (0.1512, 0.2488, 0.6) | (0.2705, 0.2295, 0.5) | (0.6512, 0.1488, 0.2) |

Table 4(c). IFS of policies for ASEAN countries (continuation from Table 4(c))

| Countries | Fisheries, $c_8$ | Agriculture, $c_9$ | Climate Change, $c_{10}$ |
|---|---|---|---|
| Malaysia | (0.0930, 0.2070, 0.7) | (0.9550, 0.0450, 0.0) | (0.0840, 0.2160, 0.7) |
| Brunei | (0.4732, 0.2268, 0.3) | (0.1768, 0.2232, 0.6) | (0.0052, 0.0948, 0.9) |
| Thailand | (0.1026, 0.1974, 0.7) | (0.8451, 0.0549, 0.1) | (0.1568, 0.2432, 0.6) |
| Philippines | (0.0774, 0.2226, 0.7) | (0.8316, 0.0684, 0.1) | (0.3882, 0.2118, 0.4) |
| Singapore | (0.0368, 0.1632, 0.8) | (0.9850, 0.0150, 0.0) | (0.0849. 0.2151, 0.7) |
| Cambodia | (0.0432, 0.1568, 0.8) | (0.4669, 0.2331, 0.3) | (0.5173, 0.1827, 0.3) |
| Myanmar | (0.0999, 0.2001, 0.7) | (0.6784, 0.1216, 0.2) | (0.6184, 0.1816, 0.2) |
| Indonesia | (0.1524, 0.2476, 0.6) | (0.2730, 0.2270, 0.5) | (0.2445, 0.2555, 0.5) |
| Vietnam | (0.0388, 0.1612, 0.8) | (0.2390, 0.2610, 0.5) | (0.3390, 0.2610, 0.4) |

Step 6 Calculate the weight of criteria.

Since the weight $w_1$ of the criterion $C_j(j=1,2,3,4,5,6,7,8,9,10)$ is completely unknown, the entropy weights are calculated using Equation (7).

$$\mathrm{H}_1 = \frac{1}{9}(0.2+0.1+0.1+0.4+0+0.6+0.6+0.4+0.6) = 0.3333,$$

$$\sum_{j=1}^{10}\mathrm{H}_j = \frac{1}{9}\mathrm{H}_1 + \frac{1}{9}\mathrm{H}_2 + \ldots + \frac{1}{9}\mathrm{H}_{10}$$

$$= \frac{1}{9}(0.2+0.1+0.1+0.4+0+0.6+0.6+0.4+0.6) + \frac{1}{9}(0+0+0.6+0.4+0+0.6+0.7+0.5+0.7)$$

$$+ \ldots + \frac{1}{9}(0.7+0.9+0.6+0.4+0.7+0.3+0.2+0.5+0.4)$$

$$= 4.5,$$

Weight of environmental burden of disease policy categories,

$$w_1 = \frac{1-0.3333}{10-4.5}$$

$$= 0.1212$$

The calculation for other policy categories is executed in the similar fashion. The weights for all ten policy categories are obtained as,

$$\mathrm{w}_1 = 0.1212, \mathrm{w}_2 = 0.1111, \mathrm{w}_3 = 0.0869, \mathrm{w}_4 = 0.0808, \mathrm{w}_5 = 0.0828,$$

$$\mathrm{w}_6 = 0.1273, \mathrm{w}_7 = 0.1091, \mathrm{w}_8 = 0.0586, \mathrm{w}_9 = 0.1354, \mathrm{w}_{10} = 0.0869.$$

Step 7 Calculate the weighted correlation coefficient.

The weighted correlation coefficient between an alternative $A_i$ and the ideal alternative $A^*$ with entropy weights for criteria, $W(A^*, A_i)(i=1,2,3,4,5,6,7,8,9)$ is measured using Equation (6). The weighted correlation coefficient $W(A^*, A_1)$ is:

$$W_i(A^*, A_1) = \frac{\sum_{j=1}^{10} w_j \mu_A(C_j)}{\sqrt{\sum_{j=1}^{10} w_j(\mu_{A_i}^2(C_j) + \nu_{A_i}^2(C_j))}}$$

$$= \frac{0.1212(0.6448) + 0.1111(0.9730) + \ldots + 0.0869(0.0840)}{\sqrt{0.1212(0.6448^2 + 0.1552^2) + 0.1111(0.9730^2 + 0.0270^2) + \ldots + 0.0869(0.0840^2 + 0.2160^2)}}$$

$$= 0.7983$$

The calculation is executed to other eight countries. Weighted correlation coefficients are listed in Table 5.

Table 5. Weighted correlation coefficient of ASEAN countries

| Countries ($A_i$) | $W(A^*, A_i)$ | New Ranking |
|---|---|---|
| Malaysia ($A_1$) | 0.7983 | 2 |
| Brunei ($A_2$) | 0.7969 | 3 |
| Thailand ($A_3$) | 0.7994 | 1 |
| Philippines ($A_4$) | 0.7778 | 4 |
| Singapore ($A_5$) | 0.7680 | 5 |
| Cambodia ($A_6$) | 0.6888 | 8 |
| Myanmar ($A_7$) | 0.6937 | 7 |
| Indonesia ($A_8$) | 0.7214 | 6 |
| Vietnam ($A_9$) | 0.6569 | 9 |

Table 5 shows the weighted correlation coefficients for each country. The larger value of weighted correlation coefficients indicates the better alternative in environmental performance.

## 5. Discussion

A new EPI ranking for ASEAN countries using weighted correlation coefficient is obtained. The new ranking was established after considering the weights in each policy category. The membership and non-membership values in IFSs are the main characteristics prior to introducing the weights. The memberships are embedded with fuzzy correlation coefficient to find the best alternatives. The best alternative represents the better country in environmental performance.

The results show that Thailand leads in the first place followed by Malaysia and Brunei while the last three countries are Myanmar, Cambodia and Vietnam. The obtained ranking is illustrated and compared with the original EPI 2012 and EPI 2012 using AHP. The new ranking and ranking from the other two methods are illustrated in Figure 3.

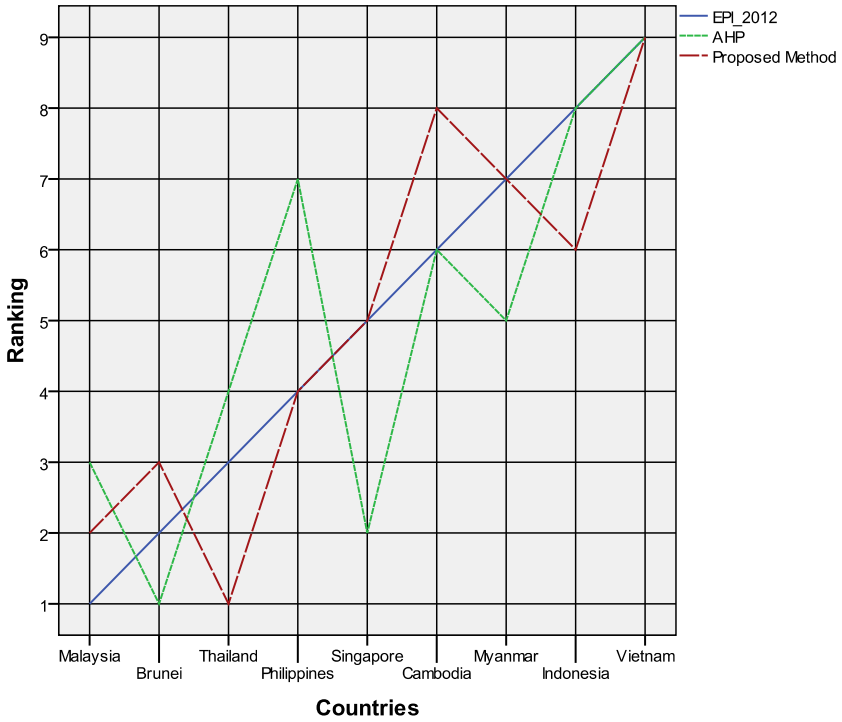



Figure 3. Ranking of EPI 2012 for ASEAN countries: comparison analysis

It can be seen that the new ranking differs from original EPI 2012 and EPI 2012 using AHP. The original EPI indicates that Malaysia is the first among ASEAN countries in environmental performance followed by Brunei and Thailand. However, Brunei takes the first spot when AHP method is used, followed by Singapore and Malaysia. Contrarily, Thailand leads in first spot followed by Malaysia and Brunei when the calculation is made using the weighted correlation coefficients. The inconsistency in ranking can be related due to membership and non-membership of IFSs.

## 6. Conclusion

This paper has shown the capability of decision tool weighted correlation coefficient of IFSs in proposing the new ranking of EPI 2012 among ASEAN countries. The original EPI is used arithmetic mean in calculating the ten policy categories. Neglecting some extreme values was among the weakness of this simple mathematical operation. Thus, weighted correlation coefficient based on IFSs was proposed. This mathematical model was successfully proposed weights in each of category policies of environmental performance. The advantage of IFSs lies on the two-sided measurements, membership and non membership values. A new ranking of EPI among ASEAN countries show that Thailand is at the first place followed by Malaysia and Brunei. The method could be extended to other countries as this method offer an alternative measure in evaluating environmental performance. As a further research, the weighted correlation coefficient may consider the extension version interval-valued intuitionistic fuzzy sets in the calculation. It might provide a comprehensive evaluation due to its interval values thereby offer a more reliable EPI.

## References

Atanassov, K. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 20*, 87-96. http://dx.doi.org/10.1016/S0165-0114(86)80034-3

Belton, V., & Gear, A. E. (1983). On a short-coming of Saaty's method of analytic hierarchies. *Omega, 11*, 228-230. http://dx.doi.org/10.1016/0305-0483(83)90047-6

Belton, V. (1986). A comparison of the analytic hierarchy process and a simple multi-attribute value function. *European Journal of Operational Research, 26*, 7-21. http://dx.doi.org/10.1016/0377-2217(86)90155-4

Burillo, P., & Bustince, H. (1996). Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets. *Fuzzy Sets and Systems, 19*, 305-316. http://dx.doi.org/10.1016/0165-0114(96)84611-2

Cui, J., & Yu, B. (2009). The regional resource supply and environmental capacity analysis based on the ecological footprint: A case study. *Modern Applied Science, 3*, 96-100.

Eco Greenwares. (2009). *Carbon and ecological footprints*. Eco Greenwares, Retrieved from http://ecogreenwares.com/biodegradable/compostable/carbon-footprints.html

Emerson, J. W., Hsu, A., Levy, M. A., Sherbinin, A. D., Mara, V., Esty, D. C., & Jaiteh, M. (2012). Environmental Performance Index and Pilot Trend Environmental Performance Index. Yale Center for Environmental and Policy, New Heaven.

Gernsternkorn, T., & Manko, J. (1991). Correlation of intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 44*, 39-43. http://dx.doi.org/10.1016/0165-0114(91)90031-K

Hersh, M. A. (2006). *Mathematical Modelling for Sustainable Development*. Germany: Springer-Verlag Berlin Heidelberg.

Macharis, C., Springael, J., Brucker, K. D., & Verbeke, A. (2004). Promethee and AHP: The design of operational synergies in multicriteria analysis. Strengthening Promethee with ideas of AHP. *European Journal of Operational Research, 15*, 307-317. http://dx.doi.org/10.1016/S0377-2217(03)00153-X

Marchettini, N., Niccolucci, V., Pulselli, F. M., & Tiezzi, E. (2007). Environmental sustainability and the integration of different method for its assessment. *Environmental Science and Pollution Research International, 14*, 227-228. http://dx.doi.org/10.1065/espr2006.12.367

Odum, H. T. (1998). Emergy evaluation. Presented at *the International Workshop on Advances in Energy Studies: Energy flows in ecology and economy*, Italy.

Saaty, T. L., & Windi, Y. (1980). Marketing applications of the analytic hierarchy process. *Management Science, 26*(7), 641-658. http://dx.doi.org/10.1287/mnsc.26.7.641

Sadiq, S., & Tesfamariam, S. (2009). Environmental decision-making under uncertainty using intuitionistic fuzzy analytic hierarchy process. *Stochastic Environmental Research and Risk Assessment, 23*, 75-91. http://dx.doi.org/10.1007/s00477-007-0197-z

Silvert, W. (2000). Fuzzy indices of environmental conditions. *Ecological Modelling, 130*, 111-119. http://dx.doi.org/10.1016/S0304-3800(00)00204-0

Wan-Ismail, W. K., & Abdullah, L. (2012). A new Environmental Performance Index using analytic hierarchy process: a case of ASEAN countries. *Environmental Skeptics and Critics, 1*(3), 39-47.

Yale Center for Enviromental Law and Policy and Center for International Earth Science Information Network. (2012). *Enviromental Performance Index*. Yale University. Retrieved from http://epi.yale.edu/epi2012/countryprofiles

Ye, J. (2010). Fuzzy decision-making method based on the weighted correlation coefficient under intuitionistic fuzzy environment. *European Journal of Operational Research, 205*, 202-204. http://dx.doi.org/10.1016/j.ejor.2010.01.019

Zadeh, L. A. (1965). Fuzzy sets. *Information Control, 8*(3), 338-353. http://dx.doi.org/10.1016/S0019-9958(65)90241-X

Zhang, L. Q., & Cai, J. M. (2012). Eco-environmental quality assessment of Xining City based on GIS and AHP. *Modern Applied Science, 6*(4), 84-100. http://dx.doi.org/10.5539/mas.v6n4p84

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# The Study of Robot Movement Inverse Solution Based on Genetic Algorithm

Shen Chao[1]

[1] University of Shanghai for Science and Technology, Shanghai, China

Correspondence: Shen Chao, College of Mechanical Engineering, University of Shanghai for Science and Technology, 334th, Jun Gong Road, Shanghai 200093, China. Tel: 86-137-6112-1824. E-mail: petershen88@126.com

Received: March 2, 2013 Accepted: April 24, 2013 Online Published: May 14, 2013

doi:10.5539/mas.v7n6p53 URL: http://dx.doi.org/10.5539/mas.v7n6p53

**Abstract**

Through the analysis of three degree of freedom of the industrial robot bar geometric parameter, the homogeneous transformation matrix method was used to establish the system of positive kinematics model. This paper will introduce genetic algorithm to solve the problem of inverse kinematics robot to plane three degree of freedom robot, to explain the genetic algorithm is applied to solve the inverse kinematics more effectively.

**Keywords:** robot, forward and inverse kinematics, genetic algorithm

## 1. Introduction

Based on the robot kinetic theory (Yu & Ma, 1999), we establish the robot kinematic model. According to the forward kinematics theory, under the conditions of known geometric parameters and the joint angle of the lever member, the position and attitude of the reference coordinate system relative to the manipulator end effector (Ma, 1991). Based on genetic algorithm, we put forward a new kind of algorithm controlling robot pose, and solve the inverse kinematics problem.

## 2. Establishment of Robot System Composition and Kinematics Model

### *2.1 Robot Kinematics Introduction (Jiang, 1994)*

Control of robot kinematics model (Chen & Zhang, 2005) is divided into two parts, forward kinematics and inverse kinematics. Solving the forward kinematics problem (Guo & Wang, 2012), a general method in engineering is homogeneous matrix transform method, presented by Denavit and Hartenberg. The core idea is build homogeneous transformation matrix for each robot joint rod, through successive matrix operations, which will use the robot manipulator end effector coordinate system representation of the vector to the base coordinate system. And the establishment of the corresponding relationship between the robot joint variables and rod of homogeneous coordinates.

### *2.2 Establishment of Robot Kinematics Model*

In order to describe the joint movement, the position and attitude of the reference coordinate system relative to the manipulator end effector, according to D-H member axes rule, each robot each rod in the joint shaft is set up by regular Descartes attached body coordinate system ($x_i$, $y_i$, $z_i$), where i=1, 2,..., n (n is the number of degrees of freedom), as Figure 1 shown below. As long as the $z_0$ axis along the first joint motion axis, the position and direction of the manipulator at 0 system on the base can be optionally. As long as the $x_n$ axis and $z_{i-1}$ axis is vertical, the last coordinates ($x_n$, $y_n$, $z_n$) can be placed in any position of the manipulator end. When the joint i movement, rod i relative to the rod member $i_{i-1}$ motion, on which the coordinates ($x_{i-1}$, $y_{i-1}$, $z_{i-1}$) relative to the rod member i-1 will also move, the coordinates ($x_{i-1}$, $y_{i-1}$, $z_{i-1}$) relative to the rod I-1 coordinate system ($x_{i-2}$, $y_{i-2}$, $z_{i-2}$) will also move.

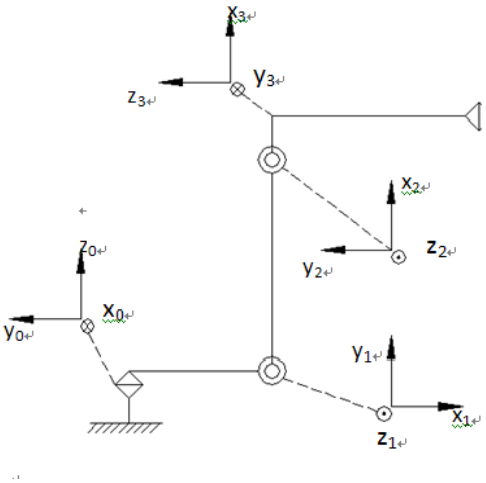



Figure 1. Robot link coordinate system

Based on the robot pole coordinate system, denoted by D-H method can be used to calculate the homogeneous transformation matrix connecting i coordinates and i-1 coordinates, a point $P_i$ represented by icoordinate can be available through coordinates i-1, just accomplished the transformation one by one:

(1) $x_{i-1}$ axis around the $z_{i-1}$ axis angle $\theta_i$, making it the same $x_i$ axis alignment;

(2) move along the $z_{i-1}$ axis translation distance $d_i$, the $x_{i-1}$ axis and $x_i$ axis coincide;

(3) move along the $x_i$ axis distance $a_i$, icoordinates and i-1 coordinates of the origin and the x axis coincide;

(4) around the $x_i$ axis angle of i, i coordinates and i-1 coordinates coincide completely.

With basic homogeneous rotation-translation matrix representation and multiplicative quadature of these four kind of action, we'll obtain the adjacent coordinates i and i-1 synthesis homogeneous transformation matrix $A_i$:

$^{i-1}A_i=Rot(z, \theta_i)Trans(0,0,d_i)Trans(a_i,0,0)Rot(x, a_i)$

Expand the above equation can be obtained:

$$^{i-1}A_i=\begin{bmatrix} \cos\theta_i & -\cos\alpha_i \sin\theta_i & \sin\alpha_i \sin\theta_i & \alpha_i \cos\theta_i \\ \sin\theta_i & \cos\alpha_i \cos\theta_i & -\sin\alpha_i \cos\theta_i & \alpha_i \sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

When the mechanical hand of each link coordinates were specified, will be able to list constant parameters in all links. The rotation joint, $a_i$, i and $d_i$ is a constant, $\theta_i$is the joint variables.

Using $^{i-1}A_i$ matrix, the robot arm on a fixed point $P_i$ in the i coordinate system of homogeneous coordinates and i-1 coordinates in the homogeneous coordinates of link, have equation:

$$P_{i-1}= (x_{i-1}, y_{i-1}, z_{i-1}, 1)^T, Pi= (x_i, y_i, z_i, 1)^T.$$

Establishment of robot forward kinematics model, as shown below:

The table shows the link parameters, each connecting transformation matrix can be obtained as follows:

$$^0T_1=\begin{bmatrix} c\theta_1 & -s\theta_1 & 0 & 330c\theta_1 \\ s\theta_1 & c\theta_1 & 0 & 330s\theta_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad ^1T_2=\begin{bmatrix} -s\theta_2 & -c\theta_2 & 0 & -300s\theta_2 \\ c\theta_2 & s\theta_2 & 1 & 300c\theta_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad ^2T_3=\begin{bmatrix} s\theta_3 & c\theta_3 & 0 & 150s\theta_3 \\ -c\theta_3 & s\theta_3 & 0 & -150c\theta_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Wherein, $^0T_1$ said connecting rod 1 coordinate {1} relative to the pose of base coordinate system {0}, $^1T_2$、$^2T_3$ and so on, $c\theta_1$, $s\theta_1$ represent $\sin\theta_1$, $\cos\theta_1$, and the rest of the analog.

Design of manipulator end pole coordinate system to the robot coordinate system homogeneous transformation matrix, then the

$$^{0}T_{3} = {}^{0}A_{1}\,{}^{1}A_{2}\,{}^{2}A_{3}$$

Any of the known robot joint angle vector substituting into the above formula, we can find the homogeneous coordinates values of robot manipulator end relative to the respective reference coordinate system. Establish the corresponding relationship between robot end-effector position and attitude and the joint variable value. In the robot joint constraints, given any set of joint variables, through the forward kinematics equation uniquely determine the end-effector pose.

## 3. Robot Gripping Control Algorithm

On the basis of establishment of robot kinematic model in the above context, this chapter is based on the genetic algorithm robot gripping control algorithm, analyze solution of the robot inverse kinematics, illustrate specific implementation process about genetic algorithmic and a planar robot with three degrees of freedom.

### *3.1 Inverse Kinematics of the Robot*

The robot kinematics equation can be obtained from the equations of robot motion:

$$^{0}\mathrm{T}_{3}=\begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = {}^{0}T_{1}(\theta_1)\,{}^{1}T_{2}(\theta_2)\,{}^{2}T_{3}(\theta_3)$$

If the robot end effector connecting rod pose has been given, i.e. n, o, a and p were known, then seek the joint variables $\theta_1$, $\theta_2$, $\theta_3$, the value of them is called the inverse solution.

Robot inverse kinematics problem is when the position, attitude, velocity, angular velocity, acceleration and angular acceleration of the robot end effector rod are presented, and can achieve these requirements of the joint position, velocity and acceleration. This article only discusses the problem in the first place, which is known as satisfy some work requirements of the spatial position and posture A of end effector, and each pole structure parameters, and the robot joint variables can be calculated.

### *3.2 Genetic Algorithm*

Genetic algorithm referred to as GA, is a kind of evolutionary algorithm first proposed by the United States of America professor J. H. Holland in 60s later period, it's the most widely used. Genetic algorithm is a kind of optimization search algorithm draw from biological natural election and natural genetic mechanisms, is an approximation algorithm.

Genetic algorithm (Zhang, 1995) the term is derived from the nature genetics. An alternative solution is called a chromosome, each chromosome consists of several genes, each gene can be used to represent a number, a number of chromosomes compose a group. Genetic algorithm (Wang & Wei, 1996) is an alternative solution for iterative computation process, and each iteration called generation. At the completion of iteration, the current group will make use of the certain evaluation function performance evaluation, and on the basis of the evaluation of new generation. Initial groups randomly determined to subjective experience, groups of fixed capacity for N, specific algorithm is shown by the following:

1) t = 0;

2) generate an initial group of G (t);

3) with a certain evaluation function to evaluate G (t);

4) satisfy the end condition, transferred to 8);

5) produce a new generation of G (t + 1) by G (t);

6) to evaluate G (t + 1);

7) back to 4);

8) stop;

The above iterative compute process to find a satisfying solution or to preset number of iterations. Algorithm of the evaluation function is mainly used to evaluate each of the chromosomes in the current generation, by evaluating the next generation of removed a number of low performance of chromosomes, keep some high-performance chromosomes, and through genetic operators added some new chromosomes, to ensure that the next generation of group contains new information, make its average performance is improved, the result was

very good, to meet the requirements of the problem solving.

In general, GA is not a simple optimization algorithm, it is not a traditional deterministic calculation tool, but it is a kind of new optimization algorithm that based on the evolutionary ideas, and it's a powerful tool to solve the problem of large-scale complex.

*3.3 Planar Robot Example*

One, parameters of robot connecting links.

Take a three planar robot as an example here. Link parameters are shown in Table 1.

Table 1. The three connecting rod mechanical parameters

| Angle | $\alpha_{i-1}$ | $\theta_i$ | $l_i$ | $\alpha_i$ |
|---|---|---|---|---|
| 1 | 0 | $\theta_1$ | $l_1$ | 0 |
| 2 | 0 | $\theta_2$+90 | $l_2$ | 0 |
| 3 | 0 | $\theta_3$-90 | $l_3$ | 0 |

Two, the derivation of end effector position.

According to the link parameter definition can get transformation matrix:

$$ {}^{i-1}A_i = \begin{bmatrix} c_i & -s_i & 0 & a_{i-1} \\ s_i & c_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

The position matrix of end effector:

$$ {}^0T_3 = {}^0T_1\,{}^1T_2\,{}^2T_3\,{}^3T_E = \begin{bmatrix} c_{123} & -s_{123} & 0 & c_{123}l_3 + c_{12}l_2 + c_1l_1 \\ s_{123} & c_{123} & 0 & s_{123}l_3 + s_{12}l_2 + s_1l_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

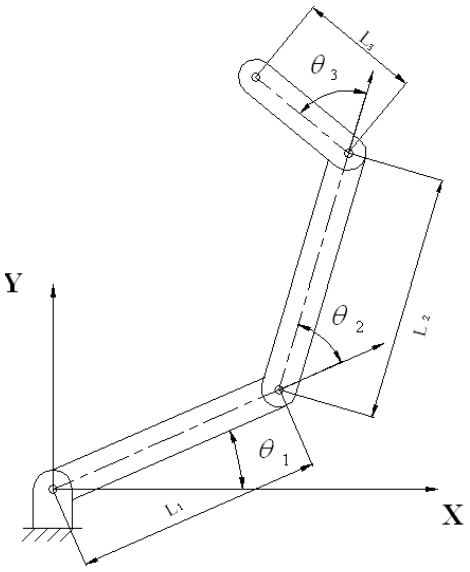



Figure 2. Diagram of three lever operating machine

At the end of three lever operating machine, the X, y coordinates can be expressed as:

$$x = l_1 \cos\theta_1 + l_2 \cos(\theta_1 + \theta_2) + l_3 \cos(\theta_1 + \theta_2 + \theta_3)$$
$$y = l_1 \sin\theta_1 + l_2 \sin(\theta_1 + \theta_2) + l_3 \sin(\theta_1 + \theta_2 + \theta_3)$$

The parameters of an operating machine of the rod respectively is $l_1$=330, $l_2$=300, $l_3$=150, firstly, the control parameters of genetic algorithm are as follows: group capacity N, genetic algebra G, crossover probability C, mutation probability M, according to the number of variables, determine the chromosome number of 3, vector situation expressed as [$\theta_1$, $\theta_2$, $\theta_3$] range of each variable is [-180°, 180°], in the absence of prior knowledge $\theta_1$, $\theta_2$, $\theta_3$ in their range, random generate chromosome N [$\theta_{1I}$, $\theta_{2I}$, $\theta_{3i}$], i=1, 2,..., N, as the initial population.

Selection of genetic parameter control:

As the above genetic algorithm parameter selection, we know, population capacity of N affects the final characteristics and effectiveness of genetic algorithm. Small group capacity will make the convergence speed slow, but the large group capacity will increase computing workload of each generation, and when N greater than 100, the convergence of the improvement is not obvious, therefore recommend selection N=100.

Crossover probability determined the crossover frequency, large crossover probability means that each generation will have more new chromosome. Recommend C=0.9.

The mutation probability determines the probability of mutation operator.

Too large Mutation probability will lead to a good chromosome loss of information. While too small mutation probability will do no help to the improvement of the convergence of the algorithm. Generally it takes M=0.03-0.07.

N=100, C=0.9, M=0.04, genetic end condition (Shi, Chen, & Yao, 2009) is:

$$J = \sqrt{(x - x_0)^2 + (y - y_0)^2} < 0.1 \qquad (*)$$

Take the position coordinates of the end effector (651,421), and then follow the above genetic algorithm genetic operation, until it meets the end condition. Finally the minimum of the group J chromosomes is considered as each rod corner solutions for robot mechanical hands.

In the solution of genetic algorithm in this paper, in order to avoid the loss of information good chromosome, during each generation of genetic, ten optimal chromosome were selected and retained to the new generation, this also ensures that the performance of a new generation is at least better than that in later generation. In addition, mutation operation method has influence on convergence speed. In order to improve the convergence speed of the algorithm, the selected gene mutation in genetic will initially replace its value of 110% or 90% of the current number (probability 50%), after each increase of 10 generations, the changes in the value of a corresponding reduction in 0.99 or 1.01 times, this makes the search range from big to small, especially with the genetic algebra increases, the method helps to accelerate the search speed, when search performance cannot be improved significantly.

The calculation results are as follows:

The G 100 generation, =0.54;

The G 150 generation, =0.092.

As can be seen the condition (*), from this group, J value of the smallest chromosome is regarded as the desired solution, then the manipulator end-effector position coordinates is (651.06, 421.07), close to the target value (651, 421).

## 4. Conclusions

According to definition of robot bar geometric parameters, analyze link parameters in three degree of freedom industrial robot, using homogeneous transformation matrix method to set up the robot forward kinematics model.

Genetic algorithm is used to solve the inverse kinematics. The robot kinematics equation is highly nonlinear and strong coupling equations, the traditional optimization method for kinematic solving has two problems: firstly, it is too complex, algorithm efficiency is low; secondly, it is easy to fall into local optimization, it is difficult to find the global optimal solution. In this paper, genetic algorithm is introduced into the solution of inverse kinematics problem, proposed using real number coding to improve the computing efficiency of the algorithm and accuracy. With an example of the plane robot which has three degrees of freedom illustrate the application of genetic algorithm to the validity of robot inverse k-inematics solution. Genetic algorithm constitute a new intelligent control method, the research in robotics has opened up a broad field, also it is bound to the

development of intelligent robots.

**References**

Chen, L. P., & Zhang, Y. Q. (2005). *Dynamics analysis of mechanical system*. Beijing: Tsinghua university press.

Guo, Y. K., & Wang, F. Z. (2012). The analysis and comparison of regional pole assignment in the robot movement. *Computer tehnology and automation, 31*(4), 21-25.

Jiang, X. F. (1994). *Introduction of robotics*. pp. 221-229. Shenyang: Liaoning Science and Technology Press.

Ma, X. F. (1991). *Robot mechanism* (1st ed.). Beijing: Mechanical Industry Press.

Shi, L., Chen, T. J., & Yao, L. N. (2009). Theory and application of Intelligent control. *Shanghai Jiaotong University Press, 4*(1), 300-303.

Wang, H. J., & Wei, X. F. (1996). Genetic algorithm Overview. *Control theory and applications, 13*(6), 697-707.

Yu, D. T., & Ma, X. F. (1999). *Application of the industrial robot engineering*. Beijing: Metallurgical Industry Press.

Zhang, Y. Q. (1999). Genetic optimization techniques for beginners. *Journal of Kunming University of Science and Technology, 24*(3), 41-45

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# Using Swarm Intelligence to Optimize the Energy Consumption for Distributed Systems

Neil Bergmann[1], Yuk Ying Chung[2], Xiangrui Yang[2], Zhe Chen[2], Wei-Chang Yeh[3], Xiangjian He[3] & Raja Jurdak[4]

[1] School of Information Technology and Electrical Engineering, The University of Queensland, Australia

[2] School of Information Technologies, University of Sydney, Australia

[3] University of Technology Sydney, Australia

[4] CSIRO ICT Centre, Pullenvale, Australia

Correspondence: Yuk Ying Chung, School of Information Technologies, University of Sydney, NSW 2006, Australia. Tel: 61-2-9036-9109. E-mail: vchung@it.usyd.edu.au

Received: November 16, 2012 Accepted: April 19, 2013 Online Published: May 21, 2013

doi:10.5539/mas.v7n6p59 URL: http://dx.doi.org/10.5539/mas.v7n6p59

**Abstract**

Large, distributed, network-based computing systems (also known as Cloud Computing) have recently gained significant interest. We expect significantly more applications or web services will be relying on network-based servers, therefore reducing the energy consumption of these systems would be beneficial for companies to save their budgets on running their machines as well as cooling down their infrastructures. Dynamic Voltage Scaling can save significant energy for these systems, but it faces the challenge of efficient and balanced parallelization of tasks in order to maximize energy savings while maintaining desired performance levels. This paper proposes our Simplified Swarm Optimization (SSO) method to reduce the energy consumption for distributed systems with Dynamic Voltage Scaling. The results of SSO have been compared to the most popular evolutionary Particle Swarm Optimization (PSO) algorithm and have shown to be more efficient and effective, reducing both the execution time for scheduling and makespan.

**Keywords:** energy optimization, evolutionary algorithm, distributed computing

## 1. Introduction

According to recent research on energy consumption, the electricity usage on servers in U.S. in 2005 represents 0.6% of the total electricity consumption of the whole country, and the number goes to 1.2% when the cooling infrastructures are also included (Koomey, 2008). In the past few years, out of the interest of large scalability, cost efficiency and performance, the concept of cloud computing has became very popular among companies that require huge amounts of computation, and nowadays the rapid growth of both experimental and commercial cloud services has brought a significant influence on building large data centers, web services, and the whole Internet, together with a tremendous growth of energy consumption on these machines.

Dynamic Voltage Scaling (Lee & Zomaya, 2009) is a power management technique that can optimize the energy efficiency on distributed systems with or without influencing the overall performance depending on our goals. In order to perform dynamic voltage scaling effectively, we need to find out the critical path in the task graph. Unfortunately, the problem of finding out the critical path itself is NP-Complete (Garey & Johnson, 1990), which means that the solutions can be hard to compute in a reasonable amount of time. In this paper, we propose Simplified Swarm Optimization (SSO) to achieve better performance. The results have been compared with the most popular evolutionary algorithm Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995). We explain the design of our experiments, how the algorithms work and why would they be effective in solving distributed computing problems.

To evaluate the performance, we have focused on three parameters: makespan, time and energy consumption. The evaluation approach will be introduced in Sections 2 and 3, and the experimental results and analysis will be presented in Section 4.

## 2. Experiment Design

### *2.1 Overview*

A distributed computing task can be presented as a directed graph with weights on each node indicating the amount of computation that is needed to complete each task respectively.

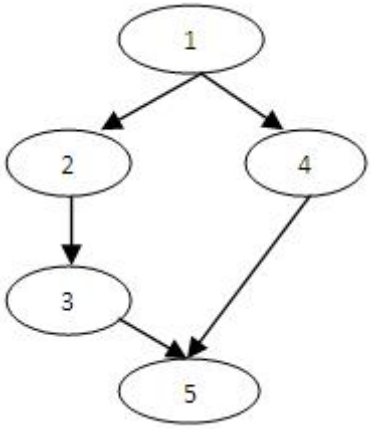



Figure 1. An example graph

Figure 1 is an example graph, with the nodes 1 to 5 indicating 5 different individual tasks that are needed for the whole computation. With dependencies to each other (shown as arrows), we know that in order to start calculating task 5, task 3 and task 4 need to be finished beforehand. Inside a distributed system, a task scheduler needs to distribute the tasks on different machines to make the overall performance efficient. For the example that is given in Figure 1, a possible solution is to utilize 2 threads: one of them is made of tasks 1, 2, 3, 5 and the other one is made of tasks 1, 4 and 5.

Parallel and distributed computing does not always provide a better performance compared to sequential computing, with the delay and overhead caused by message sending and unbalanced parallelism (http://en.wikipedia.org/wiki/Distributed_computing; http://en.wikipedia.org/wiki/Parallel_computing). Utilizing dynamic voltage scaling can help us to optimize the time and energy consumption for unbalanced tasks in a distributed system.

### *2.2 Dynamic Voltage Scaling*

Dynamic Voltage Scaling (DVS) is a power management function that allows software (e.g. port through BIOS) to scale up or scale down the voltage supply on components (such as RAM, CPU and disk) inside a computing system. The power consumption (Rabaey, 1996) P on a device can be presented as follows.

$$P = CV^2f \quad (1)$$

where C is the capacitance being switched per clock cycle; V is the supply voltage and f is the switching frequency.

DVS has proven to be a very promising technique on DVS-enabled devices. To implement DVS efficiently in a distributed computing system, we need to schedule the tasks on different processors so that we can reach the maximum optimization with minimum energy power to finish the specific task. The available voltage levels are normally pre-set according to the hardware standards. In our case, we have the following four levels according to the Intel Pentium 4 Processor:

Table 1. The pre-set voltage level for the simulated hardware (Note 1)

| Level | Voltage (V) | Relative Speed (%) | Energy Reduction (%) |
|---|---|---|---|
| 0 | 1.75 | 100% | 0% |
| 1 | 1.4 | 80% | 20% |
| 2 | 1.2 | 60% | 22% |
| 3 | 0.9 | 40% | 33% |

Unfortunately, the algorithm of finding the critical path is NP-Complete, which means that the processing time is super-polynomial to the input size. This renders the computation time prohibitive for problems with large excessively large inputs. To address this issue, we have proposed the new evolutionary optimization algorithm called Simplified Swarm Optimization (SSO), which we present in detail in Section 3. We first present the graph-based model of distributed systems and the key performance parameters to provide context on the problem that requires optimization.

*2.3 Graph-Based Representation*

We use Directed Acyclic Graph (DAG) file format to store the tasks. A DAG file indicates a tree structure of tasks that need to be computed. In our DAG files, we specify the number of processors, the dependencies between nodes, the processing time for each task on each processor, and the communication delay for each dependency.

The proposed optimization algorithm can allocate the processor for each task as well as the voltage supply for each pair of processor and task. The energy consumption will be calculated according to the voltage supply and the relative speed.

*2.4 Performance Parameters*

The evaluation will be focused on three different aspects: makespan, time, and the energy optimization.

Makespan is the amount of time needed to finish processing all the tasks including the delays for communications created during the distributed processes. The makespan is a key parameter to evaluate the quality of the task schedule. A good schedule of tasks should be relatively more balanced so that we can achieve good performance.

Time is the amount of time needed to schedule the tasks. Greedy algorithms always provide us with a solution in the shortest time, however since our problem is NP-Complete, we need algorithms that make approximations so that we can calculate the solution in a reasonable amount of time.

Energy consumption is the most important measure in our research, which includes energy needed for computing the tasks, as well as the overhead created during distributing the tasks.

Our goal is to measure these three aspects for different optimization algorithms so that we can achieve the best energy savings with the minimum performance cost.

**3. Algorithms**

This section introduces the class of swarm intelligence algorithms (http://en.wikipedia.org/wiki/Swarm_intelligence). We briefly revisit the popular Particle Swarm Intelligence approach, and then we introduce our Simplified Swarm Intelligence method.

*3.1 Introduction to Swarm Intelligence*

The inspiration of Swarm Intelligence (SI) (Beni, 1989) comes from nature. SI systems are often made up of a set of robots or agents, who are operating on an n-dimensional space interacting locally with one and another.

The following will give an introduction to the most popular evolutionary algorithm Particle Swarm Optimization (PSO) and the proposed Simplified Swarm Optimization (SSO). Both of them have been applied to the task scheduling problem in this work.

*3.2 PSO*

The original idea of Particle Swarm Optimization (PSO) was inspired by the movement of bird flocking. The PSO algorithm mimics the behavior of flying birds and their means of information exchange to solve optimization problems. Each potential solution is seen as a particle with a certain velocity, and “flies” through the problem space. Each particle adjusts its flight according to its own flying experience and its companions’ flying experience. The particle swarms find optimal regions of complex search spaces through the interaction of individuals in a population of particles. PSO has been successfully applied to a large number of difficult combinatorial optimization problems; and it often outperforms Genetic Algorithms (Goldberg, 1989).

In PSO, particles represent candidate solutions in a solution space, and the optimal solution is found through moving the particles in the solution space. An individual particle flies through an n-dimensional search space with a velocity that dynamically changes according to its own experience and other particles existing in the same search space. The velocity changes under the following rule:

$$V_{id(t)} = WV_{id(t-1)} + C_1R_1\left(P_{id(t-1)} - X_{id(t-1)}\right) + C_2R_2\left(P_{gd(t-1)} - X_{id(t-1)}\right) \quad (2)$$

$$X_{id(t)} = X_{id(t-1)} + V_{id(t-1)} \quad (3)$$

where d=1, 2, 3,…, S ; t is the round of iteration; $V_i$ and $X_i$ are the velocity and position of the $i_{th}$ particle; $P_i$ is the previous local best of particle I and is called pbest; $P_\alpha$ is the previous best position for all particles and is called gbest; $C_1$ and $C_2$ are positive constants; and $R_1$ and $R_2$ are random values. All particles will be assessed on their fitness by a function. A standard procedure for PSO can be described by the following:

1) Initialize t = 0, and S, and set P = S
2) Evaluate S and P, and define gbest and pbest
3) While t < MAX_ITERATION:
4) Update S using Equation (2) and Equation (3)
5) Evaluate S
6) Update P and redefine pbest and gbest
7) t = t+1
8) END While
9) Return gbest

*3.3 SSO*

In this paper, we propose Simplified Swarm Optimization (SSO) which is an adaptation of PSO using discrete values and modifying the mutation operation to be determined randomly between 3 user-set limits. SSO uses random populations with the mutation operation randomly changing each dimension of the particle as determined by the aforementioned limits. This allows the user to change between focusing on finding local optima and expanding the search to cover more of the problem space.

SSO is unique and effective due to its simple search methods. Prior to this, particles are mutated randomly with random dimensions being changed in each cycle. Having each dimension change toward a local optimal or global optimal dimension is very well suited to discrete data and distributed computing applications. Combining with the above algorithms gives a wide variety of random vs. controlled mutation and continuous vs. discrete data optimisation techniques.

The main difference of SSO and PSO is that, SSO does not need to use the velocity and the initial weight; instead the update positions of particles are chosen based on the relationship between the values of the new generated random variable and three pre-defined constants $C_W$, $C_D$ and $C_G$ ranging from (0, 1):

$$X_{id(t)} = \begin{cases} X_{id(t-1)} \ if \ newRandom \in [0, C_w) \\ P_{id(t-1)} \ if \ newRandom \in [C_w, C_p) \\ g_{id(t-1)} \ if \ newRandom \in \left[C_p, C_g\right) \\ x \ if \ newRandom \in [C_g, 1) \end{cases} \quad (4)$$

where X is the position, P is the local best and g is the global best. During the iterations, for each dimension of the particle, a newRandom variable is generated in the range of (0, 1), and the new position will be chosen between previous position, local best, global best and the current location based on which interval the newRandom lies in.

Based on the concept, SSO is more suitable to deal with discrete variables and PSO is more suitable to deal with continuous variables.

In section 4, we will compare the performance of PSO and the proposed SSO in our experimental model and explore the feasibility of using them as a task scheduler.

## 4. Performance Evaluation

*4.1 Benchmark*

To evaluate the performance, we need some comparative data. For example, the energy consumption usually

decreases when the execution time takes longer, and our goal is to find an operating point that balances between the performance and energy saving. Therefore, when evaluating the energy consumption, we also need to take in consideration of the speedups.

The makespan will be compared to the sequential runtime ($T_o$) of the original tasks so that we can see how different the speedups (Amdahl, 1967) are by using different optimization techniques. To calculate the speedups, we have:

$$Speedup = \frac{T_o}{makespan + time} \tag{5}$$

We will also calculate the raw energy consumption ($E_o$), to evaluate the results. The energy optimization will be $O_{algo} = E_o / E_{algo}$. We have used hundreds of simulated tasks from the DAG files to test and verify our proposed SSO system. The optimization will be mainly measured by the macro-average of $O_{algo}$.

*4.2 Test Data*

In this paper, we use a graph generator to generate a large set of DAG files for the testing data. The graph generator can generate directed acyclic graphs with specified number of edges and nodes, and the amount of computation needed for each task (node) and the communication (edge) is randomly selected from a list of commonly used constant values.

Table 2. The testing result for PSO and SSO

| | | *Makespan (ms)* | *Time (ms)* | *Energy (J)* |
|---|---|---|---|---|
| None | Overall | 10005777 | 0 | 30595583 |
| | Avg | 100% | N/A | 100% |
| PSO | Overall | 3806335 | 1079 | 23955588 |
| | Avg | 42.64% | N/A | 78.24% |
| SSO | Overall | 2476881 | 855 | 24435717 |
| | Avg | 27.02% | N/A | 79.21% |

We have tested our model using sufficiently large amount of data in our experiments. For each test case, we have calculated the makespan, time and energy consumption under no optimization and under optimization using traditional PSO and the proposed SSO. To compare the differences, we have used the macro-average of makespan and energy saving to calculate the average performance of the three testing algorithms: None (no optimization), PSO and SSO.

Table 3. The speedup result for using PSO and SSO

| | *None* | *PSO* | *SSO* |
|---|---|---|---|
| Speed up | 1 | 2.63 | 4.03 |

*4.3 Analysis*

Table 2 shows the overall statistics of our experiment. From the table, we can see that both of the two algorithms can achieve over 20% of energy savings and PSO can perform slightly (less than 1%) better than SSO. However, SSO can provide scheduled tasks with better quality, and the makespan of SSO solutions are significantly less than the makespan of PSO (from Table 2). The execution time of SSO is also around 10% smaller than PSO. Table  has shown the speedups (calculated using Equation (5)) of PSO and SSO. Our results show that SSO is definitely providing a task schedule that consumes much less time than PSO. Also, we have found that both PSO and SSO can work better when dealing with more complex tasks.

In the experiments, we have also tried various tasks of different topological structures, and the tasks are different on many aspects which include: 1) the number of available processors, 2) the dependencies between the tasks, 3) the depth of critical path, 4) the total number of possible paths. From our results, we have found that the structure of the tasks have very little effect on the energy optimization; however the effect on makespan varies a lot.

*4.4 Comparison*

4.4.1 Makespan

We have also studied the performance of PSO and SSO on individual test cases. In Figure 2, we have used a chart to compare the makespan of using SSO and PSO. In this work, we would like to test how these two algorithms are different in makespan. According to the results in the previous table (Table 2), the overall average of makespan of SSO is only 65% of PSO, and we have found that it also applies to the individual cases from Figure 2.

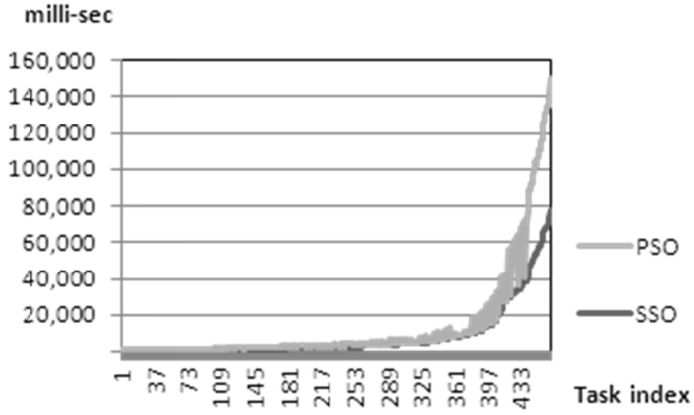



Figure 2. Makespan comparison for SSO and PSO

In Figure 2, the x coordinate indicates the index of test case, and the y coordinate indicates the measurement of makespan. As displayed in the chart, the ratio of PSO makespan and SSO makespan is also relatively stable for most of the cases.

4.4.2 Time

Figure 3 below shows the relationship of the execution time of PSO and SSO. As shown in Figure 3 the x coordinate indicates the test case index (test cases sorted by their execution time), and the y coordinate indicates the time in milli seconds. We can see from Figure 3, the ratio of time between PSO and SSO has shown that SSO is 10% faster than PSO.

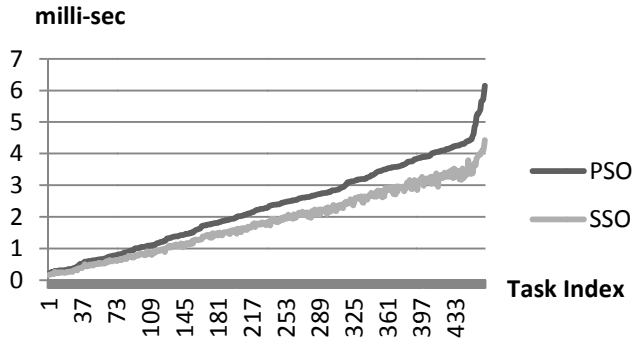



Figure 3. Execution time comparison for PSO and SSO

4.4.3 Energy

The experimental results have shown that both PSO and SSO can deliver over 20% of energy savings. We have studied the performance of PSO and SSO on energy saving. In the graph below, we are showing the $O_{PSO} / O_{SSO}$ for 500 individual test cases.

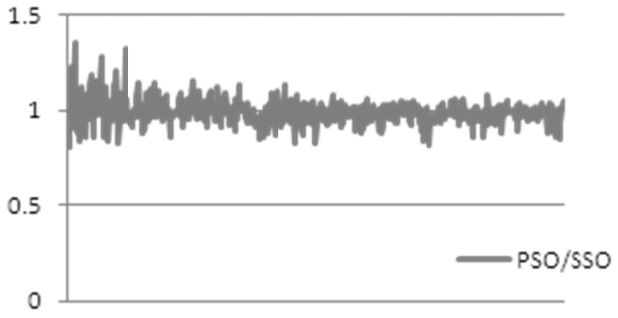



Figure 4. Energy saving comparison for PSO and SSO

Figure 4 shows that the ratio of $O_{PSO} / O_{SSO}$ is floating at around 1. Therefore, the results of PSO and SSO are very close to each other.

## 5. Conclusions and Future Work

Task scheduling is a common problem for distributed systems. In our research, we have improved the computational efficiency by implementing swarm intelligence techniques to solve the task-scheduling problem to reduce the energy cost. Even though the problem itself is NP-Complete, the proposed SSO algorithm can provide a good solution within a reasonable amount of time without delay. In our model, both PSO and SSO algorithms can improve the performance as well as energy consumption, and the amount of energy saved is approximately 21% of the entire energy cost including the pre-calculation and latency when compared with the system without using optimization algorithm.

We have also studied the differences between the proposed SSO and the traditional PSO. Our experimental results have showed that the proposed SSO tends to provide an overall better solution than PSO for energy consumption and task scheduling, though PSO is also acceptable for non time sensitive systems.

Improving the energy efficiency of computation is becoming increasingly more important nowadays with greater reliance on distributed data centers and servers. We have demonstrated that using dynamic voltage scaling and swarm intelligence can improve the energy efficiency as well as the performance. Moreover, both PSO and SSO can be used to schedule tasks for other distributed tasks such as disk drivers, and GPUs.

## References

Amdahl, G. M. (1967). Validity of the Single Processor Approach to Achieving Large-Scale Computing Capabilities. *Solid-State Circuits Society Newsletter, IEEE, 12*(3), 19-20. http://dx.doi.org/10.1109/N-SSC.2007.4785615

Beni, G. (1989). From Swarm Intelligence to Swarm Robotics. *Swarm Robotics Lecture Notes in Computer Science, 3342*, 1-9. http://dx.doi.org/10.1007/978-3-540-30552-1_1

Distributed Computing http://en.wikipedia.org/wiki/Distributed_computing

Garey, M. R., & Johnson, D. S. (1990). *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman Co.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Mass: Addison-Wesley.

Kennedy, J., & Eberhart, R. (1995). Full text access may be available. *Particle swarm optimization. IEEE International Conference on Neural Networks, 4*, 1942-1948. http://dx.doi.org/10.1109/ICNN.1995.488968

Kashan, A. H., & Karim, B. (2009). A discrete particle swarm optimization algorithm for scheduling parallel machines. *Computers & Industrial Engineering, 56*(1), 216-223. http://dx.doi.org/10.1016/j.cie.2008.05.007

Koomey, J. G. (2008). Worldwide electricity used in data centers. *Environmental Research Letters, 3*(034008). http://dx.doi.org/10.1088/1748-9326/3/3/034008

Lee, Y. C., & Zomaya, A. Y. (2009). Minimizing energy consumption for precedence-constrained applications using dynamic voltage scaling. *In Proceedings of 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*. http://dx.doi.org/10.1109/CCGRID.2009.16

Parallel Computing http://en.wikipedia.org/wiki/Parallel_computing

Rabaey, J. M. (1996). *Digital Integrated Circuits*. Upper Saddle River, NJ: Prentice Hall.

Swarm Intelligence http://en.wikipedia.org/wiki/Swarm_intelligence

**Notes**

Note 1. Notice that the energy reduction is under the ideal situation and in real world, we have to consider the communication delay

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# Update on the Statistical Analysis of Traffic Countings on Two-Lane Rural Highways

Raffaele Mauro[1] & Federico Branco[2]

[1] University of Trento, Faculty of Engineering, Trento, Italy

[2] Consulting Engineer, Trento, Italy

Correspondence: Raffaele Mauro, University of Trento, Faculty of Engineering, Trento, Italy. Tel: 39-461-282-588. E-mail: raffaele.mauro@ing.unitn.it

Received: April 7, 2013 Accepted: May 2, 2013 Online Published: May 21, 2013

doi:10.5539/mas.v7n6p67 URL: http://dx.doi.org/10.5539/mas.v7n6p67

**Abstract**

This paper deals with the problem of model identification, calibration and validation for traffic countings on two-lane rural highways. A criterion for preliminary selection of arrival laws as a function of appropriate sample statistics and a technique for deciding whether sample data sets of traffic counting are congruent with stationary time series behavior are suggested; besides arrival laws currently used in research and engineering practice, the Neyman distribution has been also applied although it is not frequently implemented in the field of traffic engineering. Moreover, this work aims at applying these methods to a set of empirical data derived from a recent survey on two two-lane rural highways; the arrival laws that best agree with the observations are found and the relations between the parameters identifying the arrival laws and the flow rates are worked out. Finally, the results have been compared to those achieved in similar observations, carried out by one of the authors in the past.

**Keywords:** two-lane rural highway, counting distribution, steady-state flow

## 1. Introduction

Traffic counting represents the first and, to this day, foremost empirical measurement designed to research on vehicular flows; counting distributions have direct relevance to discrete-time point processes applied to road traffic (Bertò, Schoen, & Speranza, 1996). Statistical models for traffic counting, also known as probabilistic arrival laws, and vehicular headway distributions (Mauro & Branco, 2012; Ha, Aron, & Cohen, 2012) are used in road and highway engineering, i.e. in traffic simulation procedures, in vehicle density estimations as well as in the study of the waiting phenomena at intersections and barriers. From the beginning of the 30s until the late 70s, special attention was devoted to theoretical and application aspects of these topics. Important studies have been made by Kinzer (1933), Adams (1936), Breiman (1962) and Gerlough and Barnes (1971). Helpful guidelines for traffic engineers about the study and application of counting distributions are also contained in Gerlough and Huber (1975) and in May (1990). After a few decades theoretical research has restarted and some works in the field have been produced by Jabari and Liu (2012), Clementi, Monti and Silvestri (2011) and Cao, Tai and Chan (2012) who have analysed some statistical models for counting distributions.

Since the 30s Poisson law has been proposed for theoretical distribution of arrivals, in that it is a peculiar flow model for discrete events under conditions of statistical regularity. Some interesting generalizations about this law have been subsequently made along with the progress of research on the issue of vehicular arrivals. Poisson law was introduced by Kinzer (1933) for the elaboration of certain aerial surveys on traffic, previously carried out by Johnson (1928). Afterwards, it was applied by Adams (1936) for further numerical exemplifications and by Greenshields, Shapiro and Ericksen (1947) in a work on intersections. A lot of applications of this traffic model have been later suggested by Gerlough, first in a book entirely on the topic (Gerlough, 1955) and more recently, in collaboration with others, in a paper that also deals with further theoretical distributions that may be used as arrival laws (Gerlough & Barnes, 1971). Finally, certain realistic models of vehicular flows leading to Poisson arrivals are described in Breiman's (1962, 1963) and Weiss and Herman's (1962) researches.

In the study of traffic conditions for which Poisson model is not useful, various scholars have suggested and sometimes verified through experimentation different counting laws. Specifically, among the most interesting

contributions to applications, Beckmann et al. have proposed a simple binomial distribution model (Beckmann, McGuire, & Winsten, 1956); Haight (1959) has introduced the generalized Poisson law and studied it with reference to real cases in collaboration with others (Haight, Whilser, & Mosher, 1961); Buckley (1965) and Drew (1965) have researched the possibility of applying the negative binomial distribution; and, once again, Buckley (1968) has further generalized the Poisson law. Theoretical works which propose counting models, barely or not at all verified through empirical validation, cannot be predominantly neglected. Oliver and Thibault (1962), Buckley (1965), Buckley (1967), and Serfling (1969) can be consulted on such a type of law.

Finally, Ha, Aron and Cohen (2012) can be referred to for most recent results; they have introduced innovative models for time headway and counting distributions, supported by empirical researches on French roads.

## 2. Probabilistic Models for Arrivals

Results from experiments have shown that according to the statistical hypothesis Poisson law is, at the roots of its deduction, correctly applicable as a model for arrivals, on one or multiple lanes, if in the observed interval:

a) the phenomenon remains stationary, i.e. no external perturbation intervenes and affects flows;

b) the gap between vehicles is such that they do not influence each other.

Under traffic conditions in which the vehicles are not farther apart than the distance at which they do not interact with each other, the circumstance b) involves that the lower the flow, the more consistent with empirical data is the model. Considering the results from this research, the authors suggest to apply the model to flow rates up to 400÷500 vph in ordinary road conditions (dry weather, daylight, pavement in good conditions). Thus, Poisson law cannot generally be used without stationarity and on high flow rates; if these two conditions are not met, other probability laws are to be adopted. The criterion for choosing alternative counting models, presented in the literature and exposed in Gerlough and Huber (1975), is described below.

If the mean $\bar{x}$ and variance $s^2$ of the sample turn out to be substantially equal, one can assume that the statistical distribution may be close to Poisson law. According to such a law, the mean $\mu$ and variance $\sigma^2$ turn out to be equal and this value completely defines the model. For the sake of simplicity, this paper does not deal with other statistical laws for which $\mu = \sigma^2$. If the mean of the sample turns out to be higher than the variance, the measure variability can be deduced as lower than that expected from purely poissonian arrivals of equal mean. Should that be the case, the empirical data can be checked to see how consistent they are with the positive binomial model or with the generalized Poisson model, according to which the mean is generally higher than the variance. Also the quantities $\mu$ and $\sigma^2$ completely identify both distributions. In terms of traffic, the circumstance $\bar{x} > s^2$ has been found on a frequent basis in flow conditions far from the free flow circulation, when flow rates usually present a very high volume. Finally, if the mean of the sample is lower than its variance-in other words, for a given mean value countings appear to be more dispersed than those derived from purely poissonian arrivals-the negative binomial distribution, also known as Pascal's law, should be generally determined, in that it presents, like the data, a lower expectation than the variance $\mu < \sigma^2$. Also in this case, mean and variance completely identify the model. Cases of traffic countings where $\bar{x} > s^2$ and in the presence of compliance with Pascal's law have been mainly observed in flow conditions developed under traffic-light regulation, although Buckley (1967) has used this distribution as a model for arrivals on roads with two or more lanes.

Table 1 summarizes the criterion for choosing a model according to the mean $\bar{x}$ and variance $s^2$ values.

Table 1. Selection criteria for traffic counting models

| Value of $s^2/\bar{x}$ | Suggested distribution |
|---|---|
| $> 1$ | negative binomial |
| $\cong 1$ | Poisson |
| $< 1$ | binomial or generalized Poisson |

Table 2 shows the probability distribution of the binomial, Poisson and negative binomial random variables with their parameter expressions as a function of the sample statistics $\bar{x}$ and $s^2$.

It is worth remembering that the choice for the length of the subdivision intervals $\Delta t$ of the observation period T is known to influence the identification of the model, as well explained by Gerlough and Barnes (1971). For example, this means that, in the same period of observation, a flow can be Poissonian if the data are recorded on

subintervals $\Delta t_1$, but on the other hand it can conform to other models if $\Delta t_2 < \Delta t_1$ are used. As for the estimates shown in this paper, the width of the interval was predetermined by the assumptions about the minimum extension of the periods of stationarity. Moreover, the duration T of the observation period, at a constant length of the subinterval of subdivision t, has been proved to affect the ratio $s^2/\bar{x}$ : the bigger the interval T, the bigger the ratio $s^2/\bar{x}$ (Miller, 1970).

Table 2. Theoretical distributions of arrival: probability distribution, mean, variance, parameters estimate

| Distribution | binomial | Poisson | negative binomial |
|---|---|---|---|
| probability distribution | $\binom{n}{x} p^x (1-p)^{n-x}$ | $\frac{\mu^x}{x!} e^{-\mu}$ | $\binom{x+k-1}{k-1} p^k (1-p)^x$ |
| mean $\mu$ | $np$ | $\mu$ | $\frac{k(1-p)}{p}$ |
| variance $\sigma^2$ | $np(1-p)$ | $\mu$ | $\frac{k(1-p)}{p^2}$ |
| $\frac{\mu}{\sigma^2}$ | $(1-p)^{-1} > 1$ | 1 | $p < 1$ |
| parameters estimate | $p = (\bar{x} - s^2)/\bar{x}$<br>$n = \bar{x}^2/(\bar{x} - s^2)$ | $\mu = \bar{x}$ | $p = \bar{x}/s^2$<br>$k = \bar{x}^2/(s^2 - \bar{x})$ |

In addition to the selection criterion indicated in Table 1, a further criterion for the preliminary choice of the model, as a function of appropriate statistical data, is shown. In some research areas (e.g. Biometrics) (Gore & Paranjpe, 2001), this criterion is more efficient than the simple comparison between the values of mean and variance: in order to have a greater amount of information than that gained only from the comparison of $\bar{I} = s^2/\bar{x}$ with $I = \sigma^2/\mu$, the parameter $\bar{L} = m_3/s^2$ can also be compared with $L = \mu_3/\sigma^2$, where $\mu_3$ is the theoretical third central moment and $m_3$ is the corresponding frequency moment. On the plane (I,O,L) the point (I,L) is placed on a different position for each different theoretical distribution (Figure 1), lying on the segment AC if it is representative of a binomial distribution, on the half-line with origin in C if it is representative of a negative binomial distribution, and coinciding with the point C if it is representative of the Poisson law (Ord, 1972).

The positioning of the sample points $(\bar{I}, \bar{L})$ on the plane (I,O,L) near the Loci previously defined can be helpful for the choice of the theoretical distribution that best fits the sample data; however, the statistical hypothesis can be verified by applying a test of hypothesis to the chosen model.

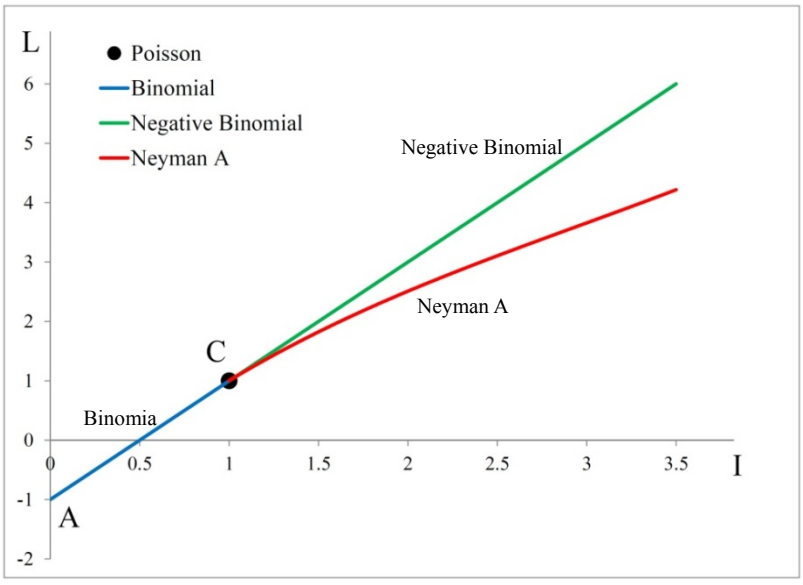



Figure 1. Probability distributions on the plane (I,O,L)

This method is very useful and easy to implement, but on the other hand it is not always univocally discriminative, as later shown in this work. In Figure 1 the curve L = f(I) of the Neyman type A biparametric distribution is also illustrated. The probability distribution of the Neyman type A is:

$$p(0) = e^{-m_1}\left(1 - e^{-m_2}\right) \tag{1}$$

$$p(n+1) = \frac{m_1 m_2 e^{-m_2}}{n+1} \sum_{k=0}^{n} \frac{m_2^k}{k!} p(n-k), \quad n = 0,1,2,... \tag{2}$$

where $m_1$ and $m_2$ are:

$$m_1 = \bar{x}^2 \Big/ \left(s^2 - \bar{x}\right) \tag{3}$$

$$m_2 = \left(s^2 - \bar{x}\right) \Big/ \bar{x} \tag{4}$$

Relations L = f(I) for the models just mentioned and deduced from the expressions of $\mu$, $\sigma^2$ and $\mu_3$, are shown in Table 3.

Table 3. Relations L = f(I) for the probability models

| Distribution | Relation L = f(I) | |
|---|---|---|
| Poisson | L = I = 1 | |
| binomial | L = 2I – 1 | $I \in [0,1]$ |
| negative binomial | L = 2I – 1 | $I \in [1,\infty[$ |
| Neyman tipe A | L = I + 1 – 1/I | $I \in [1,\infty[$ |

The negative binomial and Neyman type A biparametric distributions belong to the larger family of probabilistic models known as “aggregate or contagious distributions”. They are of great importance in Biometrics when the elements of the surveyed populations are generally found in groups or clusters. As for traffic, the equivalent situation can be platoons. Therefore, in such a circumstance (running in platoons) traffic countings should be consistent with the negative biomial or Neyman type A biparametric distributions. This side of the problem will be dealt with in the following analysis of the empirical results.

The next section firstly shows a statistical procedure for the identification of steady-state periods; in fact, model identification, calibration and validation of counting laws have to be made on these stationarity periods. These periods need to be defined so as to proceed with the search for links between the statistical parameters identifying models and flow rates, or more in general, relations between flow rates and traffic processes parameters.

## 3. Flow Stationarity Tests

T is defined as the observation period of the traffic flow in a section of a lane. If T is divided into n smaller intervals $t_i$, i = 1, 2, …, n of equal length $\Delta t$ and the mean of the process, represented by the sequence $\{X_i\}$ of the random variables defined as “number of vehicles crossing the road section during the interval $\Delta t_i$”, is constant, the flow is defined as stationary during the period T. In order to check that the realization $\{x_i\} = x_1, x_2, \ldots, x_n$ of the sequence $\{X_i\}$ is extracted by a process of traffic countings with a constant mean, a “distribution free” test is used in this paper in that, unlike other types (for instance, the sequential probability ratio test by Wald (1947)), it has the advantage of not requiring any statistical hypothesis on the arrival law. Through this test the constancy of the process mean is controlled by verifying the independence hypothesis of the sequence $\{X_i\}$ compared to the independence of the first “n” natural numbers (Kendall & Stuart, 1967).

The parameter used in this test is the correlation coefficient r between the two sequences:

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n} i x_i - \bar{x}\frac{n+1}{2}}{\left[\frac{1}{12} s^2 \left(n^2 - 1\right)\right]^{1/2}} \tag{5}$$

where $\bar{x}$ and $s^2$ are the mean and the variance of $\{X_i\}$ elements. The distribution of equation (5) is obtained by

considering that the n values of r calculated in n permutations of $\{X_i\}$ elements are equiprobable. The extreme values of "r" are evidently –1 and +1, where r = +1 if the sequence of traffic countings increases linearly with the natural numbers, r = –1 if the $\{X_i\}$ decrease when the natural numbers increase. Choosing suitably large n, typically n > 8, it is possible to limit the length of the acceptance zone of the test and to replace (5) with the parameter obtained from (5) that follows:

$$t = \left[ \frac{(n-2)r^2}{1-r^2} \right]^{1/2} \tag{6}$$

distributed according to a Student's t with (n – 2) degrees of freedom.

Through these parameters, the hypothesis of flow stationarity at level $\alpha$ is refused when

$$|t| \geq t_{\alpha/2} \tag{7}$$

One has to choose n so that the length $\Delta t$ is greater than a minimum value; in this way it is possible to consider the elements $\{X_i\}$ as independent elements and to apply to them a "distribution free" test, such as the one just mentioned.

Finally, the same length $\Delta t$ of the observation interval must be kept sufficiently small to avoid non-linear flow variations during the interval. Such fluctuations cannot be pointed out by the test adopted which is rather very useful only for the linear trend hypothesis. Consequently, the present study focuses the quantities $\Delta t$ = 20 sec and n = 30; so the observation period is equal to T = 10 min. With the foregoing assumptions, for definition the flow is considered stationary for intervals of less than 10 min.

In order to verify that the length $\Delta t$ = 20 sec is appropriate to consider the observed traffic countings as independent elements, a lag 1 serial correlation test is carried out. This verification is usually used by applying, for the elements of $\{X_i\}$, the autocorrelation coefficient given by:

$$\rho = \left( \frac{\mu_3}{6\mu^3} - \frac{(\sigma^2)^2}{4\mu^4} - \frac{1}{12} \right) \left( \frac{\sigma^2 \Delta t}{\mu^3} + \frac{1}{6} + \frac{(\sigma^2)^2}{2\mu^4} - \frac{\mu_3}{3\mu^3} \right)^{-1} \tag{8}$$

where $\mu$, $\sigma^2$ and $\mu_3$ are the mean, variance and third central moment of the time headway distribution (Cox & Lewis, 1966). Using Equation (8) it is possible to calculate $\rho$ and to evaluate if it is small enough to accept the hypothesis of absence of statistical relationships between the elements of the traffic counting sequence.

Indeed, in this research work no information about time headways is available; therefore, the lag 1 serial correlation test just above recalled is used through the quantity

$$r = \sum_{i=1}^{n} x_i x_{i+1} \tag{9}$$

Equation (9) for sufficiently large n (n > 15) is normally distributed with mean and variance given by:

$$\mu_r = \frac{s_1^2 - s_2}{n-1} \tag{10}$$

$$\sigma_R^2 = \frac{s_2^2 - s_4}{n-1} + \frac{s_1^4 - 4s_1^2 s_2 + 4s_1 s_3 + s_2^2 - 2s_4}{(n-1)(n-2)} - \mu_r \tag{11}$$

where:

$$s_k = \sum_{i=1}^{n} x_i^k \tag{12}$$

The knowledge of the theoretical distribution of r allows performing a common two-sided test at the chosen significance level $\alpha$. Through this test, the hypothesis of independence is rejected if:

$$|r| > r_{\alpha/2} \tag{13}$$

Returning now to the correlation ratio test for the verification of the flow stationarity, it has been applied in two ways to the data surveyed and with the values of n, $\Delta t$ and T previously specified.

The former way made the assumption of a test dimension equal to $\alpha$ = 0.05 and treated the countings of the first thirty $\Delta t_i$ starting from the observation instant. If the stationarity occurred, three additional consecutive subintervals were added and the first three subintervals were excluded; such a test was then repeated on the new

sequence. The iteration stopped when the non stationarity of $\{X_i\}$ was obtained. The test was then performed on arrivals per 10-minute interval, where the first minute corresponded to the last minute of the previous interval. The range of acceptance of this test, applied to the empirical data, was not small, even if the sample size n was not very large. The procedure for data sets creation was therefore modified in order to obtain realizations of gradually size increasing $\{X_i\}$ as well as increasingly narrow acceptance intervals: starting from the first instant, the countings relating to the following $\Delta t_i$ were added one by one and the test was carried out, step by step, on larger and larger samples; the process is stopped when the non-stationary flow occurred. The procedure was iterated starting from the traffic counting value which corresponded to the last minute of the previous sequence and identifying further periods at a constant flow, and so forth up to the complete analysis of the available data. The test presented a considerable inertia and revealed the presence of a trend only in very few cases.

For those reasons, the stationarity intervals identified through the first technique were then employed in the following analyses.

The methodology used here has been described by Esposito and Mauro (1994) who applied the technique for the identification of stationarity periods to empirical measures of traffic countings. After nearly two decades the authors used the same technique on newly-acquired empirical data, as specified below; the results from newly-collected data and their comparison with those obtained by Esposito and Mauro (1994) are presented in the following paragraph.

## 4. Empirical Data Analysis

The methods described in paragraphs 2 and 3 were used in traffic countings data carried out in the spring of 2012; these data refer to two sections of two-lane rural highways (one lane for each direction). In particular, the data of 1800 intervals with a width $\Delta t = 20$ sec, equal to a total of 10 hours and 3053 vehicular passages, were analysed. During these periods, the observed road section was filmed by a video camera and the countings were later carried out by means of the recordings.

Empirical data have been surveyed on two road sections of the following roads in the province of Trento, Italy:

1) Provincial Road No. 36 “delle Grazie”, in the municipality of Arco;
2) National Road No. 421 “dei Laghi di Molveno e Tenno”, in the municipality of Tenno.

The first section is set on a long, flat and straight road where overtaking is disallowed. The roadway is about 7.00 m wide, with lane width of 3.25 m and paved shoulders of 0.25 m. Table 4 shows some information gained from the countings on section 1.

Table 4. Information about the countings on the Provincial Road No. 36 “delle Grazie”

| Set | Date | Time | Direction |
|---|---|---|---|
| A | April 26th, 2012 | 10:00-12:00 | North |
| B | April 26th, 2012 | 13:00-14:00 | South |
| C | April 27th, 2012 | 16:00-17:00 | North |
| D | April 27th, 2012 | 17:00-18:00 | South |

Instead, section 2 is set on a short, straight road, with a 6% grade, where overtaking is disallowed. The carriageway is about 6.50 m wide and the paved shoulders are about 0.50 m. Table 5 summarizes some information on the survey.

Table 5. Information about the countings on the National Road No. 421 “dei Laghi di Molveno e Tenno”

| Set | Date | Time | Direction |
|---|---|---|---|
| E | May 10th, 2012 | 15:00-18:00 | South |
| F | May 10th, 2012 | 17:00-19:00 | North |

By applying the methodology above described to all the collected data, the observed intervals were divided into 23 stationarity periods, summarized in Table 6 and in Table 7, respectively for the former and the latter road section.

Table 6. Data on Provincial Road No. 36 “delle Grazie”

| Direction and date | Hour and stationarity period length | Set | Traffic counting data | | | I and L estimate | | Parameters | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | negative binomial | | binomial | | Neyman A | |
| | | | $\bar{x}$ (veh/20 sec) | $s^2$ | Flow rate (veh/h) | I | L | p | k | p | n | $m_1$ | $m_2$ |
| North 04/26/’12 | 10.00-10.33 33 min | $A_1$ | 2.389 | 5.830 | 430 | 2.44 | 3.67 | 0.410 | 1.658 | - | - | - | - |
| | 10.34-10.52 18 min | $A_2$ | 2.567 | 8.637 | 462 | 3.37 | 6.24 | 0.297 | 1.085 | - | - | - | - |
| | 10.53-11.22 29 min | $A_3$ | 1.694 | 2.434 | 305 | 1.44 | 1.47 | - | - | - | - | 3.882 | 0.436 |
| | 11.23-11.49 26 min | $A_4$ | 1.583 | 4.051 | 285 | 2.56 | 4.48 | 0.391 | 1.016 | - | - | - | - |
| South 04/26/’12 | 13.00-13.21 21 min | $B_1$ | 2.694 | 6.829 | 485 | 2.53 | 4.10 | 0.395 | 1.756 | - | - | - | - |
| | 13.22-13.34 12 min | $B_2$ | 2.617 | 3.614 | 471 | 1.38 | 1.44 | - | - | - | - | 6.865 | 0.381 |
| | 13.35-13.57 22 min | $B_3$ | 2.878 | 7.899 | 518 | 2.74 | 4.35 | 0.364 | 1.649 | - | - | - | - |
| North 04/27/’12 | 16.00-16.41 41 min | $C_1$ | 1.783 | 3.700 | 321 | 2.07 | 3.00 | 0.482 | 1.659 | - | - | - | - |
| | 16.42-17.00 18 min | $C_2$ | 2.111 | 3.600 | 380 | 1.71 | 1.78 | - | - | - | - | 2.993 | 0.705 |
| South 04/27/’12 | 17.00-17.37 37 min | $D_1$ | 2.739 | 10.430 | 493 | 3.81 | 5.98 | 0.263 | 0.975 | - | - | - | - |
| | 17.38-18.00 22 min | $D_2$ | 3.178 | 14.518 | 572 | 4.57 | 7.84 | 0.219 | 0.890 | - | - | - | - |

Table 7. Data on National Road No. 421 “dei Laghi di Molveno e Tenno”

| Direction and date | Hour and stationarity period length | Set | Traffic counting data | | | I and L estimate | | Parameters | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | negative binomial | | binomial | | Neyman A | |
| | | | $\bar{x}$ (veh/20 sec) | $s^2$ | Flow rate (veh/h) | I | L | p | k | p | n | $m_1$ | $m_2$ |
| South 05/10/’12 | 15.00-15.26 26 min | $E_1$ | 0.961 | 0.876 | 173 | 0.91 | 0.89 | - | - | 0.089 | 11 | - | - |
| | 15.27-16.03 36 min | $E_2$ | 1.117 | 1.702 | 201 | 1.52 | 2.24 | 0.656 | 2.130 | - | - | - | - |
| | 16.04-16.29 25 min | $E_3$ | 0.939 | 0.721 | 169 | 0.77 | 0.81 | - | - | 0.232 | 4 | - | - |
| | 16.30-16.58 28 min | $E_4$ | 1.083 | 0.762 | 195 | 0.70 | 0.54 | - | - | 0.297 | 4 | - | - |
| | 16.59-17.21 22 min | $E_5$ | 1.211 | 1.875 | 218 | 1.55 | 1.71 | - | - | - | - | 2.209 | 0.548 |
| | 17.22-17.48 26 min | $E_6$ | 1.272 | 2.214 | 229 | 1.74 | 1.92 | - | - | - | - | 1.719 | 0.740 |
| North 05/11/’12 | 17.00-17.18 18 min | $F_1$ | 1.133 | 1.540 | 204 | 1.36 | 2.03 | 0.736 | 3.158 | - | - | - | - |
| | 17.19-17.29 10 min | $F_2$ | 1.250 | 2.238 | 225 | 1.79 | 2.08 | - | - | - | - | 1.581 | 0.790 |
| | 17.30-17.44 13 min | $F_3$ | 1.494 | 2.980 | 269 | 1.99 | 3.00 | 0.501 | 1.503 | - | - | - | - |
| | 17.45-18.12 27 min | $F_4$ | 1.217 | 2.052 | 219 | 1.69 | 2.82 | 0.593 | 1.772 | - | - | - | - |
| | 18.13-18.31 18 min | $F_5$ | 1.550 | 3.110 | 279 | 2.01 | 3.18 | 0.498 | 1.540 | - | - | - | - |
| | 18.32-19.00 28 min | $F_6$ | 1.606 | 3.870 | 289 | 2.41 | 3.72 | 0.415 | 1.138 | - | - | - | - |

For each stationarity period such tables contain the mean (for vehicles per 20 sec and vehicles per hour), the variance and the parameters I and L. By adopting the criteria laid down in paragraph 2, the theoretical distributions that best approximate the observed data for each period with a constant flow rate were analysed; the values of the parameters obtained by the calculations are also indicated in the tables.

No empirical series is well approximated by a Poisson distribution; the observed data, instead, seem to be approximated by the negative binomial (in most cases), binomial or Neyman type A distributions. The graphical representation on the plane (I,O,L) of the sample points $(\bar{I}, \bar{L})$ has helped to identify the most suitable theoretical model; in that regard see Figure 2 and Figure 3. These models were calibrated through the criteria indicated in paragraph 2 and their parameters are shown in Table 6 and Table 7.

The validation of the chosen models to the theoretical predictions was then verified by performing a $\chi^2$ test and by calculating the relative squared index, which confirmed the above choices. More specifically, the relative squared index $I_{rs}$ is defined as follows:

$$I_{rs} = \frac{\sqrt{\frac{\sum (y_i - \hat{y}_i)}{n}}}{\frac{\sum \hat{y}_i}{n}}; \quad 1 < i < n \tag{14}$$

where $y_i$ is the observed value i; $\hat{y}_i$ is the value i of the theoretical distribution.

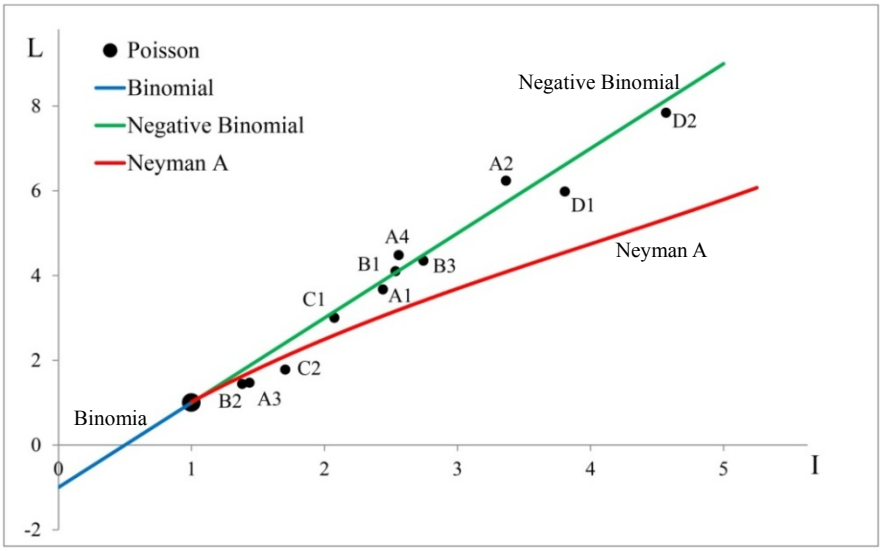



Figure 2. Plane (I,O,L) for countings on Provincial Road No. 36 "delle Grazie"

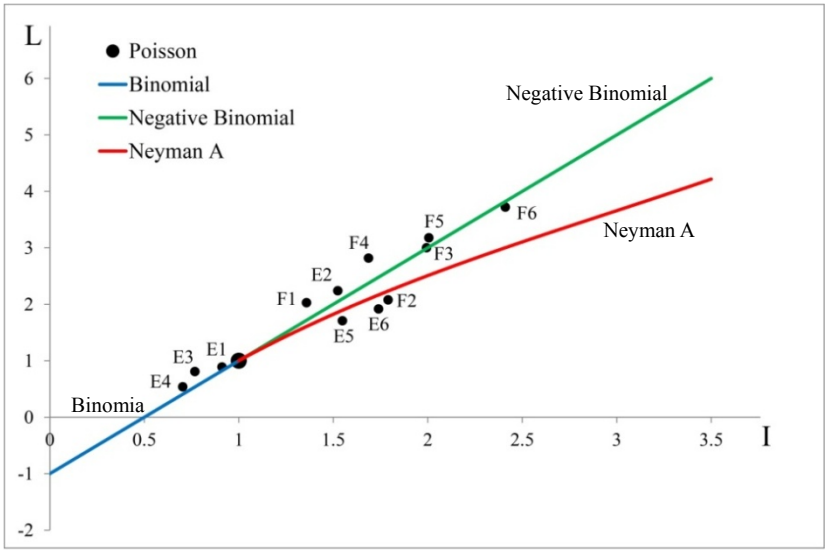



Figure 3. Plane (I,O,L) for countings on National Road No. 421 "dei Laghi di Molveno e Tenno"

By way of example, Figure 4, Figure 5 and Figure 6 show the frequency distributions of three sets of empirical traffic countings, together with the corresponding theoretical distributions that best approximate the data; the figures also show the value of the probability $p_\chi$ obtained by applying a $\chi^2$ test to assess the conformity between the sample statistic and the law chosen as its theoretical representation (note that values $p_\chi$ tending to 1 support the hypothesis verification) and the relative squared index $I_{rs}$. An example for each analysed model (negative binomial, binomial and Neyman type A) is shown.

As for the binomial and negative binomial frequency distributions, in order to avoid approximations, the binomial coefficient was calculated by generalizing the factorial operator and using the Gamma function $\Gamma(x)$ through the following definition of general validity:

$$\binom{x}{y} = \frac{\Gamma(x+1)}{\Gamma(y+1)\Gamma(x-y+1)} \tag{15}$$

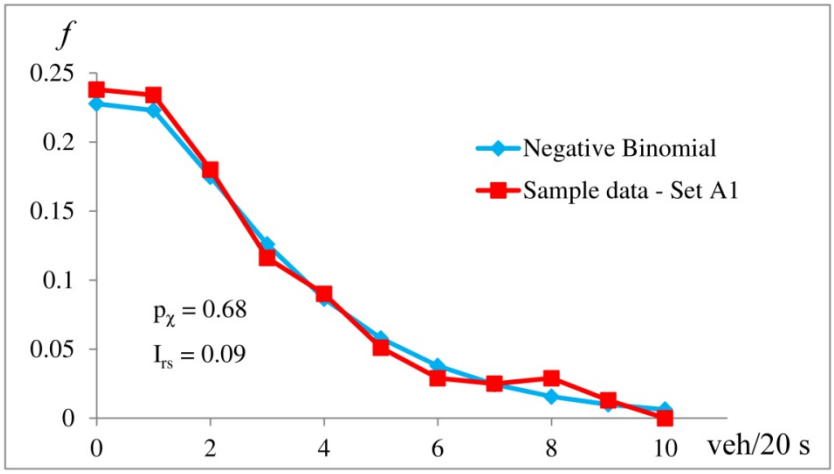



Figure 4. Frequency distributions for set A1

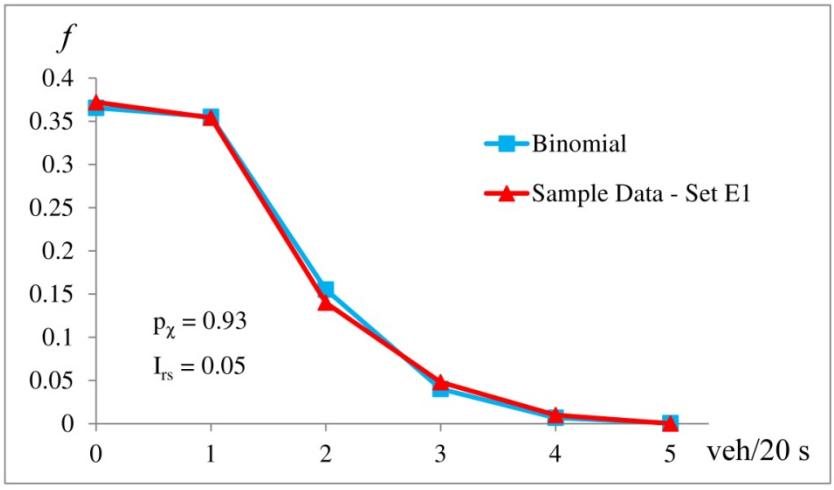



Figure 5. Frequency distributions for set E1

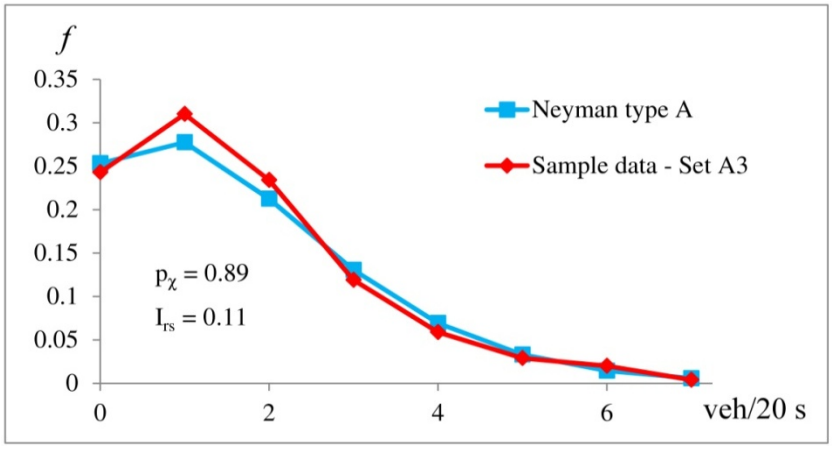



Figure 6. Frequency distributions for set A3

Finally, links between the parameters p and k, which define the negative binomial model, and the flow rate Q have been investigated by means of the least-squares method; the following equations were obtained (see also Figure 7 and Figure 8):

$$p = 0.9784e^{-0.0024Q}, Q > 50 vph; r = 0.885 \quad (16)$$

$$k = 34.93Q^{-0.543}; r = 0.557 \quad (17)$$

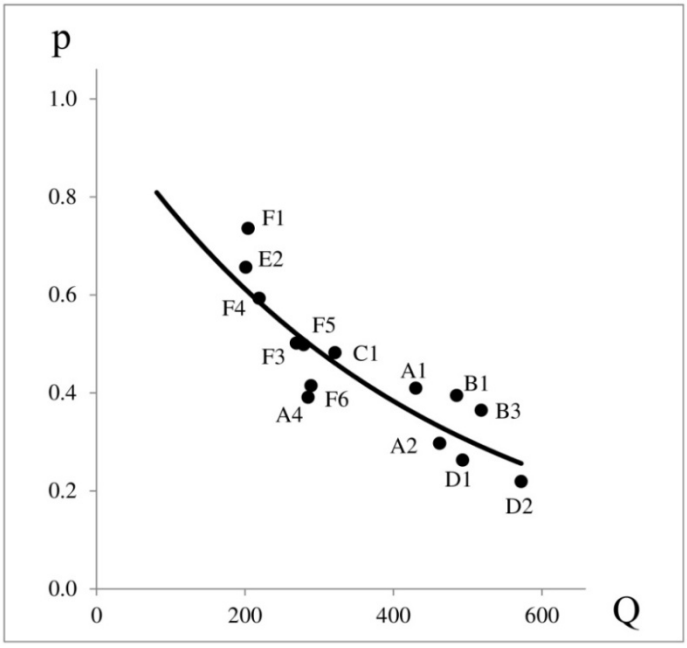



Figure 7. Regression curve between the negative binomial parameter p and the flow rate

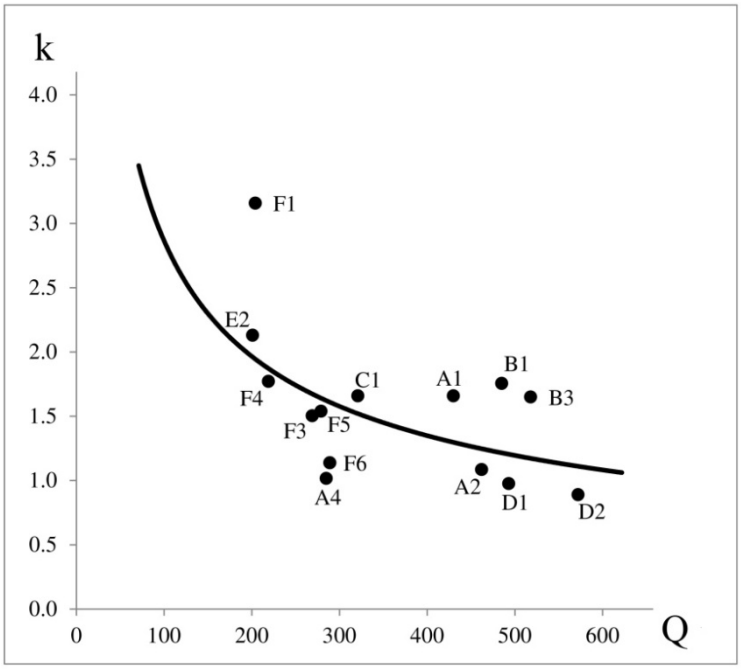



Figure 8. Regression curve between the negative binomial parameter k and the flow rate

As mentioned above, a similar analysis had been carried out by Esposito and Mauro (1994); in that paper traffic countings were carried out in 2892 intervals with a width $\Delta t = 20$ sec, for a total of 4745 vehicular passages. These data were collected on three two-lane rural highways between 1989 and 1992; each road had different geometric alignment features, as well as different overtaking rules.

Esposito and Mauro (1994) identified 33 steady-state periods; their data (minimum and maximum values) are summarised in Table 8.

Table 8. Summary of data collected by Esposito and Mauro (1994)

| Number of stationarity period | Length of stationarity period (minutes) | Flow rate (veh/h) | I and L estimate | | Parameters | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | negative binomial | | binomial | |
| | | | I | L | 28 periods | | 5 periods | |
| | | | | | p | k | p | n |
| 33 | min. 11 | min. 113 | min. 0.53 | min. 0.35 | min. 0.204 | min. 0.807 | min. 0.032 | min. 3 |
| | max. 60 | max. 630 | max. 4.91 | max. 8.49 | max. 0.828 | max. 4.808 | max. 0.475 | max. 20 |

The regression curve equations, obtained by Esposito and Mauro (1994), are as follows:

$$p = 1.154e^{-0.0028Q}, Q > 50vph; r = 0.876 \quad (18)$$

$$k = 227Q^{-0.857}; r = 0.626 \quad (19)$$

These curves are shown in Figure 9 and Figure 10, together with the regression curves of the parameters p and k obtained by jointly examining data in the paper by Esposito and Mauro (1994) and data coming from the above said surveys carried out in the spring of 2012; the following equations were obtained:

$$p = 1.1022e^{-0.0027Q}, Q > 50vph; r = 0.876 \quad (20)$$

$$k = 137.34Q^{-0.773}; r = 0.610 \quad (21)$$

By comparing the relations at different times, the curves appear to be substantially corresponding to flow rate values which are located approximately in the central part of the examined range ($300 < Q < 500$); instead, the curves move away in correspondence with the extreme values, but not in a pronounced way. The curves obtained from the complete data analysis (data collected during the years 1989-1992 and 2012) are very close to those

found in the paper by Esposito and Mauro (1994) (data collected during the years 1989-1992); this can be easily explained by considering that the sample points $(\bar{I}, \bar{L})$ obtained by traffic countings made in 2012 are the half than those resulting from traffic countings carried out by Esposito and Mauro (1994).

Although carried out at a distance of about two decades, the results obtained from the two traffic counting surveys are essentially consistent with one another; they can be used in simulations which do not allow performing a complete empirical analysis and in parameter estimation for arrival statistical distributions, on condition that there is compliance with the negative binomial theoretical distribution.

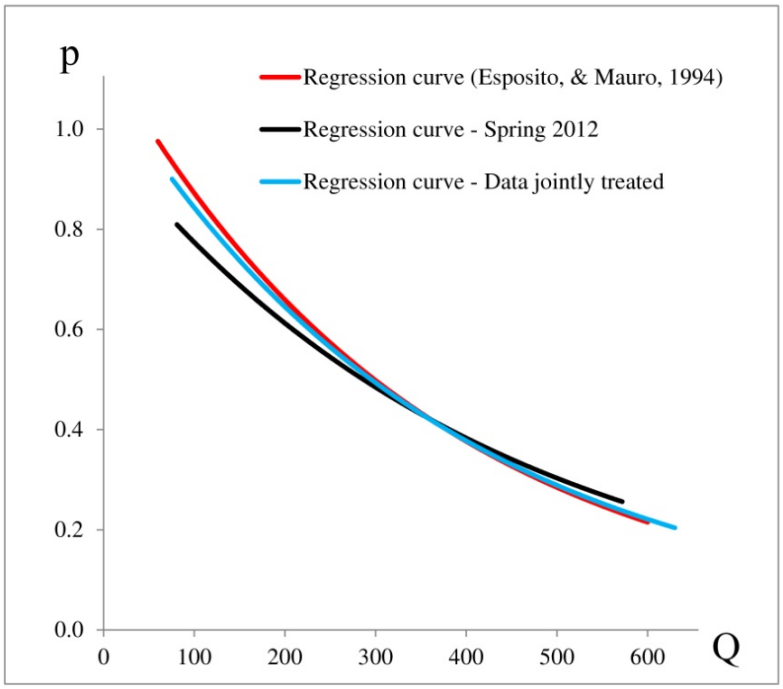



Figure 9. Regression curve between the negative binomial parameter p and the flow rate

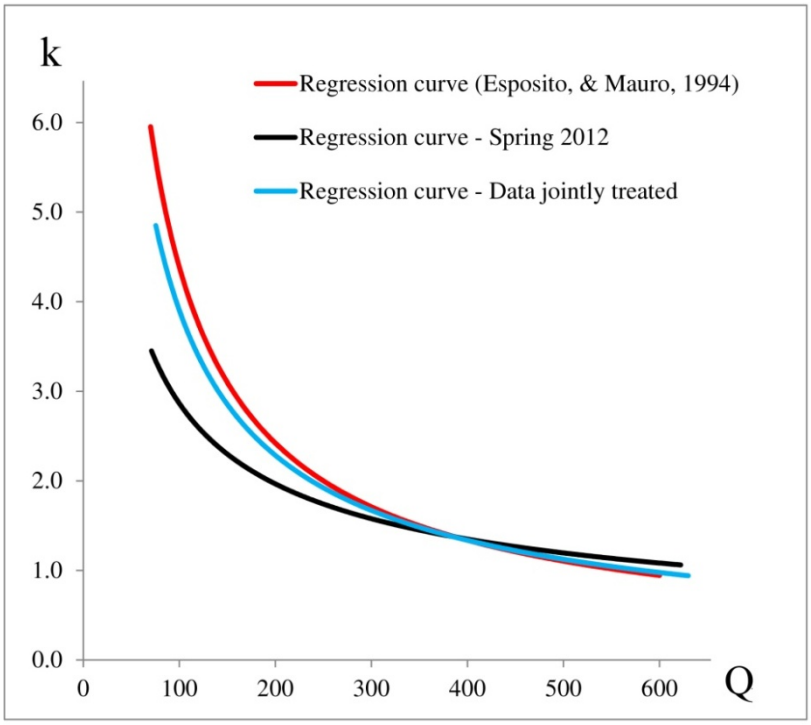



Figure 10. Regression curve between the negative binomial parameter k and the flow rate

## 5. Conclusions

The paper deals with the identification, calibration and validation of statistical models for countings of two-lane rural highways, especially with the definition and determination of the statistical stationarity periods of the flow rate, on the basis of the observed data analysis.

After a brief review of probabilistic arrival models usually used in the study of traffic phenomena, a complete statistical analysis methodology has been developed to define stationarity periods, to verify the independence of the events (the sequence of arrivals) as well as to identify the probability laws for traffic countings. The procedure was applied to some samples of traffic counting, empirically surveyed on two two-lane rural highways; then, links were established between flow rate and parameters of the negative binomial distribution (found to comply with most of the data). However, the same analysis was not performed in the binomial, Poisson and Neyman type A distributions because of the very limited availability of data. Finally, a comparison was made between the results from the 2012 data collection and those reported in the previous paper by Esposito and Mauro (1994); the latter applied the same procedure to other data sets.

The analysis of the data collected specifically for this study confirms what already mentioned by Esposito and Mauro (1994); indeed, it shows that the arrivals analysed in the two roads are mostly well modeled by the negative binomial counting distribution (which is an “aggregate” or “contagious” distribution), and other data (though insufficient for in-depth analysis) is well modeled by the Neyman type A distribution, which is defined as an “aggregate” or “contagious” distribution as well. These results are consistent with the fact that the flow mainly recorded on the two examined infrastructures is a commonly called “platoon” flow. So, in general and even for low flow values, the Poisson law seems to be unsuitable for representing the arrivals on two-lane rural highways; instead, other models are recommended in these cases. Such models have to take into account the vehicle group formation influenced by a leader.

Finally, as regards the relationships between the parameters of the negative binomial model and the flow rate, the comparison between the relations from this study and those by Esposito and Mauro (1994) has shown that the curves are substantially corresponding each other for flow rate values which are located approximately in the central part of the examined range ($300 < Q < 500$); instead, the curves move away in correspondence with the extreme values, but not in a pronounced way.

The results obtained from the two traffic measurement surveys, although carried out at a distance of about two decades, are essentially in good agreement, and they may be useful comparison tools for parameter estimates of arrival distributions, on condition that there is compliance with the negative binomial distribution. Such results can be useful in simulations when a complete empirical analysis cannot be performed. However, the limited size of the sample data analysed does not allow to extend the relationships found to other two-lane rural highways without taking proper precautions or performing adequate sensitivity analyses.

## References

Adams, W. F. (1936). Road traffic considered as a random series. *J. Inst. Civ. Engrs, 4*, 121-130. http://dx.doi.org/10.1680/ijoti.1936.14802

Beckmann, M. J., McGuire, C. B., & Winsten, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press.

Bertò, B., Schoen, F., & Speranza, M. G. (1996). Discrete time point processes: application to road traffic. In M. Papageorgiou (Eds.), *Concise Encyclopedia of traffic & transportation systems*. Pergamon Press.

Breiman, L. (1962). On some probability distributions occurring in traffic flow. *Bull. Inst. Statist. Inst.*, 155-161.

Breiman, L. (1963). The Poisson tendency in traffic distribution. *Ann. Math. Statist., 34*, 308-311. http://dx.doi.org/10.1214/aoms/1177704267

Buckley, D. J. (1965). *Inter-vehicle spacing and counting distributions*. Ph.D. Thesis, School of Traffic Engineering of New South Wales, Sydney.

Buckley, D. J. (1967). Road traffic counting distributions. *Transp. Res., 1*(2), 105-116. http://dx.doi.org/10.1016/0041-1647(67)90166-9

Buckley, D. J. (1968). A semi-Poisson model for traffic flow. *Transp. Sci., 2*(2), 107-133. http://dx.doi.org/10.1287/trsc.2.2.107

Cao, M., Tai, A. H., & Chan, L. Y. (2012). New statistical distributions for group counting in Bernoulli and Poisson processes. *Test, 21*, 29-53. http://dx.doi.org/10.1007/s11749-010-0232-0

Clementi, A., Monti, A., & Silvestri, R. (2011). Modelling mobility: A discrete revolution. *Ad Hoc Networks, 9*, 998-1014. http://dx.doi.org/10.1016/j.adhoc.2010.09.002

Cox, D. R., & Lewis, P. A. (1966). *The statistical analysis of series of events*. New York, NY: J. Wiley & S.

Drew, D. R. (1965). Application of discrete distributions to traffic. *Traff. Eng. Control, 7*(8), 24-25.

Esposito, T., & Mauro, R. (1994). Analysis of counting distributions for two-lane two-way roads (in italian). *Le Strade, 5*, 433-440.

Gerlough, D. L. (1955). *Use of Poisson distribution in highway traffic*. Saugatuck, CT: The Eno Foundation for Highway traffic control.

Gerlough, D. L., & Barnes, F. C. (1971). *The Poisson and other probability distributions in highway traffic*. Saugatuck, CT: The Eno Foundation for Highway traffic control.

Gerlough, D. L., & Huber, M. J. (1975). Traffic flow theory, a monograph. *TRB Special Report, 165*, Washington, D.C.

Gore, A., & Paranjpe, S. (2001). *A course in mathematicaland statistical ecology*. Dordrecht, The Netherlands: Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-94-015-9811-8

Greenshields, B. D., Shapiro, D., & Ericksen, E. L. (1947). *Traffic Performance at Urban Street Intersections*. New Haven, CT: Tech. Rep. No. 1, Yale Bureau of Highway Traffic.

Ha, D. H., Aron, M., & Cohen, S. (2012). Time Headway variable and probabilistic modeling. *Transportation Research Part C, 25*, 181-201. http://dx.doi.org/10.1016/j.trc.2012.06.002

Haight, F. A. (1959). The generalized Poisson distribution. *Ann. Inst. Statist. Math, 11*, 101-105. http://dx.doi.org/10.1007/BF01737397

Haight, F. A., Whilser, B. F., & Mosher, W. W. (1961). New statistical method for describing highway distribution of cars. *Proc. Highw. Res. Board, 40*, 557-563.

Jabari, E. E., & Liu, H. X. (2012). A stochastic model of traffic flow: Theoretical foundations. *Transportation Research Part B, 46*, 156-174. http://dx.doi.org/10.1016/j.trb.2011.09.006

Johnson, A. N. (1928). Maryland aerial survey of highway traffic between Baltimore and Washington. *Highw. Res. Bd., 8*, 106-115.

Kendall, H. G., & Stuart, A. (1967). *The Advanced Theory of Statistics-Vol. II*. London, UK: Griffin.

Kinzer, J. P. (1933). Application of the Theory of Probability to Problems of Highway Traffic. *Inst. Traff. Engrs., 5*, 118-124.

May, A. D. (1990). *Traffic flow fundamentals*. New Jersey, NY: Prentice Hall.

Mauro, R., & Branco, F. (2012). Two Vehicular Headways Time Dichotomic Models. *Modern Applied Science, 6*(12), 1-12. http://dx.doi.org/10.5539/mas.v6n12p1

Miller, A. M. (1970). An empirical model for multilane road traffic. *Transp. Sci., 4*(2), 164-186. http://dx.doi.org/10.1287/trsc.4.2.164

Oliver, R. M., & Thibault, B. (1962). A highflow traffic counting distribution. *Highw. Res. Board Bull., 356*.

Ord, J. K. (1972). *Families of frequency distribution*. London, UK: Griffin.

Serfling, R. J. (1969). Non-Poisson models for traffic flow. *Transp. Res., 3*(3), 299-306. http://dx.doi.org/10.1016/S0041-1647(69)80015-8

Wald, A. (1947). *Sequential Analysis*. New York, NY: J. Wiley.

Weiss, G. H., & Hermann, R. (1962). Statistical properties of low-density traffic. *Q. Appl. Math., 20*, 121-130.

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# Effect of Various Carriers and Storage Temperatures on Survival of *Azotobacter vinelandii* NDD-CK-1 in Powder Inoculant

Marisa Phiromtan[1], Thongchai Mala[2] & Peerasak Srinives[3]

[1] Agricultural Research and Development Program (International), Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen, Nakhon Pathom, Thailand

[2] Department of Soil Science, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen, Nakhon Pathom, Thailand

[3] Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen, Nakhon Pathom, Thailand

Correspondence: Peerasak Srinives, Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, Thailand. Tel: 66-8-9894-8978. E-mail: agrpss@yahoo.com

Received: March 18, 2013 Accepted: May 10, 2013 Online Published: May 21, 2013

doi:10.5539/mas.v7n6p81 URL: http://dx.doi.org/10.5539/mas.v7n6p81

*This research was supported by a grant under the Program Strategic Scholarships for Frontier Research Network for the Joint Ph.D. from the Office of the Higher Education Commission, Thailand*

**Abstract**

A study was carried out for determining the effect of various carriers and storage temperatures on survival of *Azotobacter vinelandii* NDD-CK-1. The experiment was laid out using a 4 x 5 factorial treatment arrangement in a Completely Randomized Design with three replications. The first factor is carrier with four kinds, viz. peat (Pt), peat mixed with corn stubble compost (PtCC), peat mixed with golden flamboyant leaf compost (PtLC), and Pt mixed with mushroom waste compost (PtMC). The second factor is storage temperature with five levels, viz. -16 $^{o}$C, 5 $^{o}$C, 25 ± 2 $^{o}$C, 30 ± 2 $^{o}$C and 37.5 ± 2.5 $^{o}$C. Inoculum of *Azotobacter vinelandii* NDD-CK-1 was produced by a standard method using various carriers. The results revealed that types of carrier, storage temperatures and interaction between them showed significant effect on survival of azotobacter during 7 to 90 days. The survival rate was the highest in PtLC, followed by PtCC, PtMC, and Pt which gave the log number of bacterial viable cell of 6.41, 6.02, 5.67 and 5.50, respectively. The proliferation of azotobacter decreased with time and increasing temperature. The appropriate storage temperature at 7 to 15 days was -16 $^{o}$C, while the most suitable temperatures for longer term (30 to 90 days) was 5 $^{o}$C; followed by -16 $^{o}$C, 25 ± 2 $^{o}$C, 30 ± 2 $^{o}$C and 37.5 ± 2.5 $^{o}$C. The highest survival of azotobacter was found in PtCC at -16 $^{o}$C (9.98 log cfu/g), similar to PtCC at 5 $^{o}$C, PtLC at -16 $^{o}$C, and PtLC at 5 $^{o}$C (9.92, 9.85 and 9.77 log cfu/g, respectively).

**Keywords:** azotobacter, carrier, inoculum, temperature

## 1. Introduction

Application of biofertilizer for crop production is environmental friendly and sustainable for ecological system. Several types of biofertilizer have been developed from bacteria, particularly *Rhizobium* spp., *Azospirillum* spp., and *Azotobacter* spp., and used in production of various plants (Mala, 2003; Narula, 2000; Rai, 2006). *Azotobacter* spp., a free-living $N_2$-fixing bacteria is a beneficial biofertilizer which has profitable effects on plants and soil fertility (Holt, Novel, Peter, James, & Stanley, 2000). Azotobacter inoculant is normally produced in powder form for soil and seed inoculation (Burton, 1984). The inoculant can be prepared from several types of carriers such as peat, charcoal, farmyard manure, lignite, alginate, etc., in a standard method similar to rhizobial inoculum. Although peat is an ideal carrier for rhizobial inoculant, there are many disadvantages in using it. The quality of peat varies depending on its sources. It is rather expensive commodity and yet not widely available in some countries such as India (Sadasivam et al., 1986), and Thailand, thus causing significant drawback in using peat. Nowadays, several types of agricultural waste like maize stubble, plant compost, mushroom waste, rice straw, oil palm frond and bunch can be composted and used as bioinoculant carriers for rhizobial industry and

others. This system helps reducing the pollutants, saving energy, decreasing cost of production, and utilizing natural resources to the maximum benefit. In Thailand, agricultural waste from maize production is considered a major one. Annually, over 1.07 million hectares of maize were sown with the production of 4.25 Mt. Over 50 % of the plant parts is discarded and thrown away. Similarly, waste created from mushroom cultivation is also high due to increasing in production of various mushrooms in Thailand such as shitake, Indian oyster, oyster, yanagi mutsutake, golden needle mushroom, straw mushrooms, etc. This results cause a large amount of wastes each year. These agricultural wastes are recycled in various ways, including a raw material for production of organic fertilizer and bioinoculum carriers. Inoculum carriers serve as media in bioinoculant production, controlling quality and shelf life of bacterial inoculants by serving as microenvironment for microorganisms. Besides, types of carrier and storage temperatures are important factors determining shelf life of bioinoculants (Kremer & Peterson, 1983), and acceptance of agricultural products (Bashan, 1998). Storing such inoculants in a warehouse without refrigerator in the range of -5 to 30 $^{o}$C often causes reduction in microbial longevity. Many researchers have evaluated for suitable carriers from agricultural wastes and investigated effect of temperatures on shelf life in packages at various temperatures (Thungtrakul, 1987; Rajakumar & Lakshmanan, 1995; Saleh, Nassar, & Yassen, 2001). The current study is designed to compare effect of carriers made of peat with agricultural waste composts as well as storage temperatures on longevity of azotobacter in powder inoculum.

## 2. Materials and Methods

This study was conducted during November 2009 to April 2010 in the Laboratory of Soil Microbiology, Department of Soil Science, Kasetsart University, Kamphaeng Saen Campus**,** Nakhon Pathom, Thailand. The procedures were described as followed.

### *2.1 Preparation and Analysis on Chemical Properties of Carriers*

Four types of materials including peat (Pt), corn stubble compost (CC), golden flamboyant leaf compost (LC) and mushroom media compost (MC) were used in this study. The materials were prepared following the manual of Burton (1984). The raw materials were ground, sieved with 0.5 cm mesh screen and dried in a hot air oven at 60 $^{o}$C for two days. Then, the materials were prepared into carriers of peat (Pt), Pt mixed with each agricultural waste compost at a ratio of 1:2 and named as peat with corn stubble compost (PtCC), peat with golden flamboyant leaf compost (PtLC), and peat with mushroom media compost (PtMC). The materials were autoclave at 121 $^{o}$C at a pressure of 15 psi for 30 min. The carriers were analyzed for chemical properties, including total nitrogen ($N_{tot}$) by micro Kjedahl method, total phosphorous ($P_{tot}$) by vanadomolybdate yellow color, total potassium ($K_{tot}$) by turbidimetric techniques, organic matter (OM) by Walkley and Black, pH (1:10) and electrical conductivity (EC) by standard method (Soil Science Society of America, 1996).

### *2.2 Experimental Design*

The experimental design was a 4 x 5 factorial arrangement in a Completely Randomized Design (CRD) with three replications. The first factor comprised four carriers, viz. Pt, PtCC, PtLC, and PtMC, while the second factor comprised five storage conditions, viz. deep freezing (-16 $^{o}$C), refrigerating (5 $^{o}$C), air conditioning, (25 ± 2 $^{o}$C), ambient room temperature (30 ± 2 $^{o}$C) and greenhouse temperature (37.5 ± 2.5 $^{o}$C).

### *2.3 Production of Azotobacter Inoculum*

*Azotobacter vinelandii* NDD-CK-1, a fast growing isolate screened from rhizosphere of Chinese kale at Nadindam Village, Loei Province, Thailand was chosen based on its high effectiveness. Inoculums were prepared in powder form under aseptic condition. A loopful of NDD-CK-1 pure culture was transferred into a 250 ml erlenmeyer flask containing 100 ml of Ashby's broth (20 g mannitol, 0.2 g $K_2HPO_4$, 0.2 g $MgSO_4.7H_2O$, 0.2 g NaCl, 0.1 g $K_2SO_4$, 5 g $CaCO_3$, 0.05 g $Na_2MoO_4$, and 1 g $NH_4Cl$ per liter of distilled water) and incubated at 28 ± 2 $^{o}$C on 120 rpm rotary shaker for 72 hrs. Seventy-five milliliters of broth culture was mixed thoroughly with 100 g of each sterile carrier, adjusted the moisture content to 75 % water holding capacity, packed in polyethylene bags, sealed and incubated under room temperature for five days. The inoculums were repacked into sterile polyethylene bags.

### *2.4 Evaluation for Survival of the Azotobacter During Storage in Different Temperatures*

The number of azotobacter was determined after the inoculum was subjected to different carriers and temperatures. Ten grams of each sample was taken for estimating viable cells at the initial date, 7, 15, 30, 60 and 90 days after storage using dilution plating method on Ashby's agar and incubated at 28 ± 2 $^{o}$C for 5-7 days (Mala, 2003). The number of apparent azotobacter colonies were counted, calculated into viable cells and converted to log number per gram of dry inoculum.

*2.5 Statistical Analysis*

The collected data were statistically analyzed using R-program for Windows, version 2.15.1, and the treatment means were compared using Duncan's Multiple Range Test (DMRT) method.

## 3. Results

*3.1 Chemical Properties of Inoculum Carriers*

Some chemical properties of the carriers are shown in Table 1. There was a signigificantly difference among chemical properties of the carriers. $N_{total}$ of the carriers were considered low. PtMC had the highest $N_{total}$ of 1.40 %, while PtCC and PtLC were 0.83 % and 0.86 %, respectively. For $P_{total}$ content, PtMC was the highest (0.47 %) while Pt was the lowest (0.07 %). $K_{total}$ of PtMC was the highest at 2.21 % while those of PtCC and Pt were the lowest at 0.24 % and 0.27 %, respectively. The pH of all carriers showed high acidity (< 4.5). The highest pH was found in PtLC and PtMC (3.63 and 3.56, respectively), whereas that of Pt was the lowest at 2.80. EC of PtMC was the highest at 3.12 dS/cm, while that of PtLC was the lowest at 0.45 dS/cm. All materials were high in OM, with the highest in Pt (39.87 %) and the lowest in PtLC (31.44 %).

Table 1. Chemical properties of Pt, PtCC, PtLC and PtMC used as inoculum carriers for azotobacter

| Carriers | Chemical properties | | | | | |
|---|---|---|---|---|---|---|
| | $N_{tot}$ (%) | $P_{tot}$ (%) | $K_{tot}$ (%) | OM (%) | pH (1:10) | EC (dS/cm) |
| Pt | 1.00b | 0.07d | 0.27c | 39.87a | 2.80c | 1.19b |
| PtCC | 0.83c | 0.39b | 0.24c | 37.18ab | 3.38b | 0.60c |
| PtLC | 0.86c | 0.14c | 0.98b | 31.44c | 3.63a | 0.45d |
| PtMC | 1.40a | 0.47a | 2.21a | 34.26bc | 3.56ab | 3.12a |
| Mean | 1.02 | 0.27 | 0.93 | 35.69 | 3.34 | 1.34 |
| *P*-value | <.0001 | <.0001 | <.0001 | 0.0303 | <.0001 | <.0001 |
| CV (%) | 2.60 | 2.11 | 7.13 | 7.89 | 2.87 | 4.33 |

Remark: mean values in each column followed by the same letter are not significantly different by DMRT at P≤ .05.

*3.2 Survival of Azotobacter in the Carriers During Storage in Different Temperatures*

There were significant differences in survival of *Azotobacter vinelandii* NDD-CK-1 among carriers, storage temperatures at initial date to 90 days and interaction between carriers and temperatures during 7 to 90 days after storage (Figures 1 to 4). At the initial date, PtCC gave the highest bacterial number followed by Pt, PtLC and PtMC with 12.48 (Figure 2), 12.40 (Figure 1), 11.02 (Figure 3), and 10.96 log cfu/g (Figure 4), respectively. During 7, 15, 30, 60 and 90 days across five temperature regimes (Figure 5), PtLC had the highest bacterial count (10.04, 9.36, 8.88, 7.36 and 6.41 log cfu/g), followed by PtCC (9.60, 9.18, 8.86, 7.32 and 6.02 log cfu/g, PtMC (9.63, 9.00, 8.4, 6.78 and 5.67 log cfu/g) and Pt (8.20, 7.76, 7.56, 6.15 and 5.50 log cfu/g), respectively.

For storage temperatures, storing in deep freeze gave the highest growth rate of azotobacter at 7 to 15 days after preservation. During 7 days after storage, the survival of bacteria in deep freeze were 9.72 log cfu/g, while those stored in the refrigerator, green house, ambient temperature, and air-conditioned temperature had viable cells of 9.44, 9.33, 9.22 and 9.14 log cfu/g, respectively (Figure 6). At 15 days after storing, keeping the inoculants in deep freezer gave the highest number of azotobacter (9.35 log cfu/g) similar to that in refrigerator, but higher than those under air-conditioning temperature, ambient temperature and greenhouse condition which gave 8.59, 8.49 and 8.47 log cfu/g, respectively (Figure 6). While, at 30, 60 and 90 days after storage, preserving the inoculum in refrigerator gave the highest survival rate of 9.47, 8.03 and 7.52 log cfu/g, respectively, followed respectively by those kept in deep freezer (9.15, 7.97, and 7.30 log cfu/g), air conditioning temperature (8.17, 6.22 and 5.19 log cfu/g), ambient temperature (7.90, 6.21 and 5.17 log cfu/g) and greenhouse condition (7.46, 6.08 and 4.34 log cfu/g) as shown in Figure 6.

The interaction between carrier and storage temperature during 7 to 90 days showed that PtLC stored at 37.5 ± 2.5 $^{o}$C gave the largest microbial population (10.37 cfu/g) (Figure 3), similar to that of PtCC at -16 $^{o}$C (Figure 2), PtLC at 5 $^{o}$C (Figure 3) and PtLC at -16 $^{o}$C (Figure 3) with 10.32, 10.15 and 10.14 log cfu/g, respectively. While,

Pt at 5 $^{o}$C and at 25 ± 2 $^{o}$C had the lowest survival rates at one week after storage (Figure 1). At 15 days after storage, stored PtCC at -16 $^{o}$C gave the highest survival rate (9.98 log cfu/g) (Figure 2), similar to PtCC at 5 $^{o}$C (Figure 2), PtLC at -16 $^{o}$C (Figure 3), and PtLC at 5 $^{o}$C (Figure 3) (9.92, 9.85 and 9.77 log cfu/g, respectively). Whereas maintaining Pt carrier at 37.5 ± 2.5 $^{o}$C gave the lowest number of azotobacter (7.26 log cfu/g) (Figure 1). At 30 days after preservation, maintaining the inoculant in PtCC at 5 $^{o}$C showed the highest bacterial number (10.08 log cfu/g) (Figure 2), similar to those maintained in PtLC at 5 $^{o}$C and PtLC at -16 $^{o}$C (Figure 3) (10.03 and 10.00 log cfu/g, respectively). In contrast, Pt at 30 ± 2 $^{o}$C had the lowest bacterial population (6.69 log cfu/g) (Figure 1), which was close to PtMC at 37.5 ± 2.5 $^{o}$C (Figure 4), PtLC at 37.5 ± 2.5 $^{o}$C (Figure 3), Pt at 25 ± 2 $^{o}$C and Pt at 37.5 ± 2.5 $^{o}$C (Figure 1) at 7.65, 7.48, 7.43 and 6.88 log cfu/g, respectively. Interaction between carrier and temperature had a significant effect on the survival of azotobacter stored for 60 days. The maximum number of bacteria was obtained from PtCC stored at 5 $^{o}$C (Figure 2), similar to PtLC at 5 $^{o}$C (Figure 3) and PtCC at -16 $^{o}$C (Figure 2), gave 8.93, 8.88 and 8.70 log cfu/g, respectively. While, Pt at 37.5 ± 2.5 $^{o}$C gave the lowest microbial population of 5.70 log cfu/g (Figure 1). At 90 days of preservation, the highest survival rate of azotobacter was found in PtLC at 5 $^{o}$C (8.42 log cfu/g) (Figure 3) and PtCC at 5 $^{o}$C (8.37 log cfu/g) (Figure 2) followed by PtCC at -16 $^{o}$C (7.91 log cfu/g) (Figure 2). While PtCC at 37.5 ± 2.5 $^{o}$C (Figure 2) and PtMC at 25 ± 2 $^{o}$C (Figure 4) had the lowest survival cells at 3.69 and 3.71 log cfu/g, respectively.

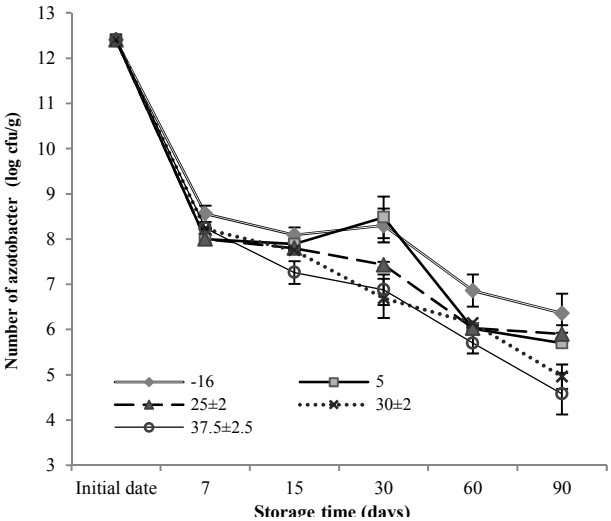



Figure 1. Population of *Azotobacter vinelandii* NDD-CK-1 existed in Pt carrier during initial date to 90 days after storage in different temperatures

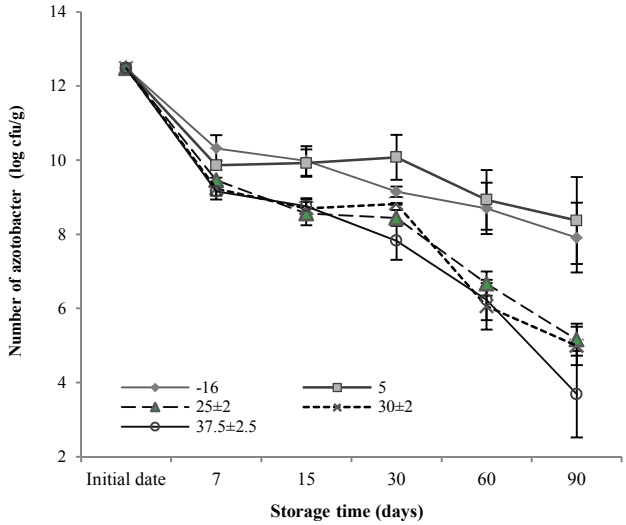



Figure 2. Population of *Azotobacter vinelandii* NDD-CK-1 existed in PtCC carrier during initial date to 90 days after storage in different temperatures

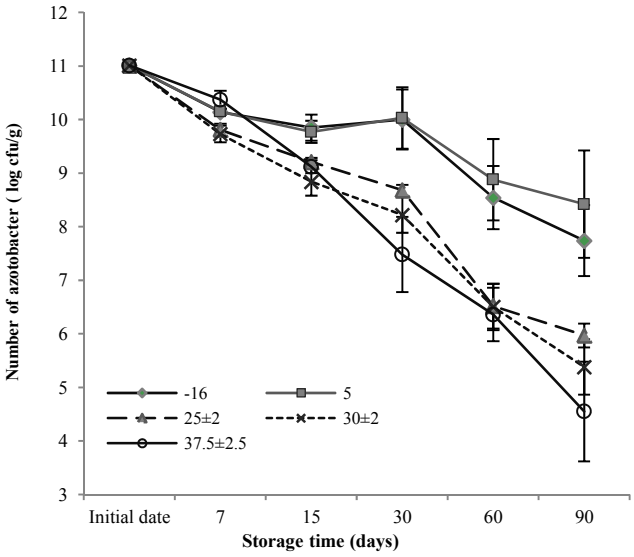



Figure 3. Population of *Azotobacter vinelandii* NDD-CK-1 existed in PtLC carrier during initial date to 90 days after storage in different temperatures

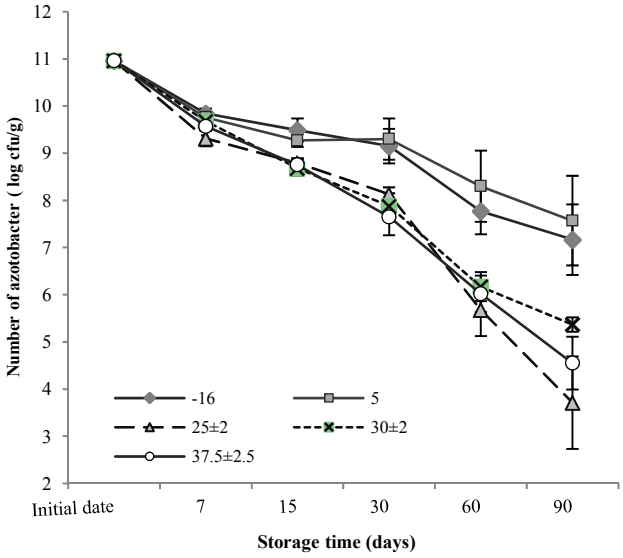



Figure 4. Population of *Azotobacter vinelandii* NDD-CK-1 existed in PtMC carrier during initial date to 90 days after storage in different temperatures

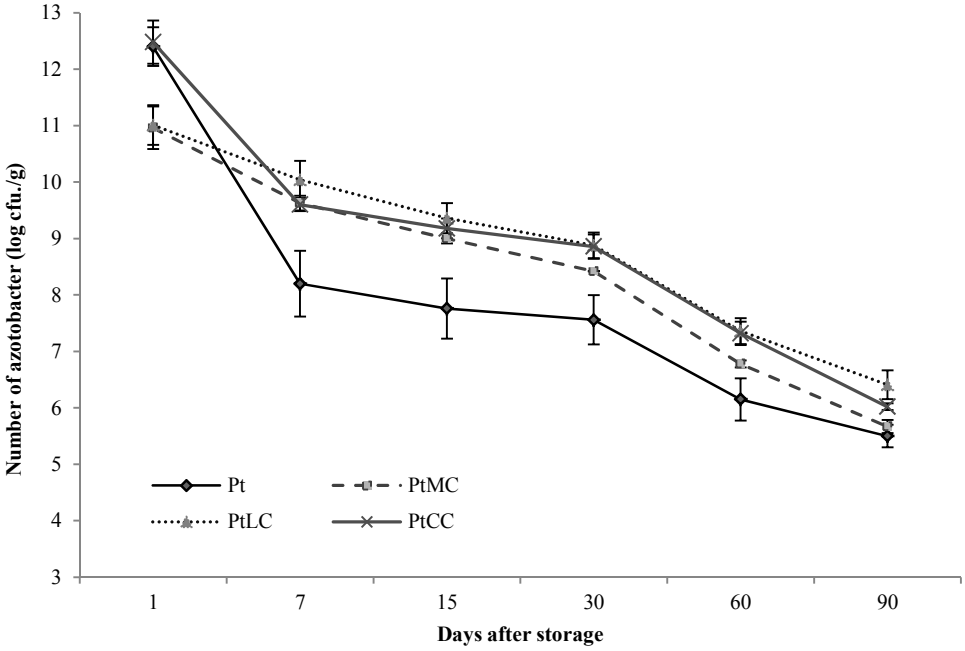



Figure 5. Effect of P, PMC, PtLC, and PCC carriers on mean number of *Azotobacter vinelandii* NDD-CK-1 averaged across temperatures

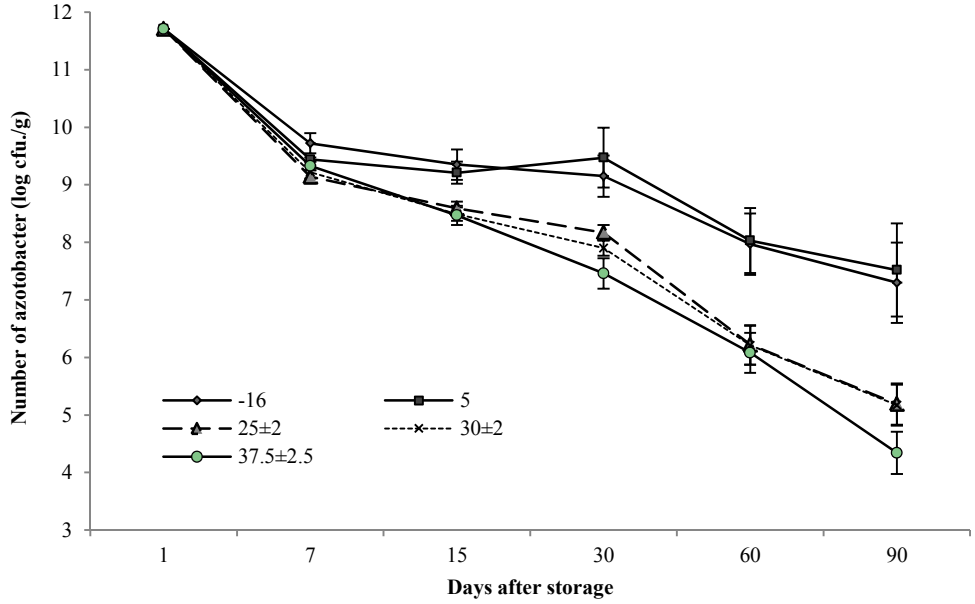



Figure 6. Effect of various storage temperature on mean number of *Azotobacter vinelandii* NDD-CK-1 averaged across carriers

## 4. Discussion

Population of *Azotobacter vinelandii* NDD-CK-1 decreased over time depending on types of carrier and storage temperatures. (Figure 5), similar to earlier reports (Youssef, Sedik, Fayez, & Hegazi, 1997; Saleh et al., 2001). Initial microbial populations of those materials were higher than those reported by Muthuselvam & Tholkappian (2008). The density of bacteria sharply dropped at 7 days due to lack of moisture and nutrients of the carriers. In this case, OM, $N_{total}$, $P_{total}$ and $K_{total}$ as well as moisture content of inoculum carriers were almost depleted by time due to bacterial activities and storage conditions while transitioning from logarithmic to stationary phase during incubation (Neidhardt, Ingraham, & Schaechter, 1990; Tate, 2000). The bacterial population slightly

declined until 15 days. During 7-30 days of storage the death rate was the smallest but greatly decreased from 30 to 60 days and leveled off at 90 days of storage (Figure 5). In this study, the survival of azotobacter in PtLC, PtCC and PtMC carriers were satisfactory in comparison with the biofertilizer standard, except in Pt carrier after storage for 90 days. However, the microbial population was less than those reported by many workers (Wangaruro, Karanja, Makatiani, Odee, & Woomer, 2000; Madan & Singh, 2010; Raja & Karmegam, 2010). They found that populations of *Azotobacter chroococcum* did not reduce below $10^8$ cfu/g within three months.

*4.1 Effect of Carriers on the Population of Azotobacter*

The results indicated that PtLC was the most suitable carrier for production of azotobacter inoculum, followed by PtCC, PtMC and Pt (Figures 1-4). LC had more nutrients, especially N content than the others possibly from leaves of the legume yellow flamboyant. PtLC also had the highest pH but lowest EC (Table 1) that was suitable for promotiing microbial growth. PtLC may also release less toxic compounds during sterilization as compared to peat (Burton, 1984; Bashan, 1998; Anonymous, 2006). This compost also had high clay mineral derived from adding of clayey soil during composting process which played a critical function in promoting physical and biochemical environment for the microbial population. The increase in high specific surface area of PtLC can promote adsorption of organic and inorganic substances, cation exchange capacity, and water holding capacity. In addition, clay particles also encourage microbial catabolism by increasing adherence and tolerance capacity of azotobacter in the PtLC under hot condition (Tate, 2000). This finding is similar to the report of Bashan (1998) that survival of bacteria was increased by adding clay to the alginate beads as compared to alginate beads alone. Our finding showed that Pt gave poor survival of azotobacter during 90 days, similar to the report of Muthuselvam & Tholkappian (2008) and Thananusont (1993). Pt had lower quality and higher acidity than that of the others because over 50% of the peat was mixed with coconut husk which is considered a poor source of carrier. Thus, adding plant leaf compost and corn stubble compost to peat can improve the inoculum quality and longevity of azotobacter. Many researchers have suggested to use various combinations of carriers to prolong shelf life of bacterium inoculant (Jauhri, Bhatnagar, & Iswaran, 1979; Jauhri & Philip, 1984; Thungtrakul, 1987; Gaind & Gaur, 2004; Madan & Singh, 2010; Raja & Karmegam, 2010).

PtCC had the highest bacterial number at the initial date possibly due to its fine structure, which gave rise to higher specific surface area, more water content and nutrients spreading more thoroughly and easily adhered to microorganisms than those of the other carriers. This condition promoted bacterial growth within a few days after inoculation as compared to the condition in larger particle size carriers as in PtLC and PtMC. In contrast, PtMC had the lowest bacterial population due to the remaining of lignin and tannin from Para rubber saw dust, the major material of MC. The saw dust blocked bacterial respiration and reduced water absorption of MC resulting in less available water and inhibiting the bacterial population during incubation period.

*4.2 Effect of Storage Temperature on Population of Azotobacter*

The storage temperature affected proliferation of *Azotobacter vinelandii* NDD-CK-1in the carriers. Our study suggested that population of azotobacter was decreased by increasing temperatures (Figure 6). The survival of bacteria considerably dropped at 7 days and slightly declined during 15 to 30 days, similar to the result of Somasegaran (1985). After one month, number of azotobacter continuously declined until reaching the lowest number of viable cells at 60 and 90 days after storage. The suitable temperature for preservation of azotobacter until 90 days was 5 $^{o}$C, followed by -16 $^{o}$C which gave more viable cells than that stored at over 20 $^{o}$C. The extreme temperatures reduced the survival rate of bacteria and its metabolism (Slonnezewski & Foster, 2010). Our finding is similar to the reports of Ben Rebah, Tyagi, & Prévost (2002) and Kibunja (1991). At the refrigerating temperature, the bacteria had lower metabolism and physiological activity which maintained high mineral contents and more available moisture than that stored over 20 $^{o}$C, similar to the report of Bozida & Vladimir (1995). Inoculum stored at 5 $^{o}$C had longer shelf life without formation of ice crystal as that stored at -16 $^{o}$C. However, storing carriers at -16 $^{o}$C gave the highest bacterial population at 7 to 15 days and sharply declined afterwards. Preservation at -16 $^{o}$C for over a month caused water to become ice in the microbial cells and carriers, thus reduced the microbial density by initial killing at the time of freezing and afterwards (Neidhardt, Ingraham, & Schaechter, 1990).

*4.3 Interaction between Carriers and Temperatures*

The interaction between types of carrier and storage temperatures was similar to the report of Thungtrakul (1987), except at the initial storing period. At 7 days, the number of azotobacter in PtLC stored at 37.5 ± 2.5 $^{o}$C was higher than those at other temperatures. It may be due to the increasing of temperature from 17-19 $^{o}$C during incubation to 37.5 ± 2.5 $^{o}$C in storage which promoted population of bacteria during the first week without limitation of water and nutrients, and then fell down sharply afterwards. A similar result was reported by Saleh et

al. (2001) that the population of *Azotobacter vinelandii* A1 in rice husk carrier rose up to 128 % from the initial population after storing at 30 $^{o}$C. At 30 days after storage, the bacterial population increased slightly in Pt and PtLC stored at -16 $^{o}$C, and Pt, PtCC, PtMC and PtLC at 5 $^{o}$C.

During 7-15 days, PtCC stored at -16 $^{o}$C gave the maximum survival rate. The explaination may be that PtCC had fine particle which acted as the protective insulation from the damage of ice crystal under sub-zero temperature. In the case of Pt carrier, population of azotobacter was maximum at -16 $^{o}$C, followed by 5 $^{o}$C (Figure 5). Among types of carrier, Pt showed a robust property in keeping high population after storage at 37.5 ± 2.5 $^{o}$C, similar to PtLC and PtMC, but was higher than PtCC. Pt has lighter particle which protected bacterial cells from extreme temperatures better than the other carriers during 3 months. However, interaction between bacteria, carrier and temperature was a complicated mechanism limiting number of azotobacter under various environments. The survival ability of the bacteria varies depending on species and strains of azotobacter which are determined by gene expression under stressful conditions, habitat, and duration of preservation period in the medium.

## 5. Conclusion

PtLC is the most suitable inoculum carrier after storage for 90 days, followed by PtCC, PtMC, and Pt. PtLC and PtCC can be used as inoculum carriers of azotobacter with an advantage of utilizing natural resources to the maximum benefits, reducing the pollutant and inoculum cost. The usage of peat in combination with agricultural wastes resulted in high quality and survival of the bacteria. The inoculum can be kept for 90 days at 5 $^{o}$C followed by -16 $^{o}$C, 25 ± 2 $^{o}$C, 30 ± 2 $^{o}$C and 37.5 ± 2.5 $^{o}$C. It can be stored up to the maximum of one month at an ambient temperature and up to three months in a refrigerator. PtLC, PtMC and PtCC inoculums can be stored at temperatures above 25 $^{o}$C for a period of 1-2 months and still maintained a minimum of $10^{7}$cells/g. Adding clayey soil in the inoculum or during composition process is a beneficial technique for improving the quality of bioinoculant and protecting target microorganisms by serving as a micro-environment. An appropriate proportion of clay is necessary for improving product quality.

## Acknowledgements

This research was supported by a grant under the Program Strategic Scholarships for Frontier Research Network for the Joint Ph.D. from the Office of the Higher Education Commission, Thailand.

## References

Anonymous. (2006). *Biofertilizer Manual*. Japan Atomic Industrial Forum (JAIF).

Bashan, Y. (1998). Inoculants of plant growth-promoting bacteria for use in agriculture. *Biotechnology Advances, 16*, 729-770. http://dx.doi.org/10.1016/S0734-9750(98)00003-2

Ben Rebah, F., Tyagi, R. D. & Prévost, D. (2002). Wastewater sludge as a substrate for growth and carrier for rhizobia: the effect of storage conditions on survival of *Sinorhizobium meliloti*. *Bioresource Technology, 83*, 145-151. http://dx.doi.org/10.1016/S0960-8524(01)00202-4

Bozida, S., & Vladimir, M. (1995). Temperatures and dissolved oxygen concentration as parameter of *Azotobacter chroococcum* cultivation for use in biofertilizers. *Biotechnology Letters, 14*, .453-458.

Burton, J. C. (1984). *Legume Inoculant Production Manual*. NifTal Center-MIRCEN, University of Hawaii, Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, Hawaii, USA.

Gaind, S., & Gaur, A. (2004). Evaluation of fly ash as a carrier for diazotrophs and phosphobacteria. *Bioresource Technology, 95*, 187-190. http://dx.doi.org/10.1016/j.biortech.2003.07.014

Holt, J. G., Novel, R. K., Peter, H. A. S., James T. S., & Stanley T. W. (2000). *Bergey's Manual of Determinative of Bacteriology* (9th ed.). Lippincott Williams and Wilkins, USA.

Jauhri, K. S., Bhatnagar, R. S., & Iswaran, V. (1979). Survival of bacterial inoculants in a cheaper carrier material. *Current Science, 48*, 170-171.

Jauhri, K. S., & Philip, K. (1984). Pressmud: A potential carrier for Rhizobium and Azotobacter: II. Influence of storage temperature on the survival of Rhizobium and Azotobacter in different carriers and nodulation efficiency of rhizobial inoculants. *Zentralblatt für Mikrobiologie, 139*(2), 97-107.

Joshi, P., Rayalu, S., Bansiwal, A., & Juwarkar A. A. (2007). Surface modified zeolite, a novel carrier material for *Azotobacter chroococcum*. *Plant and Soil, 296*, 151-158. http://dx.doi.org/10.1007/s11104-007-9305-6

Kibunja, C. N. (1991). A local carrier for rhizobium inoculants in Kenya. *Thesis*. University of Nairobi, Kenya.

Kremer, R. J., & Peterson H. L. (1983). Effects of carrier and temperature on survival of *Rhizobium* spp. in legume inocula: development of an improved type of inoculant. *Applied and Environmental Microbiology, 45*, 1790.

Mala, T. (2003). *Organic Fertilizer and Biofertilizer: Production Techniques and Usage*. Bangkok, Thailand: Kasetsart University Press.

Madan, K., & Singh, A. K. (2010). Relative nitrogen fixing efficiency of *Azotobacter* isolates and shelf life of different organic carriers. *Environment and Ecology, 28*, 1655-1659.

Muthuselvam, K., & Tholkappian, P. (2008). Vermicompost: a potential carrier material for bacterial bioinoculants. *Plant Archives, 8*, 895-898.

Narula, N. (2000). *Azotobacter in Sustainable Agriculture*. New Delhi, India: CBS Publishers and Distributors.

Neidhardt, F. C., Ingraham, J. L., & Schaechter, M. (1990). *Physiology of the bacterial cell: A molecular approach*. Sunderland, Massachesetts, USA: Sinauer Associates, Inc. Publishers.

Rai, M. K. (2006). *Handbook of microbial biofertilizers*. New York, USA: Food Products Press, Binghamton.

Raja, S. K., & Karmegam, N. (2010). Earthworm casts as an alternate carrier material for biofertilizers: Assessment of endurance and viability of *Azotobacter chroococcum*, *Bacillus megaterium* and *Rhizobium leguminosarum*. *Scientia Horticulturae, 124*(2), 286-289. http://dx.doi.org/10.1016/j.scienta.2010.01.002

Rajakumar, K., & Lakshmanan, M. (1995). Influence of temperature on the survival and nitrogen fixing ability of *Azotobacter chroococcum*. *Indian Journal of Microbiology, 35*, 25-30.

Sadasivam, K., Tyagi, R., & Ramarethinam, S. (1986). Evaluation of some agricultural wastes as carriers for bacterial inoculants. *Agricultural Wastes, 17*, 301-306. http://dx.doi.org/10.1016/0141-4607(86)90138-1

Saleh, E. A., Nassar, F. R., & Yassen, A. M. (2001). Survival of microorganisms of multi-strains inoculant on different carriers. *Annals of Agricultural Science, 39*, 2163-2169.

Slonnezewski, J. L., & Foster, J. W. (2010). *Microbiology: An Evolving Science* (2th ed.). New York, USA: W. W. Norton & Company Inc.

Soil Science Society of America. (1996). *Methods of Soil Analysis.* (Book series no. 5). Madison, Wisconsin, USA.

Somasegaran, P. (1985). Inoculant production with diluted liquid cultures of *Rhizobium* spp. and autoclaved peat: Evaluation of diluents, *Rhizobium* spp., peats, sterility requirements, storage, and plant effectiveness. *Applied and Environmental Microbiology, 50*, 398.

Tate, R. L. (2000). *Soil Microbiology*. USA: John Wiley & Son, Inc.

Thungtrakul, M. (1987). Factors affecting survival of rhizobia in carriers. *Thesis*. Kasetsart University, Thailand.

Thananusont, V. (1993). Efficiency of rhizobium in peat carrier and modified carrier In *Workshop on Research Activities of ADRC Contributed to Agricultural Development in Northeast Thailand*, p.13. Khon Kaen, Thailand, Sep. 1-3.

Wangaruro, S., Karanja, N., Makatiani, B., Odee, D., & Woomer, P. (2000). Physico-chemical properties, initial microbial population and survival of rhizobia in peat, vermiculite and filtermud. In *Proceeding of the 15th annual general meeting. soil technologies for sustainable small holder farming systems in East Africa*, Soil Science Society of East Africa, Sportman's Arms Hotel, Nyayuki, Kenya. Aug. 19-23, 404.

Youssef, H., Sedik, M. Z., Fayez, M., & Hegazi, N. A. (1997). Response of some Gramineae to peat-based inoculum of associative diazotrophs. *Bulletin of Faculty of Agriculture, Cairo University, 48*, 159-174.

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# On Solving Linear Fractional Programming Problems

P. Pandian[1] & M. Jayalakshmi[1]

[1] Department of Mathematics, School of Advanced Sciences, VIT University, Vellore, India

Correspondence: P. Pandian, Department of Mathematics, School of Advanced Sciences, VIT University, Vellore 632014, India. E-mail: pandian61@rediffmail.com

Received: March 14, 2013 Accepted: April 18, 2013 Online Published: May 24, 2013

doi:10.5539/mas.v7n6p90 URL: http://dx.doi.org/10.5539/mas.v7n6p90

**Abstract**

A new method namely, denominator objective restriction method based on simplex method is proposed for solving linear fractional programming problems. Further, another method namely, decomposition-restriction method based on decomposition principle and the denominator objective restriction method is proposed for obtaining an optimal fuzzy solution to the fully fuzzy linear fractional programming problem. The procedures for the proposed methods are illustrated with the numerical examples.

**Keywords:** linear fractional programming, optimal solution, denominator objective restriction method, fuzzy numbers, fully fuzzy linear fractional programming

## 1. Introduction

Linear fractional programming (LFP) problems are a special type of non-linear programming problems in which the objective function is a ratio of linear functions and the constraints are linear functions. In real life situations, linear fractional models arise in decision making such as construction planning, economic and commercial planning, health care and hospital planning. In the literature, several methods (Bajalinov, 2003; Stancu-Minasian, 1997, 2006) have been recommended to solve LFP problems. Isbell and Marlow (1956) first identified an example of LFP problem and solved it by a sequence of linear programming problems. Charnes and Cooper (1962) considered variable transformation method to solve LFP and the updated objective function method were developed for solving the LFP problem by Bitran and Novaes (1973). Gilmore and Gomory (1963), Martos (1964), Swarup (1965), Wagner and Yuan (1968), Pandey and Punnen (2007) and Sharma et al. (1980) solved the LFP problem by various types of solution procedures based on the simplex method developed by Dantzig (1962). Tantawy (2007, 2008) proposed two different approaches namely; a feasible direction approach and a duality approach to solve the LFP problem. Mojtaba Borza et al. (2012) solved the LFP problem with interval coefficients in objective function which is based on Charnes and Cooper technique (1962). Odior (2012) solved the LFP problem by algebraic approach which depends on the duality concept and the partial fractions.

In real life model, the possible values of coefficients of a linear programming problem are obviously unclear and vague. In fuzzy decision making problems, the idea of maximizing decision was anticipated by Bellman and Zadeh (1970). The theory of fuzzy linear programming on general level was initially proposed by Tanaka et al. (1973). Buckley and Feuring (2000) measured the fully fuzzified linear programming problem (FFLP) by transformation to multiple objective deterministic non-linear programming problems. Li and Chen (1996) solved a fuzzy linear fractional programming form by fuzzy coefficients, using the concept as well as mathematical definition of the fuzzy optimal. Jayalakshmi and Pandian (2012) have proposed a method namely, bound and decomposition method to a fully fuzzy linear programming (FFLP) problem to obtain an optimal fuzzy solution. Hashemi et al. (2006) planned a two-phase approach based on the evaluation of mean plus standard deviation of fuzzy numbers to get the optimal solutions of the FFLP problem. Mikaeilvand et al. (2008) projected a method to solve FFLP through defuzzifying with a linear ranking function. Pop and Stancu Minasian (2008) and Bogdana Stanojevi'ca and Stancu-Minasianb (2012) used deterministic multiple objective linear programming problem by quadratic constraints to work out FFLP problems. Nachammai et al. (2012) considered FFLFP problem by using ranking method based on metric distance.

In this paper, we propose a new method namely, denominator objective restriction method for finding an optimal solution to LFP problems. In this proposed method, we construct two linear programming problems from the given LFP problem such that one is of maximization type and the other is of minimization type. Then, we attain

an optimal solution to the given LFP problem from the solutions of the two constructed linear programming problems. The proposed method is based only on the simplex method which differs totally from transformation method introduced by Charnes and Copper (1962) and the fractional simplex method introduced by Swarup (1965). Further, based on the decomposition principle and the denominator objective restriction method, we develop a new method namely, decomposition-restriction method to the FFLFP problem. In the decomposition-restriction method, the fuzzy ranking function, the transformation technique and multi-objective non-linear programming technique are not used. Numerical examples are given for better understanding the solution procedures of the proposed methods.

## 2. Preliminaries

We require the following definitions of the basic arithmetic operators and partial ordering relations on fuzzy triangular numbers based on the function principle which can be established in Bellman and Zadeh (1970), Jayalakshmi and Pandian (2012) and are used in section 4.

**Definition 2.1** A fuzzy number $\tilde{a}$ is a triangular fuzzy number denoted by $(a_1, a_2, a_3)$ where $a_1, a_2$ *and* $a_3$ are real numbers and its membership function $\mu_{\tilde{a}}(x)$ is given below:

$$\mu_{\tilde{a}}(x) = \begin{cases} (x-a_1)/(a_2-a_1) & for\ a_1 \le x \le a_2 \\ (a_3-x)/(a_3-a_2) & for\ a_2 \le x \le a_3 \\ 0 & otherwise \end{cases}$$

**Definition 2.2** Let $(a_1, a_2, a_3)$ and $(b_1, b_2, b_3)$ be two triangular fuzzy numbers. Then,

(i) $(a_1, a_2, a_3) \oplus (b_1, b_2, b_3) = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$ .

(ii) $(a_1, a_2, a_3) \Theta (b_1, b_2, b_3) = (a_1 - b_3, a_2 - b_2, a_3 - b_1)$ .

(iii) $k(a_1, a_2, a_3) = (ka_1, ka_2, ka_3),\ for\ k \ge 0$.

(iv) $k(a_1, a_2, a_3) = (ka_3, ka_2, ka_1),\ for\ k < 0$ .

(v) $$(a_1, a_2, a_3) \otimes (b_1, b_2, b_3) = \begin{cases} (a_1b_1, a_2b_2, a_3b_3), & a_1 \ge 0, \\ (a_1b_3, a_2b_2, a_3b_3), & a_1 < 0, a_3 \ge 0, \\ (a_1b_3, a_2b_2, a_3b_1), & a_3 < 0. \end{cases}$$

(vi) $$If\ 0 \notin (b_1, b_2, b_3), \frac{(a_1, a_2, a_3)}{(b_1, b_2, b_3)} = \left( \frac{a_1}{b_3}, \frac{a_2}{b_2}, \frac{a_3}{b_1} \right).$$

Let $F(R)$ be the set of all real triangular fuzzy numbers.

**Definition 2.3** Let $\tilde{A} = (a_1, a_2, a_3)$ and $\tilde{B} = (b_1, b_2, b_3)$ be in $F(R)$, then

(i) $\tilde{A} \approx \tilde{B}$ if $a_i = b_i, i = 1,2,3$ ; (ii) $\tilde{A} \preceq \tilde{B}$ if $a_i \le b_i, i = 1,2,3$

(iii) $\tilde{A} \succeq \tilde{B}$ if $a_i \ge b_i, i = 1,2,3$ and $\tilde{A} \succeq \tilde{0}$ if $a_i \ge 0, i = 1,2,3$ .

## 3. Linear Fractional Programming Problems

Consider the following LFP problem:

(P) Maximize $Z = \dfrac{C^T X + \alpha}{D^T X + \beta}$, subject to $AX \le B,\ X \ge 0$

where *X, C, D* are $n \times 1$ vectors, *B* is an $m \times 1$ vector and $\alpha, \beta$ are scalars.

It is assumed that the set of feasible solutions to the problem (P), $S = \{X \in R^n : AX \le b, X \ge 0\}$ is non-empty and bounded.

Now, we can construct two single objective linear programming problem from the given problem (P) as follows:

(N) Maximize $P(X) = C^T X + \alpha$, subject to $AX \le B,\ X \ge 0$

and

(D) Minimize $Q(X) = D^T X + \beta$, subject to $AX \le B,\ X \ge 0$

Now, we prove the following two theorems connecting the solutions of the problem (P), the problem (N) and the problem (D) which are used in the proposed method.

**Theorem 3.1** Let $X_\circ$ be an optimal solution to the problem (N). If $\{X_n\}$ is a sequence of basic feasible solutions to the problem (D) such that $Z(X_k) \le Z(X_{k+1})$ for all $k = 0,1,2,...,n-1$ and $Z(X_n) \ge Z(X_{n+1})$ by simplex method considering the solution $X_\circ$ as an initial feasible solution, then $X_n$ is an optimal solution to the problem (P).

**Proof:** It is obvious that $X_n$ is a feasible solution to the problem (P)

Let U be a feasible solution to the problem (P).

This implies $Q(U) \le Q(X_n)$ or $Q(U) > Q(X_n)$.

**Case 1:** $Q(U) \le Q(X_n)$

Since $Z(X_k) \le Z(X_{k+1})$ for all $k = 0,1,2,...,n-1$ and $Z(X_n) \ge Z(X_{n+1})$ and the problem (D) is of minimization type, we have $Z(X_n) \ge Z(U)$. Therefore, $X_n$ is an optimal solution to the problem (P).

**Case 2:** $Q(U) > Q(X_n)$. That is, $Q(X_n) < Q(U)$

Since $Z(X_k) \le Z(X_{k+1})$ for all $k = 0,1,2,...,n-1$ and $Z(X_n) \ge Z(X_{n+1})$ and also, the problem (D) is of minimization type, we can conclude that $Z(X_n) \ge Z(U)$. Therefore, $X_n$ is an optimal solution to the problem (P).

Thus, $X_n$ is an optimal solution to the problem (P).

Thus, the theorem is proved.

**Theorem 3.2** Let $X_\circ$ be an optimal solution to the problem (N). If $\{X_n\}$ is a sequence of basic feasible solutions to the problem (D) such that $Z(X_k) \le Z(X_{k+1})$ for all $k = 0,1,2,...,n$ and $X_{n+1}$ is an optimal solution to the problem (D) by simplex method considering the solution $X_\circ$ as an initial feasible solution, then $X_{n+1}$ is an optimal solution to the problem (P).

**Proof:** It is observed that $X_{n+1}$ is a feasible solution to the problem (P).

Let V be a feasible solution to the problem (P).

Now, since $X_{n+1}$ is an optimal solution to the problem (D), we have $Q(V) \ge Q(X_{n+1})$.

Now, since $Z(X_k) \le Z(X_{k+1})$ for all $k = 0,1,2,...,n$ and $X_{n+1}$ is an optimal solution to the problem (D), we can conclude that $Z(X_{n+1}) \ge Z(V)$. Therefore, $X_{n+1}$ is an optimal solution to the problem (P).

Hence, the theorem is proved.

Now, we introduce a new method namely, denominator objective restriction method for finding an optimal solution to the LFP problem (P).

The proposed method proceeds as follows:

**Step 1:** Construct two single objective linear programming problems namely, the problem (N) as well as the problem (D) from the given problem (P).

**Step 2:** Compute the optimal solution to the problem (N) by means of the simplex method. Let the optimal solution to the problem (N) be $X_\circ$ and Max. $Z(X_\circ) = Z_\circ$

**Step 3:** Using the optimal table of the problem (N) as an initial simplex table to the problem (D), continue to find a sequence of improved basic feasible solutions $\{X_n\}$ to the problem (D) and the value of Z at each of the improved basic feasible solution by the simplex method.

**Step 4:** (a) If $Z(X_k) \le Z(X_{k+1})$ for all $k = 0,1,2,...,n-1$ and $Z(X_n) \ge Z(X_{n+1})$ for some n, stop the computation process and then, go to Step 5.

**Step 4:** (b) If $Z(X_k) \le Z(X_{k+1})$ for all $k = 0,1,2,...,n$ and $X_{n+1}$ is an optimal solution to the problem (D) for some *n*, stop the computation process and then, go to Step 6.

**Step 5:** $X_n$ is an optimal solution to the problem (P) and Max. $Z(X) = Z(X_n)$ by the Theorem 3.1.

**Step 6:** $X_{n+1}$ is an optimal solution to the problem (P) and Max. $Z(X) = Z(X_{n+1})$ by the Theorem 3.2.

**Remark 3.1** The maximum value for (*n* +1) is the number of the iterations to get an optimal solution to the problem (D) using simplex method.

The proposed method for solving the LFP problem is illustrated through the following examples.

**Example 3.1** Consider the following LFP problem:

$$\text{Maximize } Z = \frac{5x_1 + 6x_2}{2x_2 + 7}$$

subject to $2x_1 + 3x_2 \le 6$ ; $2x_1 + x_2 \le 3$ ; $x_1, x_2 \ge 0$ .

The following two LP problems can be obtained from the given problem:

(N) Maximize $P(X) = 5x_1 + 6x_2$

subject to $2x_1 + 3x_2 \le 6$ ; $2x_1 + x_2 \le 3$ ; $x_1, x_2 \ge 0$

and

(D) Minimize $Q(X) = 2x_2 + 7$

subject to $2x_1 + 3x_2 \le 6$ ; $2x_1 + x_2 \le 3$ ; $x_1, x_2 \ge 0$ .

Now, the optimal solution to the problem (N), by the simplex method, is given by the following table:

| | $C$ | 5 | 6 | 0 | 0 | | |
|---|---|---|---|---|---|---|---|
| $C_B$ | $X_B$ | $x_1$ | $x_2$ | $s1$ | $s2$ | Solution. | Ratio |
| 6 | $x_2$ | 0 | 1 | $\frac{1}{2}$ | $-\frac{1}{2}$ | $\frac{3}{2}$ | |
| 5 | $x_1$ | 1 | 0 | $-\frac{1}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | |
| $P_j - C_j$ | | 0 | 0 | $\frac{7}{4}$ | $\frac{3}{4}$ | $P = \frac{51}{4}$ | $Z = \frac{51}{40}$ |

Therefore, the optimal solution to the problem (N) is $x_1 = \frac{3}{4}$, $x_2 = \frac{3}{2}$ , max. $P(X) = \frac{51}{4}$ and the value of $Z = \frac{51}{40}$ .

Now, by Step 3 of the proposed method, the initial simplex table to the problem (D) is given below:

| | $D$ | 0 | 2 | 0 | 0 | | |
|---|---|---|---|---|---|---|---|
| $D_B$ | $X_B$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | Solution. | Ratio |
| 2 | $x_2$ | 0 | 1 | $\frac{1}{2}$ | $-\frac{1}{2}$ | $\frac{3}{2}$ | - |
| 0 | $x_1$ | 1 | 0 | $-\frac{1}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | 1 |
| $D_j - Q_j$ | | 0 | 0 | -1 | 1 | $Q = 10$ | $Z_o = \frac{51}{40}$ |

Now, the variable $s_2$ enters into the basis and the variable $x_1$ leaves from the basis.

**Ist iteration table:**

| | $D$ | 0 | 2 | 0 | 0 | | |
|---|---|---|---|---|---|---|---|
| $D_B$ | $X_B$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | Solution. | Ratio |
| 0 | $s_1$ | 0 | 2 | 1 | -1 | 3 | |
| 0 | $x_1$ | 1 | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ | $\frac{3}{2}$ | |
| $D_j - Q_j$ | | 0 | 0 | 0 | 0 | $Q = 7$ | $Z_1 = \frac{15}{14}$ |

Since $Z_o > Z_1$ and by the Step 4(a) of the proposed method, the optimal solution to the given linear fractional programming problem is $x_1 = \frac{3}{4}$, $x_2 = \frac{3}{2}$ and Max. $Z = \frac{51}{40}$.

**Example 3.2** Consider the following LFP problem:

Maximize $Z = \frac{x_1 + 2x_2}{2x_1 - x_2 + 2}$

subject to $-x_1 + 2x_2 \le 2$; $x_1 + x_2 \le 4$; $x_1, x_2 \ge 0$.

The following two LP problems can be obtained from the given problem:

(N) Maximize $P(X) = x_1 + 2x_2$

subject to $-x_1 + 2x_2 \le 2$; $x_1 + x_2 \le 4$; $x_1, x_2 \ge 0$.

and

(D) Minimize $Q(X) = 2x_1 - x_2 + 2$

subject to $-x_1 + 2x_2 \le 2$; $x_1 + x_2 \le 4$; $x_1, x_2 \ge 0$.

Now, by the simplex method, the optimal solution to the problem (N) is given in the following table:

| | $C$ | 1 | 2 | 0 | 0 | | |
|---|---|---|---|---|---|---|---|
| $C_B$ | $X_B$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | Solution. | Ratio |
| 2 | $x_2$ | 0 | 1 | $\frac{1}{3}$ | $\frac{1}{3}$ | 2 | |
| 1 | $x_1$ | 1 | 0 | $-\frac{1}{3}$ | $\frac{2}{3}$ | 2 | |
| $P_j - C_j$ | | 0 | 0 | $\frac{1}{3}$ | $\frac{4}{3}$ | $P = 6$ | $Z = \frac{3}{2}$ |

Therefore, the optimal solution to the problem (N) is $x_1 = 2$, $x_2 = 2$, Max. $P(X) = 6$ and the value of $Z = \frac{3}{2}$.

Now, by Step 3 of the proposed method, the initial simplex table to the problem (D) is given below:

| | $D$ | 2 | -1 | 0 | 0 | | |
|---|---|---|---|---|---|---|---|
| $D_B$ | $X_B$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | Solution. | Ratio |
| -1 | $x_2$ | 0 | 1 | $\frac{1}{3}$ | $\frac{1}{3}$ | 2 | 6 |
| 2 | $x_1$ | 1 | 0 | $-\frac{1}{3}$ | $\frac{2}{3}$ | 2 | 3 |
| $D_j - Q_j$ | | 0 | 0 | 1 | -1 | $Q = 4$ | $Z_o = \frac{3}{2}$ |

Now, the variable $s_2$ enters into the basis and the variable $x_1$ leaves from the basis.

**Ist iteration table:**

| | $D$ | 2 | -1 | 0 | 0 | | |
|---|---|---|---|---|---|---|---|
| $D_B$ | $X_B$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | Solution. | Ratio |
| -1 | $x_2$ | $-\frac{1}{2}$ | 1 | $\frac{1}{2}$ | 0 | 1 | |
| 0 | $s_2$ | $\frac{3}{2}$ | 0 | $-\frac{1}{2}$ | 1 | 3 | |
| $D_j - Q_j$ | | 0 | 0 | 0 | 0 | $Q = 1$ | $Z_1 = 2$ |

Since the I $^{st}$ iteration table is optimal and by Step 4(b) of the proposed method, the optimal solution to the given linear fractional programming problem is $x_1 = 0$, $x_2 = 1$ and Max. $Z = 2$.

## 4. Fully Fuzzy Linear Fractional Programming Problem

Consider the following FFLFP problems having $m$ fuzzy constraints and $n$ fuzzy variables:

$$\text{(FP) Maximize } \tilde{z} \approx \frac{\tilde{c}^T \tilde{x} \oplus \tilde{\alpha}}{\tilde{d}^T \tilde{x} \oplus \tilde{\beta}}, \text{ subject to } \tilde{A} \otimes \tilde{x} \{\precsim, \approx, \succsim\} \tilde{b}, \ \tilde{x} \succsim \tilde{0}$$

where $\tilde{c}^T = \left(\tilde{c}_j\right)_{1\times n}$, $\tilde{d}^T = \left(\tilde{d}_j\right)_{1\times n}$, $\tilde{A} = \left(\tilde{a}_{ij}\right)_{m\times n}$, $\tilde{x} = (\tilde{x}_j)_{n\times 1}$, $\tilde{b} = (\tilde{b}_i)_{m\times 1}$, $\tilde{\alpha}, \tilde{\beta} \in F(R)$

and $\tilde{a}_{ij}, \tilde{c}_j, \tilde{d}_j, \tilde{x}_j, \tilde{b}_i \in F(R)$, for all $1 \le j \le n$ and $1 \le i \le m$.

Let the parameters $\tilde{\alpha}, \tilde{\beta}$, $\tilde{c}_j, \tilde{d}_j, \tilde{x}_j, \tilde{a}_{ij}$ and $\tilde{b}_i$ be the triangular fuzzy numbers $(\alpha_1, \alpha_2, \alpha_3)$, $(\beta_1, \beta_2, \beta_3)$, $\left(p_j, q_j, r_j\right), \left(m_j, n_j, l_j\right), \left(x_j, y_j, t_j\right)$, $(a_{ij}, b_{ij}, c_{ij})$ and $(b_i, g_i, h_i)$ respectively. Then, the problem (FP) can be written as follows:

$$\text{(FP) Maximize } (z_1, z_2, z_3) \approx \frac{\left(\sum_{j=1}^{n} \left(p_j, q_j, r_j\right) \otimes \left(x_j, y_j, t_j\right)\right) \oplus (\alpha_1, \alpha_2, \alpha_3)}{\left(\sum_{j=1}^{n} \left(m_j, n_j, l_j\right) \otimes (x_j, y_j, t_j)\right) \oplus (\beta_1, \beta_2, \beta_3)}$$

subject to $\sum_{j=1}^{n} \left(a_{ij}, b_{ij}, c_{ij}\right) \otimes \left(x_j, y_j, t_j\right) \{\precsim, \approx, \succsim\} \left(b_i, g_i, h_i\right)$, for all $i = 1,2,...,m$, $\left(x_j, y_j, t_j\right) \succsim \tilde{0}$.

Now, since $\left(x_j, y_j, t_j\right)$ is a triangular fuzzy number, then

$$x_j \le y_j \le t_j, \text{ j } = 1,2,\ldots,\text{m.} \tag{4.1}$$

The relation (4.1) is called bounded constraints.

Now, using the arithmetic operations and partial ordering relations, we decompose the given FFLFP problem as follows:

$$\text{Max. } z_1 = \text{Lower value of } \frac{\left(\sum_{j=1}^{n} \left(p_j, q_j, r_j\right) \otimes \left(x_j, y_j, t_j\right)\right) \oplus (\alpha_1, \alpha_2, \alpha_3)}{\left(\sum_{j=1}^{n} \left(m_j, n_j, l_j\right) \otimes (x_j, y_j, t_j)\right) \oplus (\beta_1, \beta_2, \beta_3)}$$

$$\text{Max. } z_2 = \text{Middle value of } \frac{\left(\sum_{j=1}^{n} \left(p_j, q_j, r_j\right) \otimes \left(x_j, y_j, t_j\right)\right) \oplus (\alpha_1, \alpha_2, \alpha_3)}{\left(\sum_{j=1}^{n} \left(m_j, n_j, l_j\right) \otimes (x_j, y_j, t_j)\right) \oplus (\beta_1, \beta_2, \beta_3)}$$

$$\text{Max.} \quad z_3 = \text{Upper value of} \ \frac{\left(\sum_{j=1}^{n}\left(p_j,q_j,r_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\alpha_1,\alpha_2,\alpha_3\right)}{\left(\sum_{j=1}^{n}\left(m_j,n_j,l_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\beta_1,\beta_2,\beta_3\right)}$$

subject to

$$\sum_{j=1}^{n}\text{lower value of}\left((a_{ij},b_{ij},c_{ij})\otimes(x_j,y_j,t_j)\right)\{\leq,=,\geq\}b_i, \quad \text{for all} \quad i=1,2,...,m;$$

$$\sum_{j=1}^{n}\text{middle value of}\left((a_{ij},b_{ij},c_{ij})\otimes(x_j,y_j,t_j)\right)\{\leq,=,\geq\}g_i, \quad \text{for all} \quad i=1,2,...,m;$$

$$\sum_{j=1}^{n}\text{upper value of}\left((a_{ij},b_{ij},c_{ij})\otimes(x_j,y_j,t_j)\right)\{\leq,=,\geq\}h_i, \quad \text{for all} \quad i=1,2,...,m$$

and all decision variables are non-negative.

From the above decomposition problem, we construct the following three crisp LFP problems namely, middle level problem $(MLP)$, upper level problem $(ULP)$ and lower level problem $(LLP)$ as follows:

$$(MLP)\ \text{Max.}\ z_2 = \text{Middle value of} \ \frac{\left(\sum_{j=1}^{n}\left(p_j,q_j,r_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\alpha_1,\alpha_2,\alpha_3\right)}{\left(\sum_{j=1}^{n}\left(m_j,n_j,l_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\beta_1,\beta_2,\beta_3\right)}$$

subject to constraints in the decomposition problem in which at least one decision variable of the $(MLP)$ occurs and all decision variables are non-negative.

$$(ULP)\ \text{Max.} \quad z_3 = \text{Upper value of} \ \frac{\left(\sum_{j=1}^{n}\left(p_j,q_j,r_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\alpha_1,\alpha_2,\alpha_3\right)}{\left(\sum_{j=1}^{n}\left(m_j,n_j,l_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\beta_1,\beta_2,\beta_3\right)}$$

subject to $z_3 \geq z_2^{\circ}$;

constraints in the decomposition problem in which at least one decision variable of the $(ULP)$ occurs and are not used in $(MLP)$;

all variables in the constraints and objective function in $(ULP)$ must satisfy the bounded constraints;

replacing all values of the decision variables which are obtained in $(MLP)$ and all decision variables are non-negative.

And

$$(LLP)\ \text{Max.} \quad z_1 = \text{Lower value of} \ \frac{\left(\sum_{j=1}^{n}\left(p_j,q_j,r_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\alpha_1,\alpha_2,\alpha_3\right)}{\left(\sum_{j=1}^{n}\left(m_j,n_j,l_j\right)\otimes\left(x_j,y_j,t_j\right)\right)\oplus\left(\beta_1,\beta_2,\beta_3\right)}$$

subject to $z_1 \leq z_2^{\circ}$;

constraints in the decomposition constraints in which at least one decision variable of the $(LLP)$ occurs which are not used in $(MLP)$ and $(ULP)$;

all variables in the constraints and objective function in $(LLP)$ must satisfy the bounded constraints;

replacing all values of the decision variables which are obtained in the $(MLP)$ and $(ULP)$ and all decision variables are non-negative, where $z_2^{\circ}$ is the optimal objective value of $(MLP)$.

Now, we prove the following theorem which is used in the proposed method to solve the FFLFP problem.

**Theorem 4.1** Let $[x_M^{\circ}] = \{x_j^{\circ}, x_j^{\circ} \in M\}$ be an optimal solution of $(MLP)$, $[x_U^{\circ}] = \{x_j^{\circ}, x_j^{\circ} \in U\}$ be an optimal solution of $(ULP)$ and $[x_L^{\circ}] = \{x_j^{\circ}, x_j^{\circ} \in L\}$ be an optimal solution of $(LLP)$ where $L$, $M$ and $U$ are sets of

decision variables in the $(LLP),(MLP)$ and $(ULP)$ respectively. Then $\left\{\tilde{x}_j^\circ = (x_j^1, x_j^2, x_j^3),\ j = 1,2,...,n\right\}$ is an optimal fuzzy solution to the given problem (FP) where each one of $x_j^{\circ 1}, x_j^{\circ 2}$ and $x_j^{\circ 3},\ j = 1,2,...,n$ is an element of $L$, $M$ and $U$.

**Proof:** Let $[\tilde{y}_j] = \{\tilde{y}_j, j = 1,2,...,n\}$ be a feasible solution of (FP). Clearly, $[y_M], [y_U]$ and $[y_L]$ are feasible solutions of $(MLP),(ULP)$ and $(LLP)$ respectively.

Now, since $[x_M^\circ], [x_U^\circ]$ and $[x_L^\circ]$ are optimal solutions of $(MLP),(ULP)$ and $(LLP)$ respectively, we have $Z_1([x_L^\circ]) \geq Z_1([y_L]); Z_2([x_M^\circ]) \geq Z_2([y_M])$ and $Z_3([x_U^\circ]) \geq Z_3([y_U])$.

This implies that $Z([\tilde{x}_j^\circ]) \geq Z([\tilde{y}_j])$, for all feasible solution of the problem (P).

Therefore, $\{\tilde{x}_j^\circ \approx (x_j^{\circ 1}, x_j^{\circ 2}, x_j^{\circ 3}), j = 1,2,...,n\}$ is an optimal fuzzy solution to the given problem (FP) where each one of $x_j^{\circ 1}, x_j^{\circ 2}$ and $x_j^{\circ 3},\ j = 1,2,...,n$ is an element of $L$, $M$ and $U$.

Hence the theorem is proved.

**Remark 4.1** In the case of LFP problem involving trapezoidal fuzzy numbers and variables, we decompose it into four crisp LFP problems and then, we solve the middle level problems (second and third problems) first. Then, we solve the upper level and lower level problems and then, we obtain an optimal fuzzy solution to the given LFP problem involving trapezoidal fuzzy numbers and variables.

**Remark 4.2** In the case of LFP problem involving interval fuzzy numbers and/or interval variables, we decompose it into two crisp LP problems and then, we solve the upper level problem first. Then, we solve the lower level problem and then, we attain the interval optimal solution to the given LFP problem involving interval numbers and / or interval variables.

Now, we introduce a new method namely, decomposition-restriction method to the FFLFP problem (FP) for obtaining an optimal fuzzy solution, based on the decomposition principle and the denominator objective restriction method.

The proposed method proceeds as follows:

**Step 1:** Construct three crisp LFP problems namely Middle level problem, Upper level problem and Lower level problem from the given FLFP problem.

**Step 2:** Solve the Middle level problem by the denominator objective restriction method.

**Step 3:** Using the results of Step 2 and the denominator objective restriction method, solve the Upper level problem.

**Step 4:** Using the results of Step 2, Step 3 and the denominator objective restriction method, solve the Lower level problem.

**Step 5:** Using the results of Step 2, Step 3 and Step 4, obtain an optimal fuzzy solution to the given FFLFP problem by the Theorem 4.1.

The proposed method is illustrated with the following example.

**Example 4.1** Consider the following FFLFP problem:

$$\text{Maximize}\ \ \tilde{Z} = \frac{(0,1,2)\,\tilde{x}_1 \oplus (-2,-1,0)\,\tilde{x}_2 \oplus (0,1,2)}{(0,1,2)\,\tilde{x}_1 \oplus (0,1,2)\,\tilde{x}_2 \oplus (1,2,3)}$$

subject to $(0,1,2)\,\tilde{x}_1 + (0,1,2)\,\tilde{x}_2 \preceq (1,2,3)$, $(0,1,2)\,\tilde{x}_1 + (-2,-1,0)\,\tilde{x}_2 \preceq (0,1,2)$; $\tilde{x}_1, \tilde{x}_2 \geq 0$.

Let $\tilde{x}_1 = (x_1, y_1, t_1)$, $\tilde{x}_2 = (x_2, y_2, t_2)$ and $\tilde{z} = (z_1, z_2, z_3)$.

Now, the decomposition problems of the given FFLFP problem are given below:

$$\text{Maximize}\ \ z_1 = \frac{-2t_2}{2t_1 + 2t_2 + 3}$$

$$\text{Maximize}\ \ z_2 = \frac{y_1 - y_2 + 1}{y_1 + y_2 + 2}$$

$$\text{Maximize}\ \ z_3 = 2t_1 + 2$$

subject to $0x_1 + 0x_2 \le 1; -2t_2 \le 0;\ y_1 + y_2 \le 2; y_1 - y_2 \le 1;\ 2t_1 + 2t_2 \le 3; 2t_1 \le 2;\ x_1, x_2 \ge 0$ , $y_1, y_2 \ge 0$ , $t_1, t_2 \ge 0$ .

Now, the Middle Level problem is given below:

$$\text{(MLP) Maximize } z_2 = \frac{y_1 - y_2 + 1}{y_1 + y_2 + 2}$$

subject to $y_1 + 2y_2 \le 2; y_1 - y_2 \le 1 ; y_1, y_2 \ge 0$ .

Now, solving the problem (MLP) by the denominator objective restriction method, we attain an optimal solution $y_1 = 1; y_2 = 0$ and $z_2 = \frac{2}{3}$ .

Now, the Upper Level problem is given below:

$$\text{(ULP) Maximize } z_3 = 2t_1 + 2$$

subject to $2t_1 + 2 \ge \frac{2}{3}; 2t_1 + 2t_2 \le 3; t_1 \le 1;\ t_1 \ge y_1; t_2 \ge y_2 ; t_1, t_2 \ge 0$ .

Now, solving the problem (ULP) with $y_1 = 1; y_2 = 0$ using the denominator objective restriction method, he optimal solution to the problem (ULP) is $t_1 = 1; t_2 = 0$ and $z_3 = 4$ .

Now, the Lower Level problem:

$$\text{(LLP) Maximize } z_1 = \frac{-2t_2}{2t_1 + 2t_2 + 3}$$

subject to $\frac{-2t_2}{2t_1 + 2t_2 + 3} \le \frac{2}{3}; 2t_1 + 2t_2 \le 3; t_1 \le 1;\ t_1 \ge y_1; t_2 \ge y_2 ; t_1, t_2 \ge 0$ .

Now, substituting $t_1 = 1$ and $t_2 = 0$ in the problem (LLP), the optimal solution is $t_1 = 1, t_2 = 0$ and $z_1 = 0$ .

Now, since $x_1 \le y_1$ , $x_2 \le y_2$ and $x_1, x_2 \ge 0$ , we can conclude that $x_2 = 0$ and $x_1 = x_0$ where $x_0 \in [0,1]$ .

Therefore, an optimal fuzzy solution to the given FFLFP problem is $\tilde{x}_1 \approx (x_0, 1, 1)$ , $\tilde{x}_2 \approx (0, 0, 0)$ and Max. $\tilde{z} \approx \left(0, \frac{2}{3}, 4\right)$ where $x_0 \in [0,1]$ .

**Remark 4.3** In Bogdana Stanojevi'ca and Stancu-Minasianb (2012), the optimal fuzzy solution to the FFLFP problem (Example 4.1.) is $\tilde{x}_1 \approx (0, 1, 1)$ , $\tilde{x}_2 \approx (0, 0, 0)$ and the maximum value of $\tilde{z} \approx (0, 0.55, 1.09)$ by a new method which is based on Charnes-Copper method and multiobjective nonlinear programming, but by the decomposition-restriction method, the optimal fuzzy solution to the same FFLFP problem is $\tilde{x}_1 \approx (x_0, 1, 1)$ , $\tilde{x}_2 \approx (0, 0, 0)$ and the maximum value $\tilde{z} \approx \left(0, \frac{2}{3}, 4\right)$ where $x_0 \in [0,1]$ .

## 5. Conclusion

In this paper, the denominator objective restriction method is developed for solving LFP problems based only on simplex method in. It is easy to understand, compute and also, to interpret. In the second part of the proposed method, that is, solving the problem (D), each iteration solution is a next level accepted solution to the LFP problem that may be used by decision makers according to their situations if they need. Further, the decomposition-restriction method is used to solve the FFLFP problem in which the fuzzy ranking function, transformation technique and multi-objective non-linear programming technique are not used, but it is based on the decomposition principle and the denominator objective restriction method. Both the methods can serve decision makers by providing an appropriate best solution to a variety of linear fractional programming models having crisp or fuzzy parameters and variables in a simple and effective manner.

**Acknowledgements**

We owe a lot to Dr. G. Natarajan and Prof. M.A. Mohamed Sahul Hameed, VIT University, Vellore in reviewing our paper which helped us to complete this work.

## References

Bajalinov, E. B. (2003). *Linear-fractional-Programming Theory, Methods, Applications and Software*. Boston: Kluwer Academic publishers.

Bellman, R. E., & Zadeh, L. A. (1970). Decision Making in a Fuzzy Environment. *Management Science, 17*, 141-164. http://dx.doi.org/10.1287/mnsc.17.4.B141

Bitran, G. R., & Novaes, A. J. (1973). Linear programming with a fractional objective function. *Journal of Operations Research, 21*, 22-29. http://dx.doi.org/10.1287/opre.21.1.22

Buckley, J. J., & Feuring, T. (2000). Evolutionary Algorithm solution to fuzzy problems: fuzzy linear programming. *Fuzzy Sets and Systems, 109*, 35-53. http://dx.doi.org/10.1016/S0165-0114(98)00022-0

Charnes, A., & Cooper, W. W. (1962). Programming with linear functional. *Naval Research Logistics Quarterly, 9*, 181-186. http://dx.doi.org/10.1002/nav.3800090303

Dantzig, G. B. (1962). Linear Programming and extension. *Princeton University Press, Princeton, New Jersey*.

Gilmore, P. C., & Gomory, R. E. (1963). Linear programming approach to the cutting stock problem-Part 2. *Operations Research, 11*, 863-867. http://dx.doi.org/10.1287/opre.11.6.863

Hashemi, S. M., Modarres, M., Nasrabadi, E., & Nasrabadi, M. M. (2006). Fully fuzzified linear programming, solution and duality. *Journal of Intelligent and Fuzzy Systems, 17*, 253-261.

Isbell, J. R., & Marlow, W. H. (1956). Attrition games. *Naval Research Logistics Quarterly, 3*, 1-99. http://dx.doi.org/10.1002/nav.3800030108

Jayalakshmi, M., & Pandian, P. (2012). A New Method for Finding an Optimal Fuzzy Solution Fully Fuzzy Linear Programming Problems. *International Journal of Engineering Research and Applications, 2*, 247-254.

Li, D. F., & Chen, S. (1996). A fuzzy programming approach to fuzzy linear fractional programming with fuzzy coefficients. *The Journal of Fuzzy Mathematics, 4*, 829-834.

Martos, B. (1964). Hyperbolic programming. *Naval Research Logistics Quarterly, 11*, 135-155. http://dx.doi.org/10.1002/nav.3800110204

Mikaeilvand, N., Allahviranloo, T., Hosseinzadeh, F., & Khorasany Kiasary, M. (2008). Fully fuzzy linear programming problem with positive or negative core. *Far East Journal of Applied Mathematics, 33*, 337-350.

Mojtaba, B., Azmin, S. R., & Mansour, S. (2012). Solving linear fractional programming problems with interval coefficients in the objective function-A new approach. *Applied Mathematical Sciences, 6*, 3442-3452.

Nachammai, Al., & Thangaraj, P. (2012). Solving fuzzy linear fractional programming problem using metric distance ranking. *Applied Mathematical Sciences, 6*, 1275-1285.

Odior, A. O. (2012). An approach for solving linear fractional programming problems. *International Journal of Engineering and Technology, 1*, 298-304.

Pandey, P., & Punnen, A. P. (2007). A simplex algorithm for piecewise-linear fractional programming problems. *European Journal of Operational Research, 178*, 343-358. http://dx.doi.org/10.1016/j.ejor.2006.02.021

Pop, B., & Stancu-Minasian, I. M. (2008). A method of solving fully fuzzified fractional programming problems. *J. Appl. Math. Comput., 27*, 227-242. http://dx.doi.org/10.1007/s12190-008-0052-5

Schaible, S. (1981). Fractional programming: Applications and Algorithms. *European Journal of Operational Research, 7*, 111-120. http://dx.doi.org/10.1016/0377-2217(81)90272-1

Sharma, J. K., Gupta, A. K., & Gupta, M. P. (1980). Extension of simplex technique for solving fractional programming problems. *Indian Journal of Pure and Applied Mathematics, 11*, 961-968.

Stancu-Minasian, I. M. (1997). Fractional programming: Theory, methods and applications. *Kluwer Academic Publishers, Dordrecht.* http://dx.doi.org/10.1007/978-94-009-0035-6

Stancu-Minasian, I. M. (2006). A sixth bibliography of fractional programming. *Optimization, 55*, 405-428. http://dx.doi.org/10.1080/02331930600819613

Stanojevi'c, B., & Stancu-Minasian, I. M. (2012). Evaluating fuzzy inequalities and solving fully fuzzified linear fractional program. *Yugoslav Journal of Operations Research, 22*, 41-50. http://dx.doi.org/10.2298/YJOR110522001S

Swarup, K. (1965). Linear fractional functional programming. *Operation Research, 13*, 1029-1036. http://dx.doi.org/10.1287/opre.13.6.1029

Tanaka, H., Okuda, T., & Asai, K. (1973). On Fuzzy Mathematical Programming. *Journal of Cybernetics and Systems, 3*, 37-46. http://dx.doi.org/10.1080/01969727308545912

Tantawy, S. F. (2007). Using feasible directions to solve linear fractional programming problems. *Australian Journal of Basic and Applied Sciences, 1*, 109-114.

Tantawy, S. F. (2008). A new procedure for solving linear fractional programming problems. *Mathematical and Computer Modeling, 48*, 969-973. http://dx.doi.org/10.1016/j.mcm.2007.12.007

Wagner, H. M., & Yuan, J. S. C. (1968). Algorithm equivalence in linear fractional programming. *Management Science, 14*, 301-306. http://dx.doi.org/10.1287/mnsc.14.5.301

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# Saddlepoint Method to Cumulative Distribution Function for Poisson-Binomial Model

Al Mutairi Alya O.[1,2] & Heng Chin Low[1]

[1] School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia

[2] Fuculty of Applied Science, Taibah University, AlMadinah-M., Kingdom of Saudi Arabia

Correspondence: Al Mutairi Alya O., Applied Statistics Department, Faculty of Applied Science, Taibah University, AlMadinah-M., Kingdom of Saudi Arabia. E-mail: afaaq99@hotmail.com

Received: April 9, 2013 Accepted: April 30, 2013 Online Published: May 24, 2013

doi:10.5539/mas.v7n6p101 URL: http://dx.doi.org/10.5539/mas.v7n6p101

*The study is supported financially by Taibah University in Al-Madinah, Kingdom of Saudi Arabia*

**Abstract**

The randm sum distribution plays an important key in statistical science as well as with insurance program, biotechnology and applied medical science. Saddlepoint methods are considered to be random sum variables with dependent elements supposing presence of the Moment Generation Function (MGF). Saddlepoint methods are influential instruments for getting precise terms for distribution functions in closed form. However, the paper also, discusses the Saddlepoint methods to the Cumulative Distribution Function (CDF) for Poisson-Binomial model in discrete form.

**Keywords:** random-sum distribution, poisson-binomial model, saddlepoint approximation, cumulative distribution function

## 1. Introduction

Saddlepoint method is influential way in having precise terms for distribution function which is not recognized in closed sitting. Saddlepoint approximation almost surpass other techniques regarding calculating expenses; while it does not inevitably surpass them concerning accuracy. The most basic Saddlepoint method was launched by Daniels (1954) and is fundamentally expression for approximating CDF and discrete distribution function through its MGF. Saddlepoint methods are constructed by supposing existence of the MGF or, equally, the CGF, of random variable. However, for improvement to the Saddlepoint methodology and associated techniques, as references, Skovagaard in 1987 proposed a conditional version of this approximation, Daniels (1954, 1987) reported details concerning density and mass approximation, in 1999 Borowaik discussed a tail-area approximation with a uniform relative error, Reid in 1998 indicated applications to inference, and Terrell (2003) proposed a stabilized Lugannani-Rice formula. However, this paper will use Saddlepoint method to estimate the random-sum based on MGF for Poisson-Binomial model. For all real values of $x$, random variable $X$ was presumed to have mass function $p(x)$ identified. Subsequently, the MGF is identified as

$$M(s) = E(e^{sx}) = \sum_{x=-\infty}^{\infty} e^{sx} p(x), \text{ (Hogg \& Craig, 1978)} \quad (1)$$

Over values of $s$ for which the integral converges and the convergence is constantly certain at $s = 0$ and it should be supposed that $M(s)$ coverage over largest open neighborhood zero as $(a,b)$. However, the CGF is known as

$$K(s) = \ln M(s), s \in (a,b), \text{ (Johnson et al., 2005)} \quad (2)$$

For discrete integral-valued random variable $X$, the saddlepointa approximation for its mass function $p(x)$, based on the CGF $K$ is given by

$$\hat{p}(x) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp(k(\hat{s}) - \hat{s}x) \quad (3)$$

where $\hat{s} = \hat{s}(x)$ denote the unique outcome to the Saddlepoint equation

$$K'(\hat{s}) = x\,,\ x \in \tau_x$$

$x \in \tau_x$ and $\tau_x$ is the inner part of the span that powered by $X$ (Butler, 2007). Saddlepoint expression (3) is calculable for any value in $\tau_x$, however, the plot of $\hat{p}(x)$ is significant as an approximation to $p(x)$ merely for integer-value.

Daniels (1987) initiated, two stability approved adjustments for discrete integral-valued random variable $X$ for univariate cumulative distribution functions (CDF) that are presented below.

*1.1 First Continuity-Modification*

Assume $x \in \tau_x$, in order that the saddlepoint equation could be solved at value $x$, the first approximation is

$$\hat{P}_{r_1}(X \geq x) = \begin{cases} 1 - \Phi(\hat{w}) - \varphi(\hat{w})\left(\frac{1}{\hat{w}} - \frac{1}{\hat{u}}\right) & if\ x \neq \mu \\ 0.5 - \frac{1}{\sqrt{2\pi}}\left\{\frac{K'''(0)}{6K''(0)^{\frac{3}{2}}} - \frac{1}{2\sqrt{K'(0)}}\right. & if\ x = \mu \end{cases} \tag{4}$$

$\hat{w}$ and $\hat{u}_1$ are known by

$$\hat{w} = \text{sgn} = (\hat{s})[\hat{s}x - K(\hat{s})]^{0.5}$$

$$\hat{u}_1 = (1 - \exp(-\hat{s}))[K''(\hat{s})]^{0.5}$$

and the saddlepoint $\hat{s}$ solves $K(\hat{s}) = x$. The symbol $\phi$ and $\Phi$ indicate the normal probability distribution function (PDF) and the CDF correspondingly as well as $\text{sgn}(\hat{s})$ takes sign $(\pm)$ for $\hat{s}$ (Butler, 2007).

*1.2 Second Continuity-Modification*

Describe $x^- = x - .05 \in \tau_x$ as the continuity-modification or offset value of $x$. The second approximation solves the offset value of saddlepoint equation $K'(\tilde{s}) = x^-$. The saddlepoint $\tilde{s}$ and $x^-$ are employed to change the inputs into the cumulative distribution function approximation consistent with

$$\begin{aligned} \tilde{w}_2 &= \text{sgn}(\tilde{s})\sqrt{2[\tilde{s}x^- - K(\tilde{s})]} \\ \tilde{u}_2 &= 2\sinh\left(\frac{\tilde{s}}{2}\right)\sqrt{K''(\tilde{s})} \end{aligned} \tag{5}$$

This leads to the second continuity-modified approximation

$$\hat{P}_{r_2}(X \geq x) = \begin{cases} 1 - \Phi(\tilde{w}_2) - \varphi(\tilde{w}_2)\left(\frac{1}{\tilde{w}_2} - \frac{1}{\tilde{u}_2}\right) & if\ x^- \neq \mu \\ \frac{1}{2} - \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{\frac{3}{2}}} & if\ x^- \neq \mu \end{cases}, \text{ (Butler, 2007)} \tag{6}$$

*1.3 Third Approximation*

This approximation is indicated as $\hat{P}_{r_3}(X \geq x)$ and employs expression (6) with $\tilde{w}_2$ as in (5) and $\tilde{u}_2$ substituted with

$$\tilde{u}_3 = \tilde{s}\sqrt{K''(\tilde{s})}\,, \text{ (Butler, 2007)} \tag{7}$$

**2. Claim Frequency Models (Daniel, 2008)**

Numerous procedures in daily life that calculate events up to a special stage in time could be precisely explained by the supposed Poisson process; now a Poisson process will be defined. Firstly, a Poisson process known as a compilation of random variables $N(t)$ per $t$ about particular group. Explicitly, Poisson process is a reliable form: in every $t > 0$ they calculate the incidents that occur through time 0 and time $t$. The type of incidents relies in the function. One may desire to calculate how many incidents cases claimed via a private made, or how many calls for assistance and how many persons quit from a special company etc. No matter what you may consider by an "incidents", $N(t)$ indicates the incident numbers that happened following time 0 to that moment and comprising time $t > 0$. Consider that $N(0)$ is regarded to exact 0 signifies no incidents could be happened sooner than you begin calculating. As when the incident take a place and supposed to be a random, $N(t)$ is a random variable to every value of $t$. Consider that $N$ itself is presented *random process,* however, differentiating it from random variable $N(t)$ at every value of $t > 0$. Furthermore, to recognize the counting process, it required to comprehend the connotation. In addition, the probability response through the amount of

$N(t+h)-N(t)$ in time $t$ to time $t+h$, where $h>0$ and certainly $t\geq 0$. Because $N(t+h)$ equals to the number of incidents up through $t+h$ and $N(t)$ equals the random number up through *t*, the growth is merely the random number of incidents happening firmly after time *t* and up through and comprising time $t+h$. Consider that $N(t)$ itself could be regarded as an increase, explicitly from time (0) to time *t*, given that $N(t)=N(t)-N(0)$ (Daniel, 2008).

In Poisson process the balance that incidents are generally happening is very significant.

The proportion of function $\lambda$ reports the average as $\lambda(t)$ at time *t*. Consider that the average could vary with time, however, the rapidity of a vehicle the average at which it is passing distance could differ with time. You obtain the total distance passed by a vehicle through a time period by multiplying the average by the length of the time interval, provided the average is a constant; if the average varies, to merge the average function over the interval, which gives the same outcome as the simpler formula when the average is constant. As well as, to merge Poisson process's average function over an interval we gain the average of incidents in that period (Daniel, 2008).

*2.1 Definition and Properties of Poisson Random Variable*

A Poisson random variable *Y* with mean $\lambda$ is a random variable with the subsequent characteristics:

1) It is only potential values are the positive integers 0, 1, 2, 3, ... and

2) $\Pr[Y=x]=\frac{e^{-\lambda}}{x!}\lambda^{x}$

Within these characteristics we can find that

3) $\mathrm{E}[Y]=\mathrm{Var}[Y]=\lambda$.

The meaning of the previous explanation of Poisson processes.

*2.2 Definition and Characteristics of Poisson Process*

A Poisson process *N* with rate function $\lambda$ has the next properties:

1) *N* is a counting process $N(0)=0$ and for $t>0$ $N(t)$ is non decreasing and accept only positive integer values 0, 1, 2, 3, ... and, thus, it could be regarded as the random number of incidents of significance happening after time 0 and by time *t*.
2) *N* has separate increases of any group of increases $N(t_i+h_i)-N(t_i)$ for $i=1,2,...,n$ is separate for any positive integer *n*, given that the time intervals $(t_i,t_i+h_i)$ are interference (interference at an endpoint is acceptable).
3) Every $t\geq 0$ and $h>0$, the increases $N(t+h)-N(t)$ is a Poisson random variable with mean $\lambda=\int_{t}^{t+h}\lambda(z)dz$.

4) If the rate function $\lambda$ is indeed stable, afterward, *N* is described a homogeneous Poisson process. In usual practice, Poisson process has typically implied "homogeneous Poisson process", whereas "non-homogeneous Poisson process" has been employed to signify a rate function that isn't stable. However, the terms must not puzzle you, because you could constantly observe to perceive whether $\lambda$ is stable. Moreover, you observe declarations such as "events happen at the Poisson rate 3 per hour"; this is stenography for "events are happening consistent with a Poisson process with stable rate function $\lambda=3$ per hour".

*2.3 Random-Sum Poisson Process*

Companies offer medical insurance to their staff, they are certainly worried from frequent prompts, the random number of prompts occurred. Furthermore, they are worried from prompts strictness, the random size of every claim. However, they are particularly worried from the total sum of every one of the prompts. The total of the random numbers of random variables, and as it should really difficult to be analyzed. However, the probability distribution is identified as the random sum distribution.

2.3.1 Definition of Random-Sum Poisson Process

The random-sum Poisson process *Y* has the subsequent properities:

1) For $t>0, Y(t)=\sum_{j=1}^{N(t)}X_j$.

2) $N$ is the Poisson process with rate function $\lambda$;
3) Every random variable $X_j$ has similar distribution as a single random variable $X$;
4) For all $t$, the random variables $N(t)$ and all the $X_j$ from independent group;
5) If $N(t)$ equal 0 for a special value of $t$, after that the empty sum $Y(t)$ is regarded to be equal 0, thus, specifically, $Y(0) = 0$.

The number of $N$ is linked with the frequent prompts in the specified group of the policies and regulation. The general distribution of $X_1, X_2, \ldots$ is indicated by $X$. Consider $X$ forms the sum of a random prompts made in this file of insurance contract. Once the frequent prompts $N$ with a constant parameter $\lambda$ follow a Poisson distribution, the total prompts $Y$ is supposed to have a random-sum Poisson distribution with mean $\lambda = E[N]$ and the variance $Var[Y] = \lambda E[X^2]$. The MGF is given by

$$M_Y(s) = M_N[\ln M_X(s)] = \exp[\lambda(M_X(s) - 1)] \tag{8}$$

**3. Numerical Example**

Assume that vehicle accident prompts are submitted with a vehicle insurer at the Poisson rate $\lambda = 5$ per hour, and that independent $X$ of persons badly hurt in every accident are Binomial random variables with parameters $n = 2$ and $p = 0.2$. Subsequently, the total number $Y$ of those badly hurt is a random-sum Poisson process. As previously showed, random sum Poisson process are extremely difficult and hard to investigate, and, therefore, approximation techniques are frequently employed. Saddlepoint methods defeat this difficulty. Saddlepoint methods are influential tool for getting precise expressions for distribution function that isn't recognized in closed form. Saddlepoint methods roughly surpass other techniques regarding calculating expenses, while it does not inevitably surpass them concerning correctness.

The total $Y(t)$ are supposed to have a Poisson-Binomial random-sum distribution when i.i.d. random variables $X_i's$ follow Binomial $(n, p)$ distribution. The MGF of a total of prompts $Y$ is identified as

$$M_{Y(t)}(s) = M_N[\ln M_X(s)], \text{ (Hogg \& Tanis, 1983)} \tag{9}$$

where "ln" is the natural log function.

Furthermore, the CGF of a total of prompts $Y$ identified as the CGF for $N$ is known by

$$K_N(s) = \ln[M_N(s)] = \lambda(e^s - 1) \tag{10}$$

And for $X_i's$ follow Binomial $(n, p)$ distribution, the CGF is described as

$$K_X(s) = \ln[M_X(s)] = \ln(pe^s + q)^n \tag{11}$$

Where $p + q = 1$. Subsequently, it possible to drive the CGF for the Poisson-Binomial random sum distribution as

$$K_{Y(t)}(s) = K_N(K_X(s)) = \lambda[(pe^s + q)^n - 1)] \tag{12}$$

In this case the saddlepoint equation is

$$K'_{Y(t)}(\hat{s}) = n\lambda pe^{\hat{s}}(pe^{\hat{s}} + q)^{n-1} \tag{13}$$

Next, the saddlepoint can be found as $\hat{s} = \hat{s}(x)$ which indicates as the only conclusion to the saddlepoint formula $K'(\hat{s}) = x$.

However, second and third derivative of the CGF $K''(\hat{s})$ as well as $K'''_Y(\hat{s})$ also

$$K''_{Y(t)}(\hat{s}) = n(n-1)\lambda p^2 e^{2\hat{s}}(pe^{\hat{s}} + q)^{n-2}$$

$$K'''_{Y(t)}(\hat{s}) = n(n-1)(n-2)\lambda p^3 e^{3\hat{s}}(pe^{\hat{s}} + q)^{n-3}$$

This leads to the saddlepoint mass function for Poisson-Binomial random sum which is known as

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi n(n-1)\lambda p^2 e^{2\hat{s}}(pe^{\hat{s}} + q)^{n-2}}} \exp(\lambda[(pe^{\hat{s}} + q)^n - 1) - \hat{s}x) \tag{14}$$

Also the saddlepoint method to CDF for Poisson- Binomial random sum presented below.

*3.1 First Continuity-Modification*

The first continuity-modification will be described as:

$$\hat{w} = \text{sgn}(\hat{s})\sqrt{2(x\hat{s} - \lambda[(pe^{\hat{s}} + q)^n - 1)])}$$
$$\tilde{u}_1 = (1 - \exp(-\hat{s})\sqrt{n(n-1)\lambda p^2 e^{2\hat{s}}(pe^{\hat{s}} + q)^{n-2}} \quad (15)$$

This leads to the first continuity-modified method known in expression (4).

*3.2 Second Continuity-Modification*

The second continuity modification will describe $x^- = x - .05 \in \tau_x$ as the continuity-modified or offset value of $x$. The second method solves the offset saddlepoint equation $K'(\tilde{s}) = x - 0.5$ ; and the saddlepoint $\tilde{s}$ and $x^-$ are employed to change the inputs into the CDF method consistent with

$$\hat{w}_2 = \text{sgn}(\hat{s})\sqrt{2\{\hat{s}(x - 0.5) - \lambda[(pe^{\hat{s}} + q)^n - 1)]\}}$$
$$\tilde{u}_1 = 2\sinh(\frac{\tilde{s}}{2})\sqrt{n(n-1)\lambda p^2 e^{2\hat{s}}(pe^{\hat{s}} + q)^{n-2}} \quad (16)$$

Subsequently, the second continuity-modified approximation specified in expression (6) will be found.

*3.3 Third Approximation*

This approximation is indicated as $\hat{P}_{r_3}(X \geq x)$ and employs expression (6) with $\tilde{w}_2$ as in (16) and $\tilde{u}_2$ substituted with

$$\tilde{u}_3 = \tilde{s}\sqrt{n(n-1)\lambda p^2 e^{2\hat{s}}(pe^{\hat{s}} + q)^{n-2}} \quad (17)$$

## 4. Conclusion

This study indicated that, the saddlepoint methods to the CDF for random sum Poisson-Binomial distribution in discrete form. Furthermore, the methods estimate the random-sum variable components assuming presence of the MGF.

**Acknowledgements**

The study is supported financially by Taibah University in Al-Madinah, Kingdom of Saudi Arabia. The authors Thanks the referees for constructive cimments.

## References

Borowiak, D. S. (1999). A saddlepoint approximation for tail probabilities in collective risk models. *Journal of Actuarial Practice, 7*, 1 - 11.

Butler, R. W. (2007). *Saddlepoint approximations with applications*. USA: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511619083

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics, 25*, 573-650. http://dx.doi.org/10.1214/aoms/1177728652

Daniels, H. E. (1987). Tail probability approximations. *International Statistical Review, 55*, 37-48. http://dx.doi.org/10.2307/1403269

Daniel, J. W. (2008). *Poisson processes (and mixture distributions)*. Retrieved from http://www.actuarialseminars.com/Misc/PPjwd.pdf

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (3rd ed.). John Wiley & Sons, Inc., pp. 386-388. http://dx.doi.org/10.1002/0471715816

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). USA: Collier Macmillan Publisher.

Hogg, R. V., & Tanis, E. A. (1983). *Probability and Statistical Inference* (2nd ed.). New York: Macmillan Publishing Co.

Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci., 3*, 213-238. http://dx.doi.org/10.1214/ss/1177012906

Skovgaard, I. M. (1987). Saddlepoint expansions for conditional distributions. *Journal of Applied Probability, 24*, 875-887. http://dx.doi.org/10.2307/3214212

Terrell, G. R. (2003). A stabilized Lugannani-Rice formula aymposium on the Interface. *Computing Science and Statistics, Annual report department of statistics Virginia tech March 14*.

Modern Applied Science; Vol. 7, No. 6; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

# Statistical Measures of Fidelity Applied to Diagnostic Species in Plant Sociology

Manuel Peinado[1], Gustavo Díaz[2], Francisco Manuel Ocaña-Peinado[3], Juan Luis Aguirre[2], Miguel Ángel Macías[4], José Delgadillo[5] & Alejandro Aparicio[2]

[1] Franklin Institute of North American Studies, University of Alcalá, Alcalá de Henares, Spain

[2] Cátedra de Medio Ambiente, University of Alcalá, Alcalá de Henares, Spain

[3] Departament of Statistics and Operational Research, University of Granada, Granada, Spain

[4] Departament of Environmental Science, University of Guadalajara, Guadalajara, JAL, Mexico

[5] Herbarium BCMEX, Autonomous University of Baja California, Ensenada, BC, Mexico

Correspondence: Manuel Peinado, Franklin Institute of North American Studies, University of Alcalá, Alcalá de Henares 28805, Spain. Tel: 34-91-885-4945. E-mail: manuel.lorca@uah.es

Received: March 25, 2013 Accepted: May 7, 2013 Online Published: May 24, 2013

doi:10.5539/mas.v7n6p106 URL: http://dx.doi.org/10.5539/mas.v7n6p106

**Abstract**

The idea of a diagnostic species is an important concept in plant sociology. However, since over a century ago, when the term "association" was introduced, the identification of diagnostic species has been among the most controversial topics in phytosociological practice. With the aim of promoting methodological standardization in plant sociology, this paper addresses: 1) the need to distinguish between the concepts and methods involved in the definition of syntaxa (analysing relevés, characterization, diagnosis, naming and typification), and 2) the need to support and improve existing syntaxonomical classification schemes using statistical measures of fidelity to identify diagnostic species. The phytosociological literature describes numerous different approaches to the designation of diagnostic species. Here, we examine two such approaches to determine diagnostic species using as an example the class *Atriplici julaceae-Frankenietea palmeri* within the context of a data set of 5092 relevés taken of coastal plant communities distributed along the Pacific rim of North America. Diagnostic species were determined using both the phi-coefficient of association to detect differential species and the Ochiai index to designate character species. Our findings support the results obtained by combining classic phytosociological methods (expert knowledge, rearrangement of relevé tables, presence tables, etc.) with clustering methods.

**Keywords:** association, diagnostic species, fidelity measures, ochiai index, phi-coefficient, vegetation classification

## 1. Introduction

Over one hundred years have passed since the Third International Botanical Congress held in Brussels in 1910 coined the first formal definition of the plant association, which marked the birth of plant sociology (Blasi, Biondi, & Izco, 2011). At the congress, amidst intense dispute between plant sociologists and physiognomists, Flahault and Schröter obtained unanimous approval for a definition of association as an abstract vegetation unit that features: 1) a definite floristic composition, 2) a uniform physiognomy, and (3) its occurrence in uniform habitat conditions (Flahault & Schröter, 1910). Given its complete nature, this definition continues to be valid. Indeed, in his revision of the phytosociological association concept, Willner (2006) suggested going back to Flahaut and Schröter's definition, and a similar definition of association was provided by the US National Vegetation Classification (Jennings, Faber-Langendoen, Loucks, Peet, & Roberts, 2009). Despite being such a comprehensive definition, there is still no consensus as to the practical application of the concept (Willner, 2006; Biondi, 2011). This is in large measure because the original concept required at least one clarification: What exactly is a "definite floristic composition"?

Three years after that congress, it was Braun-Blanquet who for the first time selected one characteristic feature of an species he termed "fidelity", as being the most valuable for association diagnosis (Braun-Blanquet & Furrer, 1913). By this, Braun-Blanquet refers to the complete or partial restriction of certain species of narrow ecological amplitude to one particular association. In later publications, however, Braun-Blanquet substantially modified his initial concept. First he introduced the "characteristic species combination" as the main feature for

the diagnosis of an association (Braun-Blanquet, 1925); and later he adopted Koch's (1926) idea of differential species to distinguish between two syntaxonomically related units and restricted the geographical validity of character species to relatively small and ecologically homogeneous regions (Braun-Blanquet & Moor, 1938).

Despite numerous discussions in congresses and meetings in the middle of the last century, a summary of which may be found in Westhoff and van der Maarel (1973: 625), ample consensus was reached for considering that the floristic composition that defines an association, and, by extension, the rest of syntaxonomic hierarchies, is comprised of a diagnostic species that includes character and differential species, along with constant companions. Thus, Syntaxonomy can be regarded as a hierarchy system whose units are defined by groups of diagnostic species and not only by its character species in the strict sense.

Plant communities can be viewed as a hypothesis that predicts the conditions of a habitat (or viceversa) within a given area (Pignatti, 1980). Floristic difference contains no useful information as long as it cannot be interpreted as a reflection of a different habitat or a different vegetation history. This is because one of the basic objectives of Syntaxonomy is to establish a hierarchical system reflecting the patterns of similarity and dissimilarity between vegetation types (Braun-Blanquet, 1964; Westhoff & van der Maarel, 1978) in an effort to reduce the diversity of vegetation to a level that is easily comprehensible by the human mind (Mirkin, 1989; Moravec, 1989), thus facilitating communication among plant scientists (Willner, 2006). The goal is analogous to that described by Stuessy (1997) for plant taxonomy: to provide "a biological classification of high predictive value".

The phytosociological literature contains numerous different approaches to the designation of diagnostic species. Frequently, these results in discrepancies between the lists of diagnostic species published for the same community (Chytrý et al., 2002b; Khan et al., 2013a). The fact that species differ in their diagnostic value is reflected in the concept of fidelity, i.e., the extent to which a species is concentrated in a given vegetation unit. The fidelity of a species determines whether it can be considered a differential or character species or just a companion or accidental species (Braun-Blanquet, 1918). The faithful species concept *sensu* Braun-Blanquet positions itself either in the context of a given phytogeographic unit whose extension may vary within conventional phytogeographic classification systems, or alternatively in the ecological context, which is similarly variable, given that a faithful species can be so, due to very broad factors (macroclimate or limestone soils would be two general examples) or highly specific factors (microclimate or microedaphic, for instance).

Despite its usefulness, the concept of fidelity has inherent an essential concept problem, the old Aristotelian dilemma of the circular argument: a vegetation unit is defined by diagnostic species (excluding constant companions) and at the same time diagnostic species are those that show a preference for this vegetation unit. Although this circularity is not resolvable using logical arguments (Poore, 1955), this does not prevent the quantification of fidelity from being optimised if valid algorithms are found to identify species with statistically significant fidelity values (Bruelheide, 2000; Khan et al., 2011) that support other numerical classification methods based on a full comparison of the floristic assemblage (De Cáceres & Wiser, 2011; Khan et al., 2013b).

## 2. Measures of Fidelity

In this section, the following terminology is used. Generally speaking, the reader may consider field observations as sampling units such as phytosociological relevés, or any other type of sampling unit that can be used in presence/absence data sets. Here, we distinguish between the syntaxa whose fidelity we wish to assess, or target syntaxa (TSY), and the remaining relevés of the data set, which will serve as samples to compare with. These samples are here designated as reference groups (REF).

The first quantitative method used to assess fidelity was described in detail by Szafer and Pawłowski (1927), who prepared a numerical table in which five fidelity classes were defined. With slight modifications, this table appears in almost every review and textbook of Phytosociology (Becking, 1957; Westhoff & van der Maarel, 1973; Braun-Blanquet, 1979; Dierschke, 1994). Although Szafer and Pawłowski's guide to fidelity was valid at the time, its shortcomings were well known and its practical application to select diagnostic species involved long optimization processes that included several subjective choices within different contexts (Koči, Chytrý, & Tichý, 2003). These shortcomings have been discussed in reports in which alternative measures of fidelity are proposed based on cover (Barkman, 1989), constancy or frequency (Bergmeier et al., 1990; Dengler, 2003) data. In general, although tables with numerous relevés were considered, these contained either ecologically-related syntaxa or ecologically and floristically different syntaxa within a reduced geographical area. In either case, such highly intuitive measures of fidelity were of low statistical value and poorly reliable (Chytrý & Otýpková, 2003).

A characteristic species can be interpreted as a special kind of differential species: a differential species may occur in one or more vegetation unit, whereas a characteristic species should occur in only a single vegetation unit (Barkman, 1989). In reality, both character species and differential species are types of diagnostic species

but viewed in different contexts (Chytrý, Tichý, Holt, & Botta-Dukát, 2002a; De Cáceres, Font, & Oliva, 2008). The context of differential species in the sense of Koch (1926) can be achieved in two main ways: (a) by comparing a given TSY using as REF immediately higher syntaxa: associations within their respective alliances, alliances within their respective orders and orders within their respective classes; or (b) using as REF the syntaxon that most resembles the TSY floristically, as proposed by Becking (1957: 447) and Barkman (1989: 109). Both approaches can be used to search out differential species among different syntaxa, but it should be noted that all species not included in such syntaxa will be excluded from the comparison, as will the rest of the vegetation units occurring in the area under study. As a result, the measure of fidelity of any species will be limited by its context: i.e., that of the vegetation units compared. If the species shows broad ecological amplitude, it could appear in other vegetation units not used as references and consequently its fidelity values will be of little statistical significance.

To circumvent this problem, if the number of relevés in the data set is maximized, this will both broaden the ecological and geographical context and we will only have to syntaxonomically delineate the TSY; the rest of the data set will serve as a reference unit without the need for its syntaxonomic organization. If, what is more, the method used to calculate fidelity is based on presence/absence data rather than a quantitative measure, we could even use as reference samplings comprising quantitative data obtained by methods that considerably differ from strictly phytosociological relevés, provided that deviations that could provoke differences in the sampling plot size are considered (Dengler, Löbel, & Dolnik, 2009).

Fidelity is a relative measure: it compares the presence of a given species within a given TSY with that of a REF. Working on cenological fidelity, Juhász-Nagy (1964) distinguished three forms of fidelity, which were summarized by De Cáceres et al. (2008) as: (1) the asymmetric fidelity of the TSY to the species, when all relevés belonging to the TSY contain the species; (2) the asymmetric fidelity of the species to the TSY, when the species occurs only in relevés belonging to the TSY; and (3) mutual fidelity, when both elements are symmetrically faithful to each other.

Optimal measures of fidelity are those that tend towards maximal symmetry, that is, to mutual fidelity (Dufrêne & Legendre, 1997). However, although special attention should be paid to mutual fidelity, the role of the ecological and geographical context also needs to be assessed, since the fidelity of diagnostic species statistically depends on the number of samples in the reference data set in that the more relevés used as references, the greater will be the statistical significance of the measure of fidelity. Hence, the first question to consider is how to determine the context in which to evaluate measures de fidelity.

The methods used traditionally by phytosociologists to select diagnostic species either limit the ecological setting and broaden the geographical setting (e.g., basophilic beech woods in Europe), or restrict the geographical context and expand the ecological setting (as usually done in phytosociological guides for a given region in which ecologically diverse communities are compared). In both cases, the diagnostic species will be delimited by one or the other context such that their validity as indicators will be poor, as will their statistical validity (Chytrý et al., 2002a, 2002b). Obviously, if we increase the number of samples used to compare with, the statistical significance of the diagnostic species detected will also increase.

Whatever the geographic setting, several criteria can be used to select the relevés to include in a reference data set (REF) with which we will compare the given target syntaxon (TSY). De Cáceres et al. (2008) propose the use of a dual strategy with two different objectives: 1) to detect differential species depending on the context, which was the criterion used by Becking (1957) and Barkman (1989); and 2) to assess the diagnostic value of the species regardless of the context such that this is as significant as possible, aimed at identifying the truly faithful diagnostic species, i.e., the character species.

To contextually search out differential species is interesting from a standpoint of syntaxonomic classification. However, the results are highly dependent on the context and, though fidelity measures are comparable against each other, they are not statistically representative and thus their predictive value as indicators is not too reliable. In contrast, fidelity analyses based on data sets for a high number of relevés fulfil the second objective and offer a significant diagnostic value within a given geographical setting, though they do not detect differential species. This problem is resolved by the combined use of the phi-coefficient and Ochiai index (*OI*) since the former detects the differential species, while the latter does so to find the diagnostic species within a given geographical context. Obviously, the greater the geographical context and the more ecologically diverse the syntaxa comprising the REF, the more representative will be the diagnostic species. If in addition these symmetric measures of fidelity are complemented with measures of asymmetric fidelity, an extremely reliable analysis is obtained of the character species of any syntaxon (Table 1).

Table 1. Summary of main features of the diagnostic species and the methods usable for their identification (Modified from De Cáceres et al., 2008)

| Type | Statistical measure | Description |
|---|---|---|
| 1.- Diagnostic species | Ochiai Index, *OI*, Є [0.1] | Any species whose presence or absence in a given vegetation unit can be used to indicate it belongs or not to a syntaxon. |
| 1a.- Constant species | Constancy $c = n_p/N_p$, Є [0.1] | A species showing a significantly high frequency in a given TSY. Its diagnostic value derives from the fact that its absence is evidence for not assigning the type to the sampled community. In a non-strictly phytosociological context, it can also be called the "sensitivity" of the indicator. |
| 1b.- Asymmetrically faithful species | Presence $p = n_p/n$, Є [0.1] | A species whose occurrence is significantly restricted within the faithful species communities of the target unit. Its diagnostic value derives from the fact that its presence in a sampled community provides a strong basis for assigning the unit to the community. This diagnostic value can also be referred to as a positive predictive value). |
| 2. Differential species | $\phi$-Coefficient, Є [-1.1] | A species whose presence in a given vegetation unit can be used to differentiate between that unit and similar types. |

In this table and throughout the rest of this article, we use the same notations as Bruelheide (2000): $N$ = total number of relevés of the whole data set (TSY + REF); $N_p$ = number of relevés belonging to the TSY; $n$ = number of species occurrences in the whole data set; $n_p$ = number of species occurrences in the TSY.

*2.1 The Phi-Coefficient of Association*

Chytrý et al. (2002a) compared the different binary coefficients commonly used to determine mutual fidelity in plant communities and found that the index that provided the best results was the phi-coefficient of association ($\phi$) defined by Sokal and Rohlf (1995):

$$\Phi = \frac{N \cdot n_p - n \cdot N_p}{\sqrt{n \cdot N_p \cdot (N-n) \cdot (N-N_p)}}$$

The phi-coefficient takes values ranging from -1 (maximum negative fidelity) and +1 (maximum positive fidelity). Positive values indicate that the species and the vegetation unit co-occur more frequently than would be expected by chance. Larger values indicate a greater degree of mutual fidelity. A value of 1 indicates that the species and the vegetation unit are completely faithful to each other, because ($np = n = Np$), i.e., the species occurs at all sites of the vegetation unit. For the identification of diagnostic species, positive $\Phi$-values are of particular importance, although negative $\Phi$-values can be also used for negative differentiation of community types, especially if there are not too many site groups in the given typology (Tichý & Chytrý, 2006; Khan et al., 2011).

The phi-coefficient depends on the size of the TSY ($N_p$), which may vary from 1 to ($N$ - 1), and could lead to unreliable conclusions when the data sets contain site groups of unequal size, commonly found in the

phytosociological literature. Tichý and Chytrý (2006) resolved the problem by equalizing the size of the site group to the size of $N_p$ for all the TSY within the data set ($N$), such that the fidelity measures obtained by calculating $\phi$ are statistically meaningful. Table 2 provides a practical example of the use of this method on the relevés of the different TSY included in this article.

Phi-coefficient values were calculated using the program JUICE 7.0 (www.sci.muni.cz/botany/juice.htm), which, besides estimating other statistical measures useful for the analysis of plant communities (*cf*. Tichý, 2002a), is also a powerful tool for ordering tables containing data for any number of relevés and to obtain conventional frequency synoptic tables when only absence/presence data are introduced, or mean frequency or cover tables, when working with cover or abundance-dominance data.

*2.2 The Ochiai Index* (*OI*)

Following the first proposal of the use of the phi-coefficient, a series of works of its practical application ensued (Cerná & Chytrý, 2005; Knollová, Chytrý, Tichy, & Hajek, 2005; Koci et al., 2003; Petrík & Bruelheide, 2006). These studies described the method as extremely useful for identifying mutual fidelity but, owing to its independence of the context, it could not be employed to detect differential species, given that these by definition are dependent on the setting. Moreover, for large data sets, the phi-coefficient can be affected to the extent that its statistical power is low.

When $N$ increases, the number of species deemed significantly diagnostic will naturally increase, although this does not necessary imply changes in $\phi$-values. Phytosociological data sets contain many different vegetation types. Thus, when $N$ increases, the ecological context of the data set broadens and the frequency of almost all the species diminishes. Indeed, when dealing with large data sets of increasing ecological range, the $n$-value of any given species will eventually stop growing. Even at this point, however, more relevé data could still be added. The $\Phi$-values obtained would increase for all species, because adding double zeros increases the correlation between two binary variables (De Cáceres et al., 2008).

For large data sets, when $N$ tends towards infinity, this statistical distortion is avoided using *OI*, an index first used by Ochiai (1957) in a study on fish populations off the Asian Pacific coast and then used by Janson and Vegelius (1981) in other ecological association studies:

$$OI = \sqrt{\frac{n_p^2}{n \cdot N_p}}$$

Since the Ochiai index is independent of $N$, it is a measure of mutual fidelity between a given taxon and the TSY that excludes those relevés not belonging to the TSY. If its two components are, however, separated, the first is a measure of constancy, an indicator of the asymmetrical fidelity of the TSY to the species (Juhász-Nagy's asymmetric fidelity type 1), and the second is an asymmetrical measure of fidelity type 2, i.e., of the species to the TSY. An additional value of the *OI* is its close relationship with an index that assesses the indicator role of the species, *IndVal,* which is much used in ecological studies (De Cáceres & Legendre, 2009).

De Cáceres et al. (2008) proposed an alternative method for statistical measures that basically consists of two measures: the phi-coefficient to find the differential species (thus inverting its original intention) and *OI*, used both integrally and decomposing it into both components to search out the regional diagnostic species (Table 1). The setting of the region would obviously vary according to the reference sample from which the relevés arose.

The aim of the present study was to promote methodological standardization in syntaxonomy by stressing two main points: the need to explicitly distinguish between the procedures involved in the definition of syntaxa, and the need to support and improve the syntaxa defined using statistical measures of fidelity, a process included within the broad concept of "consistency in assignment" (De Cáceres & Wiser, 2011). With such an objective in mind, we here assess the efficacy of the four fidelity measures provided in Table 1, using as TSY communities whose syntaxonomical scheme has been established by us in earlier works (Peinado, Aguirre, Delgadillo, & Macías, 2008; Peinado, Ocaña-Peinado, Aguirre, Delgadillo, & Díaz Santiago, 2011; Peinado, Aguirre, Macías, & Delgadillo, 2011). Owing to their peculiar floristic composition and restricted geographical distribution these target syntaxa are a useful test of how such measures vary according to both the geographical and ecological context.

## 3. Material and Methods

### *3.1 Data Sets*

The data for the TSY, comprising 224 relevés belonging to the class *Atriplici julaceae-Frankenietea palmer*, are provided in Table 2. The endemic assemblage of the communities of this class is remarkably high (Peinado et al., 2008, 2011a). This makes these data highly representative for a comparative analysis within its own internal context to detect differential species, and for stepwise ever-wider range comparisons both in the geographic and ecological context to detect diagnostic species.

The whole data set was obtained by combining three successive partial data sets that step-by-step expanded the geographical and ecological contexts with respect to the TSY. The first partial data set was prepared by combining the data for TSY with REF-1, comprising 252 relevés from other coastal communities of Baja California. The second partial data set combined the relevés in the first partial data set with a second reference group (REF-2), and comprised the 724 relevés taken of beach and dune vegetation along the Pacific coast from California to Alaska. Thus, the groups REF-1 and REF-2 include a total de 976 relevés of psammophilous communities taken from the southern tip of Baja California to Cook Inlet, Alaska. Within the third reference group (REF-3), were included 418 relevés related to successional stages of dune forests of the North American Pacific coast. Lastly, within the fourth reference group (REF-4), we included 3461 relevés obtained in all types of communities in western North America, 2909 of which were obtained from the literature and a further 552 from our own unpublished relevés. Thus, the entire data set was comprised of 5092 relevés encompassing 2620 vascular taxa. Although for all the relevés, original data were Braun-Blanquet abundance/dominance data, these were all transformed to presence (1) or absence (0) data.

Table 2. Syntaxonomy, codes and number of relevés of the target syntaxa

| | Associations | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | *Np* | *s* | *N'p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **O-1** | ***Frankenietalia palmeri*** | | | | | | | | | | | | | 101 | 0.3 | 51 |
| **Al-1** | ***Atriplici julaceae-Frankenion palmeri*** | | | | | | | | | | | | | | | |
| | *Atriplex julacea* | | 81 | 88 | 90 | | 25 | 52 | 62 | | 42 | 36 | 100 | | | |
| | *Frankenia palmeri* | 100 | 100 | 100 | 43 | | | 19 | | 29 | 33 | | 24 | | | |
| | *Lycium californicum* | | 28 | 61 | 100 | | 13 | | | | 17 | | | | | |
| | *Euphorbia misera* | | 16 | 100 | 52 | | | | | | | | | | | |
| | *Suaeda taxifolia* | | | | 33 | | 13 | | | | | | 14 | | | |
| **As-1** | ***Atriplici linearis-Frankenietum palmeri*** | | | | | | | | | | | | | 15 | 0.3 | 3.8 |
| | *Atriplex* s. *linearis* | 100 | | | | | | | | | | | | | | |
| **As-2** | ***Atriplici julaceae-Frankenietum palmeri*** | | | | | | | | | | | | | 32 | 0.3 | 8 |
| **As-3** | ***Euphorbio miserae-Lycietum californici*** | | | | | | | | | | | | | 33 | 0.3 | 8.3 |
| **As-4** | ***Dudleyo cultratae-Lycietum californici*** | | | | | | | | | | | | | 21 | 0.3 | 5.3 |
| | *Dudleya cultrata* | | | | 90 | | 25 | 33 | | | | | | | | |
| | *Mirabilis* v. *californica* | | | | 38 | | | | | | | | | | | |
| **O-2** | ***Camissonio crassifoliae-Isocometalia menziesii*** | | | | | | | | | | | | | 123 | 0.5 | 62 |
| **Al-2** | ***Heliantho nivei-Isocomion menziesii*** | | | | | | | | | | | | | 55 | 0.3 | 18 |
| | *Camissonia crassifolia* | | | | | 85 | 13 | 57 | | 80 | 75 | | 24 | | | |
| | *Isocoma* v. *menziesii* | | | | | 100 | 13 | 62 | 54 | | | 36 | 14 | | | |
| | *Atriplex* s. *canescens* | | | | | | 50 | 10 | | 25 | 17 | 45 | 19 | | | |
| | *Helianthus* s. *niveus* | | | | | | 75 | 90 | 85 | | | 27 | | | | |
| | *Ephedra californica* | | | | 14 | | 25 | 33 | | | | 100 | | | | |
| | *Cynanchum peninsulare* | | | | | 54 | 13 | | | | 25 | 18 | | | | |
| | *Distichlis spicata* | | | | | | 38 | | | | | | 14 | | | |
| | *Lotus distichus* | | | | | | 25 | 33 | | | | | | | | |
| | *Camissonia* s. *suffruticosa* | | | | | | 38 | 19 | | | | | | | | |
| | *Lotus rigidus* | | | | | | 25 | 10 | | | | | | | | |
| | *Carpobrotus chilensis* | | | | | | 25 | | 46 | | | | | | | |
| **As-5** | ***Loto bryanthii-Isocometum menziesii*** | | | | | | | | | | | | | 13 | 0.3 | 3.3 |
| | *Lotus bryantii* | | | | | 85 | | | | | | | | | | |
| | *Amaranthus watsonii* | | | | | 85 | | | | | | | | | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Dalea brandegeei* | | | 31 | | | | | | | | | | |
| **As-6** | ***Heliantho nivei-Isocometum vernonioidis*** | | | | | | | | | | | 8 | 0.3 | 2 |
| | *Isocoma* v. *vernonioides* | | | | 100 | | | | | | | | | |
| **As-7** | ***Camissonio crassifoliae-Helianthetum nivei*** | | | | | | | | | | | 21 | 0.3 | 3.3 |
| **As-8** | ***Heliantho nivei-Astragaletum anemophili*** | | | | | | | | | | | 13 | 0.3 | 3.3 |
| | *Astragalus anemophilus* | | | | | | 100 | | | | | | | |
| **Al-3** | ***Encelion ventori*** | | | | | | | | | | | 36 | 0.3 | 12 |
| **As-9** | ***Camissonio crassifoliae-Encelietum ventori*** | | | | | | | | | | | 24 | 0.5 | 12 |
| **As-10** | ***Sphaeralceo fulvae-Encelietum ventori*** | | | | | | | | | | | 12 | 0.5 | 6 |
| | *Encelia ventorum* | | | | | | | 100 | 75 | | | | | |
| | *Sphaeralcea fulva* | | | | | | | | 100 | | | | | |
| **Al-4** | ***Lycion richii*** | | | | | | | | | | | 32 | 0.3 | 11 |
| **As-11** | ***Ephedro californicae-Lycietum richii*** | | | | | | | | | | | 21 | 0.5 | 11 |
| **As-12** | ***Lycietum brevipedis*** | | | | | | | | | | | 11 | 0.5 | 5.5 |
| | *Lycium richii* | 19 | | 38 | 10 | | | 17 | 25 | 100 | 100 | | | |

The first 12 columns show the syntaxonomical scheme of the class *Atriplici julaceae-Frankenietea palmeri* summarized from Table S-10 in Peinado et al. (2011a). Scores are presence percentages (rounded) for each association. The last three columns indicate: *Np*, number of relevés belonging to each target syntaxon; *s*, standardization factor for syntaxa of varying size belonging to the same syntaxonomical hierarchical level; *N'p*, number of standardized relevés after multiplying *Np* by *s*. The following abbreviations are used in tables 2, 3 and 4: *Atriplex canescens* subsp. *canescens* (*Atriplex* s. *canescens*), *Atriplex canescens* subsp. *linearis* (*Atriplex* s. *linearis*), *Camissonia cheiranthifolia* subsp. *suffruticosa* (*Camissonia* s. *suffruticosa*), *Helianthus niveus* subsp. *niveus* (*Helianthus* s. *niveus*), *Isocoma menziesii* var. *menziesii* (*Isocoma* v. *menziesii*), *Isocoma menziesii* var. *vernonioides* (*Isocoma* v. *vernonioides*), *Mirabilis californica* var. *californica* (*Mirabilis* v. *californica*).

### *3.2 Stage 1: Internal Analysis of Fidelity and Differential Species*

To evaluate the results obtained in the syntaxonomic classification undertaken using the floristic-sociological approach combined with cluster analysis (*cf*. Peinado et al., 2008, 2011a, 2002b), whose results we summarize in Table 2, in this initial stage, we estimated measures of fidelity ($\phi$ coefficient) within the class by comparing: a) orders against orders; b) alliances against alliances within the same order and; c) associations against associations within the same alliance. The fidelity analysis was not applied to the association *Atriplici julaceae-Frankenietum palmeri,* because this association is the "central association" of the class, which is primarily defined according to the absence of positive diagnostic species (Dierschke, 1988).

Taking into account the different sizes of the TSY, to calculate the phi-coefficient these were all equalized applying factor *s* to the syntaxa compared (Table 2). For the statistical basis of this factor, the reader is referred to Tichý and Chytrý (2006). Confidence intervals for Phi-values were established using Fisher's *z* transformation. To estimate the standard deviation, Fieller and Pearson (1961) correction was used. The program JUICE 7.0 was used to calculate the phi-coefficient.

### *3.2 Stages 2, 3, 4 and 5: Identifying Diagnostic Species*

*OI* was calculated to identify the diagnostic species of the class and to monitor changes in this index in four successive scenarios that each time extended the ecological and geographic context. This was done by detecting changes in the values of *OI, c* and *p* appearing in Table 1. Comparisons were made by considering each TSY versus the set of REF corresponding to each stage:

Stage 2: TSY versus REF-1.

Stage 3: TSY versus REF-1 + REF-2.

Stage 4: TSY versus REF-1 + REF-2 + REF-3.

Stage 5: TSY versus REF-1 + REF-2 + REF-3 + REF-4.

Since *OI* is independent of the relative number of relevés within groups, no measure of standardization was applied. For the confidence intervals of *OI*, the asymptotic approximation of the variance proposed by Janson and Vegelius (1981) was used.

Plant nomenclature follows Wiggins (1980), except for *Helianthus* (Heiser, Smith, Clevenger, & Martin, 1966) and *Isocoma* (Nesom, 1991).

Table 3. Results of the analysis of fidelity within the class *Atriplici julaceae-Frankenietea palmeri*. Syntaxa codes as in Table 2. *Φ*, phi-coefficient of association. UL and LL, upper and lower limits, respectively, of the confidence intervals (95%) for *Φ*. All values are multiplied by 1000 and rounded. Only species with *Φ*-values > 400 are shown

| | LL | *ϕ* | UL |
|---|---|---|---|
| **O-1** | | | |
| *Frankenia palmeri* | 643 | 715 | 775 |
| *Euphorbia misera* | 426 | 530 | 620 |
| *Lycium californicum* | 407 | 513 | 605 |
| **O-2** | | | |
| *Isocoma* v. *menziesii* | 313 | 429 | 532 |
| *Helianthus* s. *niveus* | 288 | 406 | 512 |
| **Al-2** | | | |
| *Helianthus* s. *niveus* | 510 | 632 | 729 |
| *Isocoma* v. *menziesii* | 402 | 543 | 659 |
| **Al-3** | | | |
| *Encelia ventorum* | 916 | 941 | 959 |
| *Camissonia crassifolia* | 678 | 765 | 831 |
| *Sphaeralcea fulva* | 288 | 445 | 579 |
| **Al-4** | | | |
| *Lycium richii* | 699 | 781 | 843 |
| **As- 1** | | | |
| *Atriplex* s. *linearis* | | 1000 | |
| **As-3** | | | |
| *Euphorbia misera* | 350 | 486 | 602 |
| *Fouquieria diguetii* | 301 | 443 | 566 |
| *Errazurizia benthamii* | 269 | 415 | 542 |
| **As-4** | | | |
| *Dudleya cultrata* | 424 | 550 | 655 |
| *Lycium californicum* | 362 | 497 | 611 |
| *Pachycereus schottii* | 331 | 470 | 589 |
| *Mirabilis* v. *californica* | 293 | 436 | 560 |
| *Isocoma* v. *menziesii* | 313 | 429 | 532 |
| *Helianthus* s. *niveus* | 288 | 406 | 512 |
| **As-5** | | | |
| *Lotus bryantii* | 842 | 884 | 915 |
| *Isocoma* v. *menziesii* | 418 | 545 | 651 |
| *Proboscidea althaefolia* | 415 | 542 | 649 |
| *Cynanchum peninsulare* | 313 | 454 | 575 |

| | LL | $\phi$ | UL |
|---|---|---|---|
| **As-6** | | | |
| *Isocoma* v. *menziesii* | 859 | 897 | 925 |
| *Distichlis spicata* | 351 | 487 | 603 |
| *Abronia maritima* | 346 | 483 | 600 |
| *Camissonia* s. *suffruticosa* | 269 | 415 | 542 |
| *Helianthus* s. *niveus* | 253 | 401 | 530 |
| **As-7** | | | |
| *Helianthus* s. *niveus* | 455 | 576 | 676 |
| *Lotus distichus* | 302 | 444 | 567 |
| **As-8** | | | |
| *Astragalus anemophilus* | 960 | 971 | 979 |
| *Carpobrotus chilensis* | 395 | 525 | 635 |
| *Helianthus* s. *niveus* | 347 | 484 | 601 |
| **As-9** | | | |
| *Camissonia crassifolia* | | 1000 | |
| *Encelia ventorum* | 702 | 777 | 835 |
| *Haplopappus sonorensis* | 289 | 433 | 558 |
| **As-10** | | | |
| *Sphaeralcea fulva* | 923 | 944 | 959 |
| *Encelia ventorum* | 324 | 464 | 584 |
| **As-11** | | | |
| *Lycium richii* | 451 | 573 | 674 |
| **As-12** | | | |
| *Ephedra californica* | 499 | 613 | 706 |
| *Lycium richii* | 384 | 516 | 627 |

## 4. Results

Table 3 provides the results of the internal analysis of fidelity of the class *Atriplici julaceae-Frankenietea palmeri*. Table 4 shows the values of *OI,* constancy and frequency for the ten species showing the highest values and that may be considered diagnostic of the class. This table summarizes the changes produced in *OI* in the stages 2 to 5. These changes are illustrated in Figure 1.

## 5. Discussion

If the frequencies provided in Table 2 are compared with the values of $\Phi$ appearing in Table 3, it may be clearly seen that the latter confirm the syntaxonomic classification based on classic phytosociological methods backed by clustering methods. However, using phi-coefficients some differential species were detected that had been overlooked by the previous analyses due to the size of the data set (1730 relevés). Such are the cases of *Fouquieria diguetii* and *Errazurizia benthamii,* differential species in association As-3*,* of *Pachycereus schottii* in As-4, of *Distichlis spicata* and *Abronia maritima* in AS-6, and of *Haplopappus sonorensis* in AS-9.

The results of the internal fidelity analysis reveal that the more heterogeneous a syntaxon is, the lower are the phi-coefficient values of association. Thus, when the two orders are compared, the $\Phi$-values for the two differential species of the order O-2, are lower than those corresponding to the differential species of O-1, despite both species being exclusive to the order (Table 2). In contrast, the phi-coefficient of *Frankenia palmeri* is almost double that of the differential species of order O-2, despite the fact that it appears in some associations of this last order. The increased heterogeneity as the cause of the decline in the phi-coefficient also emerges when we compare the three alliances of O-2. Thus, the coefficients of the differential species of alliance Al-2,

which contains four associations, are much lower than those of alliances Al-3 and Al-4, each of which encompasses two associations.

The internal heterogeneity of a given syntaxon gives rise to marked drops in the *Φ*-values of those plants that, although being practically character species of this syntaxon, are absent from some of its components. Such is the case of *Euphorbia misera* in the order O-1, despite not appearing in O-2. However, its absence from the association As-1, comprised solely of two plants owing to its extreme habitat, is the cause of its relatively low *ϕ*-coefficient.

The use of the phi-coefficient alone to confirm the differential value of a particular taxon is almost absolute in cases of highly delimited ecological or geographical contexts. With regard to the ecological context, this is the case of *Atriplex canescens* subsp. *linearis* in As-1 and of *Isocoma menziesii* var. *vernonioides* in As-6, given they both prosper in soils moistened by brackish water, a different habitat to the rest of the communities of the class. Both species have a wide distribution area, but they grow in settings ecologically differentiated from those of the remaining associations of their own alliance. The geographic context is apparent for the associations characterized by a microendemism, such as in the case of *Astragalus anemophilus* (As-8), *Lotus bryanthii* (As-5), *Camissonia crassifoli*a (As-9) and *Sphaeralcea fulva* (As-10). The fact that some of these species fail to attain a fidelity value of 1 can be attributed to their occasional presence in zones of contact with neighbouring associations or, in the case of *L. bryanthii,* to its absence in one relevé of As-5.

When the *Φ*-values are not so high, the differential role of certain species is reinforced by combining these data with the frequency data. Such are the cases of *Isocoma menziesii* var. *menziesii* and *Helianthus niveus* subsp. *niveus* in the order O-2. Their fidelity values are not relatively very high due to the heterogeneity of the order, which has three alliances. If we examine the frequencies of these two species in Table 2, it becomes clear that they never occur in the communities of the order O-1, such that their fidelity is negative with respect to the latter.

Table 4. Search results for diagnostic species in the four successive stages examined (see methods for details). *Np*, number of relevés in each partial data set; *n*, number of species appearances in each partial data set; c, constancy (*np/Np*), where *np* is the number of species appearances in the TSY; *f*, frequency (*np/n*); *OI*, Ochiai index. LL and UL, lower and upper limits, respectively, of the confidence intervals (95%) for *OI*. Values multiplied by 1000 and rounde

| Stages | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 | 2 | | | 3 | | | 4 | | | 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Np* | 489 | 1213 | 1631 | 5092 | 489 | 1213 | 1631 | 5092 | 489 | 1213 | 1631 | 5092 | 489 | | | 1213 | | | 1631 | | | 5092 | | |
| | *n* | *n* | *n* | *n* | *c* | *c* | *c* | *c* | *f* | *f* | *f* | *f* | LL | *OI* | UL | LL | *OI* | UL | LL | *OI* | UL | LL | *OI* | UL |
| *Atriplex julacea* | 144 | 144 | 144 | 150 | 0.6 | 0.6 | 0.6 | 0.6 | 0.9 | 0.93 | 0.93 | 0.89 | 728 | 746 | 764 | 735 | 746 | 757 | 737 | 746 | 756 | 730 | 731 | 742 |
| *Frankenia palmeri* | 116 | 116 | 116 | 121 | 0.5 | 0.49 | 0.49 | 0.49 | 0.9 | 0.94 | 0.94 | 0.9 | 658 | 676 | 695 | 665 | 676 | 687 | 667 | 676 | 686 | 661 | 662 | 673 |
| *Lycium californicum* | 69 | 69 | 69 | 91 | 0.3 | 0.25 | 0.25 | 0.25 | 0.8 | 0.81 | 0.81 | 0.62 | 432 | 450 | 469 | 439 | 450 | 462 | 441 | 450 | 460 | 391 | 392 | 403 |
| *Euphorbia misera* | 54 | 54 | 54 | 79 | 0.2 | 0.24 | 0.24 | 0.24 | 1 | 0.98 | 0.98 | 0.67 | 464 | 482 | 500 | 471 | 482 | 493 | 472 | 482 | 491 | 397 | 398 | 410 |
| *Lycium richii* | 51 | 51 | 51 | 52 | 0.2 | 0.22 | 0.22 | 0.22 | 1 | 0.98 | 0.98 | 0.96 | 449 | 468 | 486 | 457 | 468 | 479 | 458 | 468 | 477 | 462 | 463 | 474 |
| *Helianthus* s. *niveus* | 41 | 44 | 44 | 44 | 0.2 | 0.18 | 0.18 | 0.18 | 1 | 0.93 | 0.93 | 0.93 | 410 | 428 | 446 | 402 | 413 | 424 | 403 | 413 | 423 | 412 | 413 | 424 |
| *Isocoma* v. *menziesii* | 41 | 46 | 46 | 52 | 0.2 | 0.18 | 0.18 | 0.18 | 1 | 0.89 | 0.89 | 0.79 | 410 | 428 | 446 | 393 | 404 | 415 | 394 | 404 | 414 | 379 | 380 | 391 |
| *Encelia ventorum* | 34 | 34 | 34 | 34 | 0.2 | 0.15 | 0.15 | 0.15 | 1 | 1 | 1 | 1 | 371 | 390 | 408 | 378 | 390 | 401 | 380 | 390 | 399 | 388 | 390 | 401 |
| *Dudleya cultrata* | 29 | 29 | 29 | 29 | 0.1 | 0.13 | 0.13 | 0.13 | 1 | 1 | 1 | 1 | 342 | 360 | 378 | 349 | 360 | 371 | 350 | 360 | 369 | 358 | 360 | 371 |
| *Camissonia crassifolia* | 24 | 24 | 24 | 24 | 0.1 | 0.11 | 0.11 | 0.11 | 1 | 1 | 1 | 1 | 309 | 327 | 346 | 316 | 327 | 338 | 318 | 327 | 337 | 326 | 327 | 338 |

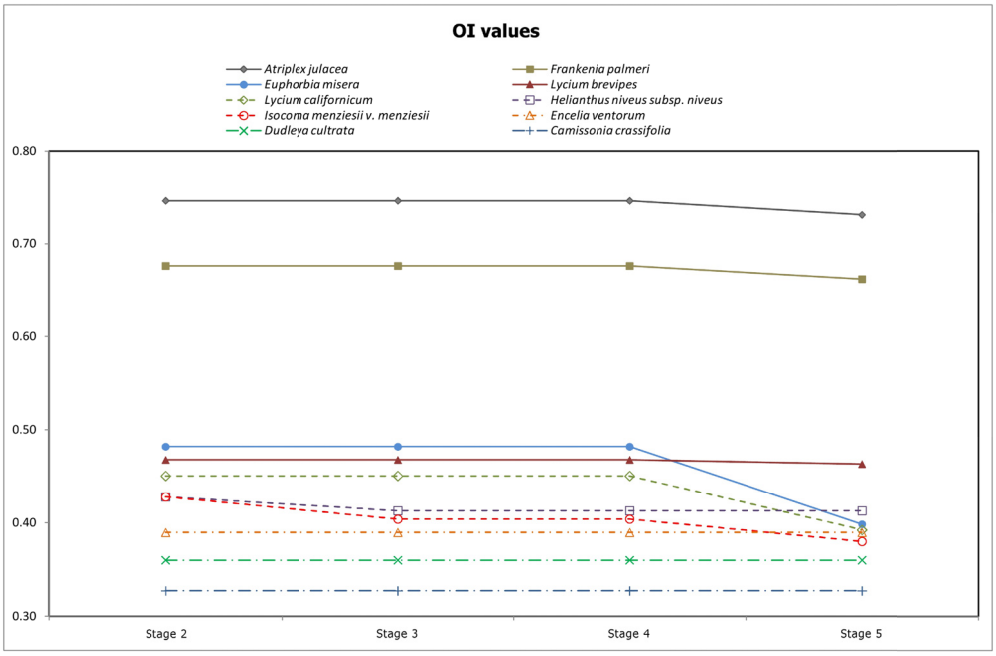



Figure 1. Ochiai index (OI) values recorded for the ten species showing the highest phi-coefficients in the successive stages of increasing the data set

Table 4 (represented in Figure 1) shows the ten species with the highest *OI* values, which are the diagnostics of the class *Atriplici julaceae-Frankenietea palmeri*. Values of asymmetric fidelity are as we would expect. Constancy (*c*) values are lower the more characteristic of a given syntaxon of inferior hierarchy to the class are the plants considered. The *p* value of asymmetric fidelity of a given species to a given syntaxon is highly significant. With the exception of *Lycium californicum,* the ten diagnostic species show values above 0.9 although some decrease considerably in Stage 5 for reasons that we discuss below.

It should be noted that in Stage 2, we included 265 relevés taken of coastal communities of Baja California that ecologically and geographically border with those of the class *Atriplici julaceae-Frankenietea palmeri*. On the Baja Californian coast, it is common to find rocks containing different proportions of sodium, which determines that the most frequent species in the class, *Atriplex julacea* and *Frankenia palmeri*, bioindicators of alkaline substrates, may sporadically appear in communities belonging to other classes. Due to this, both these species show the lowest *p* values among the 10 diagnostic species. In contrast, species of a more reduced ecological context, i.e., the character species of the order O-2 and of its three alliances, show *p* values of 100%.

In Figure 1, it may be seen that *OI* values remain practically constant as the number of relevés in the successive stages increases along with their geographic and ecological contexts. The following are, however, the exception. In the case of *A. julacea* and *F. palmeri*, a small decrease in the index occurs in Stage 5, since the 3461 relevés of REF-4 added in this stage include some taken along the Baja Californian coast that occasionally contain both plants. The drop in *OI* is most marked for *Lycium californicum* and *Euphorbia misera*, two more widely distributed species that appear in communities of the class *Prosopido torreyanae-Fouquieretea splendentis*, 247 of whose relevés were introduced in Stage 5.

The case of *Isocoma menziesii* var. *menziesii* is different to that of the rest of the diagnostic species because its *OI* goes down both in Stages 2 and 5. This is a strict psammophilous, whose distribution extends from Bahía Magdalena on the central Baja Californian coasts to central California. In Baja California it is an exclusive element of secondary dunes and, as such, a species regionally characteristic of the order O-2. Its distribution in California is wider and it inhabits both coastal sands within communities of the class *Ambrosietea chamissonis* (Peinado et al., 2011a) and interior zones (USDA, 2011). Its *OI* drops in Stage 2, when the 522 relevés of *Ambrosietea chamissonis* are incorporated and slightly drops again in Stage 5 for the same reasons described above for *L. californicum* and *E. misera*.

## 6. Conclusions

The use of fidelity measures supports and improves the results of phytosociological classification based on comparing more or less numerous sets of relevés that are, nevertheless, always limited by the ecological and geographical contexts. If we use as reference groups data sets with many relevés, both these contexts will be expanded with the consequence of the increased diagnostic and bioindicator value of a given species. With the recent introduction of large phytosociological databases, it should be possible to preferentially identify diagnostic species of more general validity in data sets that include relevés of most vegetation types occurring across a wide area.

In phytosociological and syntaxonomical practice, the diagnostic value of a species is established using standardized phytosociological methods (comparisons among relevés, table rearrangements, and expert knowledge) supported by numerical methods that lead to a final syntaxonomic classification. Once these syntaxonomic hierarchies have been defined, their diagnostic value as abstract vegetation units and the bioindicator values of their species can be reinforced using the two measures of mutual fidelity analysed here: the phi coefficient and Ochiai index.

The phi-coefficient is dependent on the context and can therefore be used to assess fidelity in previously classified vegetation units with the aim to evaluate the fidelity of the differential species in the syntaxa being compared.

The *OI* is a measure of fidelity that excludes "double zeros" or "double absences". In other words, when the size of the reference data set is increased by incorporating relevés lacking any TSY species, its value remains unchanged. Thus, since *OI* is independent of the size of the TSY and of the whole data set, it is only affected by those relevés that contain the species for which this index is being calculated. As more relevés lacking a given reference species are added, the *OI* will not change, no matter how many relevés the TSY is compared with. However, independently of the size of the latter, when relevés that include species present in the TSY are added to the data set, this causes changes in the *OI*, such that this index is able to discriminate diagnostic species with more precision than the phi-coefficient.

The methods used to assess fidelity are not in themselves classification systems, but can be extraordinarily useful for three essential purposes: a) to *a posteriori* check and improve a classification undertaken using classic phytosociological methods by identifying highly statistically significant diagnostic species independently of the context. This is very useful both in Europe, where the phytosociological approach has given rise to very complete syntaxonomical schemes, and in other world zones with scarce phytosociological tradition but that have available data from surveys conducted through other methods compatible with these new approaches; b) to detect species with high fidelity values in groups of relevés classified or ordered according to ecological or phytogeographical factors when data sets for many relevés are included, a task that is extremely cumbersome or impossible using intuitive or deductive traditional classification methods; and c) to use the taxa returning higher fidelity values as ecological or biogeographical indicators.

## Acknowledgements

The authors thank Miquel de Cáceres and Milan Chytrý for their useful comments on previous versions of this manuscript. This research was supported by a grant from the Franklin Institute of North American Studies (awarded to the University of Alcalá). We also thank Ana Burton for translating the original text.

## References

Barkman, J. J. (1989). Fidelity and character-species, a critical evaluation. *Vegetatio, 85*, 105-116. http://dx.doi.org/10.1007/BF00042260

Becking, R. W. (1957). The Zurich-Montpellier School of Phytosociology. *Botanical Review, 23*(7), 411-488. http://dx.doi.org/10.1007/BF02872328

Bergmeier, E., Härdtle, W., Mierwald, U., Nowak, B., Peppler, C., & Flintrop, T. (1990). Vorschläge zur syntaxonomischen Arbeitsweise in der Pflanzensoziologie. *Kieler Notizen zur Pflanzenkunde in Schleswig-Holsten und Hamburg, 20*, 92-103.

Biondi, E. (2011). Phytosociology today: Methodological and conceptual evolution. *Plant Biosystems, 145*(Supplement 1), 19-29. http://dx.doi.org/10.1080/11263504.2011.602748

Blasi, C., Biondi, E., & Izco, J. (2011). 100 years of plant sociology: A celebration. *Plant Biosystems, 145, Supplement.* http://dx.doi.org/10.1080%2F11263504.2011.602865

Braun-Blanquet, J. (1918). Eine pflanzensoziologische Exkursion durcharacter species Unterengadin und in den schweizerischen Nationalpark. *Beitrage zur geobotanischen Landesaufnahme der Schweizz, 4*, 1-80.

Braun-Blanquet, J. (1925). Zur Wertung der Gesellschaftstreue in der Pflanzensoziologie. *Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich, 70*, 122-149.

Braun-Blanquet, J. (1979). *Bases para el estudio de las comunidades vegetales*. Madrid: Blume.

Braun-Blanquet, J., & Furrer, S. (1913). Remarques sur l'étude des groupements des plantes. *Bulletin de la Société de Géographie du Languedoc, 36*, 20-41.

Braun-Blanquet, J., & Moor, M. (1938). *Verband des Bromion erecti. Prodromus der Pflanzengesellschaften, 5*. Montpellier.

Bruelheide, H. (2000). A new measure of fidelity and its application to defining species groups. *Journal of Vegetation Science, 11*(2), 167-178. http://dx.doi.org/10.2307/3236796

Bruelheide, H., & Chytrý, M. (2000). Towards unification of national vegetation classification: A comparison of two methods for analysis of large data sets. *Journal of Vegetation Science, 11*(2), 295-306. http://dx.doi.org/10.2307/3236810

Černá, L., & Chytrý, M. (2005). Supervised classification of plant communities with artificial neural networks. *Journal of Vegetation Science, 16*(4), 407-414. http://dx.doi.org/10.1111/j.1654-1103.2005.tb02380.x

Chytrý, M., Exner, A., Hrivnák, R., Ujházy, K., Valachovič, M., & Willner, W. (2002b). Context-dependence of diagnostic species: A case study of the Central European spruce forests. *Folia Geobotanica Phytotaxonomica, 37*(40), 403-417. http://dx.doi.org/10.1007/BF02803255

Chytrý, M., & Otýpková, Z. (2003). Plot sizes used for phytosociological sampling of European vegetation. *Journal of Vegetation Science, 14*(4), 563-570. http://dx.doi.org/10.1111/j.1654-1103.2003.tb02183.x

Chytrý, M., & Rafajová, M. (2003). Czech National Phytosociological Database: Basic statistics of the available vegetation-plot data. *Preslia, 75*, 1-15.

Chytrý, M., Tichý, L., Holt, J., & Botta-Dukát, Z. (2002a). Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science, 13*(1), 79-90. http://dx.doi.org/10.1111/j.1654-1103.2002.tb02025.x

De Cáceres, M., Font, X., & Oliva, F. (2008). Assessing species diagnostic value in large data sets: A comparison between phi-coefficient and Ochiai index. *Journal of Vegetation Science, 19*(6), 779-788. http://dx.doi.org/10.3170/2008-8-18446

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology, 90*(12), 3566-3574. http://dx.doi.org/10.1890/08-1823.1

De Cáceres, M., Legendre, P., & Moretti, M. (2009). Improving indicator species analysis by combining groups of sites. *Oikos, 119*.

De Cáceres, M., & Wiser, S. K. (2011). Towards consistency in vegetation classification. *Journal of Vegetation Science, 23*.

Dengler, J. (2003). *Entwicklung und Bewertung neuer Ansätze in der Pflanzensoziologie unter besonderer Berücksichtigung der Vegetationsklassifikation*. Nümbrecht: Archiv Naturwissensch.

Dengler, J., Löbel, S., & Dolnik, C. (2009). Species constancy depends on plot size. A problem for vegetation classification and how it can be solved. *Journal of Vegetation Science, 20*(4), 754-766. http://dx.doi.org/10.1111/j.1654-1103.2009.01073.x

Dierschke, H. (1989). Zur Benennung zentraler Syntaxa ohne eigene Kenn- und Trennarten. *Tuexenia, 8*, 381-382.

Dierschke, H. (1994). *Pflanzensoziologie. Grundlagen und Methoden*. Stuttgart: Ulmer.

Dufrêne, M., & Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs, 67*.

Fieller, E. C., & Pearson, E. S. (1961). Tests for rank correlation coefficients (II). *Biometrika, 48*. http://dx.doi.org/10.1093%2Fbiomet%2F48.1-2.29

Flahaut, C. H., & Schröter, C. (1910). Rapport sur la nomenclature phytogéographique. *Actes III Congrès International de Botanique de Bruxelles*, *1*, 131-164.

Heiser, C. B., Smith, D. M, Clevenger, S. B, & Martin, W. C. (1966). The North American Sunflowers (Helianthus). *Memoirs of the Torrey Botanical Club, 22*(3), 1-218.

Janson, S., & Vegelius, J. (1981). Measures of ecological association. *Oecologia, 49*(3), 371-376. http://dx.doi.org/10.1007/BF00347601

Jennings, M. D., Faber-Langendoen, D., Loucks, O. L, Peet, R. K., & Roberts, D. (2009). Standards for associations and alliances of the U.S. National Vegetation Classification. *Ecological Monographs, 79*(2), 173-199. http://dx.doi.org/10.1890/07-1804.1

Juhász-Nagy, P. (1964). Some theoretical models of cenological fidelity 1. *Acta Botanica Debrecina, 3*, 33-43.

Khan, S. M., Harper, D. M., Page, S., & Ahmad, H. (2011). Species and Community Diversity of Vascular Flora along environmental gradient in Naran Valley: A multivariate approach through Indicator Species Analysis. *Pakistan Journal of Botany, 43*, 2337-2346.

Khan, S. M., Page, S., Ahmad, H., & Harper, D. M. (2013). Identifying plant species and communities across environmental gradients in the Western Himalayas: Method development and conservation use. *Ecological Informatics, 14*, 99-103. http://dx.doi.org/10.1016/j.ecoinf.2012.11.010

Khan, S. M., Page, S., Ahmad, H., Ullah, Z., Shaheen, H., Ahmad, M., & Harper, D. M. (2013b). Phyto-climatic gradient of vegetation and habitat specificity in the high elevation Western Himalayas. *Pakistan Journal of Botany, 45*, 223-230.

Knollová, L., Chytrý, M., Tichy, L., & Hajek, O. (2005). Stratified resampling of phytosociological databases: Some strategies for obtaining more representative data sets for classification studies. *Journal of Vegetation Science, 16*(4), 479-486. http://dx.doi.org/10.1111/j.1654-1103.2005.tb02388.x

Koch, W. (1926). Die Vegetationseinheiten der Linthebene. *Jahrbush der St. Gallischen Naturwissenschaftlichen Gesellschaft, 61*, 1-146.

Kočí, M., Chytrý, M., & Tichý. L. (2003). Formalized reproduction of an expert-based phytosociological classification: A case study of subalpine tall-forb vegetation. *Journal of Vegetation Science, 14*(4), 601-610. http://dx.doi.org/10.1111/j.1654-1103.2003.tb02187.x

Mirkin, B. M. (1989). Plant taxonomy and syntaxonomy: A comparative analysis. *Vegetatio, 82*(1), 35-40. http://dx.doi.org/10.1007/BF00217980

Moravec, J. (1989). Influences of the individualistic concept of vegetation on syntaxonomy. *Vegetatio, 81*(1-2), 29-39. http://dx.doi.org/10.1007/BF00045511

Murtaugh, P. A. (1996). The Statistical Evaluation of Ecological Indicators. *Ecological Applications, 6*(1), 132-139. http://dx.doi.org/10.2307/2269559

Nesom, G. L. (1991). Taxonomy of Isocoma (Compositae: Astereae). *Phytologia, 70(2)*, 69-114.

Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. *Bulletin of the Japanese Society of Science and Fisheries, 22*.

Peinado, M., Aguirre, J. L., Delgadillo, J., & Macías, M. Á. (2008). A phytosociological and phytogeographical survey of the coastal vegetation of western North America. Part I: Plant communities of Baja California, Mexico. *Plant Ecology, 196*(1), 27-60. http://dx.doi.org/10.1007/s11258-007-9334-5

Peinado, M., Aguirre, J. L., Macías, M. Á., & Delgadillo, J. (2011b). A phytosociological survey of the dune forests of the Pacific Northwest. *Plant Biosystems, 145*(1), 105-117. http://dx.doi.org/10.1080/11263504.2011.602741

Peinado, M., Ocaña-Peinado, F. M., Aguirre, J. L., Delgadillo, J., & Díaz-Santiago, G. (2011a). A phytosociological and phytogeographical survey of the coastal vegetation of western North America: Beach and dune vegetation from Baja California to Alaska. *Applied Vegetation Science, 14*.

Petřík, P., & Bruelheide, H. (2006). Species groups can be transferred across different scales. *Journal of Biogeography, 33*(9), 1628-1642. http://dx.doi.org/10.1111/j.1365-2699.2006.01514.x

Pignatti, S. (1980). Reflections on the phytosociological approach and the epistemological basis of vegetation science. *Vegetatio, 42*(1-3), 181-185. http://dx.doi.org/10.1007/BF00048885

Poore, M. E. D. (1955). The use of phytosociological methods in ecological investigations: 1. The Braun-Blanquet system. *Journal of Ecology, 43*(1), 226-244. http://dx.doi.org/10.2307/2257132

Sokal, R. R., & Rohlf, E. J. (1995). *Biometry. The principles and practice of statistics in biological research* (3rd ed.). New York, NY: Freeman.

Stuessy, T. F. (1997). Classification: More than just branching patterns of evolution. *Aliso, 15*, 113-124.

Szafer, W., & Pawlowski, B. (1927). Die Pfanzenassoziationen des Tatra-Gebirges. Bemerkungen über die angewandte Arbeitstechnik. *Bulletin International de Academie Polonaise des Sciences et des Lettres, 3*, 1-12.

Tichý, L. (2002). JUICE, software for vegetation classification. *Journal of Vegetation Science, 13*(3), 451-453. http://dx.doi.org/10.1111/j.1654-1103.2002.tb02069.x

Tichý, L., & Chytrý, M. (2006). Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science, 17*.

Westhoff, V., & van der Maarel, E. (1973). The Braun-Blanquet approach. In R. H. Whittaker (Ed.). *Classification and Ordination of Communities* (pp. 619-659). The Hague, NE: Dr. W Junk.

Wiggins, I. L. (1980). *Flora of Baja California*. Stanford, CA: Stanford University Press.

Willner, W. (2006). The association concept revisited. *Phytocoenologia, 36*(1), 67-76. http://dx.doi.org/10.1127/0340-269X/2006/0036-0067

# Call for Manuscripts

*Modern Applied Science (MAS)* is an international, double-blind peer-reviewed, open-access journal, published by the Canadian Center of Science and Education. It publishes original research, applied, and educational articles in all areas of applied science. It provides an academic platform for professionals and researchers to contribute innovative work in the field. The scopes of the journal include, but are not limited to, the following fields: agricultural and biological engineering, applied mathematics and statistics, applied physics and engineering, chemistry and materials sciences, civil engineering and architecture, computer and information sciences, energy, environmental science and engineering, mechanics. The journal is published in both print and online versions. The online version is free access and download.

We are seeking submissions for forthcoming issues. All manuscripts should be written in English. Manuscripts from 3000–8000 words in length are preferred. All manuscripts should be prepared in MS-Word format, and submitted online, or sent to: mas@ccsenet.org

**Paper Selection and Publishing Process**

a) Upon receipt of a submission, the editor sends an e-mail of confirmation to the submission's author within one to three working days. If you fail to receive this confirmation, your submission e-mail may have been missed.

b) Peer review. We use a double-blind system for peer review; both reviewers' and authors' identities remain anonymous. The paper will be reviewed by at least two experts: one editorial staff member and at least one external reviewer. The review process may take two to four weeks.

c) Notification of the result of review by e-mail.

d) If the submission is accepted, the authors revise paper and pay the publication fee.

e) After publication, the corresponding author will receive two hard copies of the journal, free of charge. If you want to keep more copies, please contact the editor before making an order.

f) A PDF version of the journal is available for download on the journal's website, free of charge.

**Requirements and Copyrights**

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the authorities responsible where the work was carried out, and that, if accepted, the article will not be published elsewhere in the same form, in English or in any other language, without the written consent of the publisher. The editors reserve the right to edit or otherwise alter all contributions, but authors will receive proofs for approval before publication.

Copyrights for articles are retained by the authors, with first publication rights granted to the journal. The journal/publisher is not responsible for subsequent uses of the work. It is the author's responsibility to bring an infringement action if so desired by the author.

**More Information**

E-mail: mas@ccsenet.org

Website: www.ccsenet.org/mas

Paper Submission Guide: www.ccsenet.org/submission

Recruitment for Reviewers: www.ccsenet.org/reviewer

The journal is peer-reviewed
The journal is open-access to the full text
The journal is included in:

CABI
Chemical Abstracts database
DOAJ
EBSCOhost
Excellence in Research Australia (ERA)
Google Scholar
LOCKSS
Open J-Gate
ProQuest
Scopus
Standard Periodical Directory
Ulrich's
Universe Digital Library

Modern Applied Science
Monthly

Publisher Canadian Center of Science and Education
Address 1120 Finch Avenue West, Suite 701-309, Toronto, ON., M3J 3H7, Canada
Telephone 1-416-642-2606
Fax 1-416-642-2608
E-mail mas@ccsenet.org
Website www.ccsenet.org/mas

ISSN 1913-1844