IDŐJÁRÁS

QUARTERLY JOURNAL
OF THE HUNGARIAN METEOROLOGICAL SERVICE

CONTENTS

****

http://www.met.hu/firat/ido-e.html

VOL. 102 * NO. 1 * JANUARY–MARCH 1998

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service

Editorial Office: P.O. Box 39, H-1675 Budapest, Hungary or
Gilice tér 39, H-1181 Budapest, Hungary
E-mail: gmajor@met.hu or *antal@met.hu*
Fax: (36-1) 290-7387

Subscription by

mail: IDŐJÁRÁS, P.O. Box 39, H-1675 Budapest, Hungary;
E-mail: gmajor@met.hu or *antal@met.hu*; *Fax: (36-1) 290-7387*

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 102, No. 1, January–March 1998, pp. 1–18

Generation, detection and granulometry of nanoparticles in the air

M. B. Attoui[1], **A. Renoux**[1] and **D. Boulaud**[2]

[1]*Laboratoire de Physique des Aérosols et de Transfert des Contaminations, Université Paris XII, Avenue du Général de Gaulle, 94010 Creteil, France*

[2]*IPSN CE Saclay 91191 Gif S/Yvette, France*

(Manuscript received 16 June 1997; in final form 10 November 1997)

Abstract—Nanoaerosols are present in the atmosphere in high concentrations. They are formed from the gas phase by homogeneous or heterogeneous condensation. Natural radioactive aerosol (radon daughter elements) can also be of nanometric or even sub-nanometric size. Nanometric aerosols are also found in some industrial conditions such as: furnaces, combustion, chemical reactors for synthetic materials, nanostructures, ceramics, etc. In this paper we review the nanometric aerosol generators for polydisperse aerosols based on spraying and evaporation-condensation. We also consider electric selection for monodisperse generation (with the Differential Mobility Analyzer and the Spectomètre de Mobilité Electrique Circulaire). We examine the detection of this aerosol using the *u*ltrafine *c*ondensation *n*ucleus *c*ounter (UCNC) and its sizing with the *d*ifferential *m*obility *a*nalyzer (DMA) or diffusion battery. Finally we review another device for determining the diameter of nanometric and subnanometric aerosols: the supersonic impactor.

Key-words: nanoparticles, aerosols, granulometry.

1. Introduction

Nanometric aerosols are increasingly studied by specialists both in aerosol science and nanostructure technology. In fact, it is an area where a number of different domains overlap: the vast field of aerosol mechanics (studied by aerosol physicists for more than a quarter of a century), solid-state physics and chemistry at nanometric level. However, these aerosols have been known and studied since the last century (*Coulier*, 1875; *Aitken*, 1888). Aitken was the first to measure the concentration of condensation nuclei in air (sometimes called Aitken nuclei) in 1910. In 1951 Pollack and Murphy confirmed his results. *O'Connor et al.* (1961) demonstrated that the majority of these nuclei,

resulting from gas-particle conversion, have a diameter of less than 11 nm and are electrically neutral. The aerosol produced by algae left by the sea (low tide) is also nanometric and present in high concentration in the air, more than 10^6 cm^{-3} according to the work of *Paugam* (1978).

On the other hand, the sizes of natural radioactive aerosol particles (radon daughter products) in the air in a free state vary from a few angstroms to a few nanometers (*Malet et al.,* 1996a and b; *Hopke*, 1990). Natural radioactive aerosol conforms to the laws of aerosol physics, of course, just as non-radioactive aerosols and, in addition, can be detected using techniques for radioactivity (*Renoux*, 1996; *Kerouanton et al.*, 1996; *Mesbah et al.*, 1996; *Olawin et al.*, 1995). Gas particle conversion in sea air over continental coasts gives rise to high nanometric aerosol concentration (*Tymen et al.*, 1995). Nanometric aerosols are also found in many other conditions (combustion process, chemical reactors, electric ovens, electrical discharges, etc.).

Due to its size this aerosol is, of course, greatly influenced by the activity of the carrier gas molecules. It is therefore subject to Brownian movement. It obeys Einstein's law which expresses its average quadratic movement in terms of time and its diffusion coefficient, the latter dependent upon the size of the aerosol (*Malet et al.,* 1996b).

In this article we will pass in review the means of generating this aerosol along with techniques for measuring its size and its concentration in the air.

2. Detection of the nanometric aerosol

To detect this aerosol we use either its capacity to "grow" to a size which is relatively easy to exploit or its capacity to carry electric charges producing a measurable electric current. There are, therefore, two particle concentration measuring devices based on these two properties which can detect nanometric aerosols. The condensation nucleus counter enlarges the aerosol before detecting it optically. The electrometer measures an electric current (weak in general) induced by an air flow, the aerosols having previously been electrically charged.

2.1 The condensation nucleus counter (CNC)

The condensation nucleus counter is the device used for detecting fine, ultrafine and nanometric particles. It can "see" particles down to one nanometer in diameter. This device works according to the same principle as the first counter (*Aitken*, 1888, 1891), i.e. the particles are enlarged by condensable vapor before counting. Aitken used water vapor to form water droplets by condensation on each of the nuclei, which he then counted individually and visually.

The Pollak counter (*Nollan* and *Pollak*, 1946) was the first photoelectric counter based on the attenuation of a light ray by droplets, again, of water. These counters, along with the commercial versions which followed, use adiabatic expansion to make water vapor condense on the particles. *Bricard et al.* (1972, 1974, 1976) in France, *Sinclair* and *Hoops* (1975) in the United States propose a so-called "continuous flow" counter using butyl vapor to enlarge the aerosol to be detected. The air-borne aerosols, introduced above a tank of butanol, enter a heated horizontal tube to be saturated in vapor and then arrive in a second tube maintained at a low temperature. This tube, called a condenser, is where the butanol condenses on the aerosol particles. The droplets formed in this way, in the ten to twelve microns range (*Agarwal et al.*, 1977), are detected by an optical system. This device described and calibrated by *Agarwal* and *Sem* (1980) has been marketed by the American firm TSI (Thermo System Incorporation) under the name of CNC 3020. Its counting efficiency is only 50% for diameters of 10 nm with a sampling rate of 0.3 liter per minute. TSI has introduced two counting systems into this device (photometric and individual) which allow the measurement of concentration between 0.01 to 10^7 particles per cubic centimeter. A few years later the addition of an optical laser and a bypass loop on the air circuit allowed TSI to reduce the detectable diameter to 7 nm and increase the sampling rate to 1.4 liter per minute. The rate for enlarged aerosols remained to be 0.3 liter per minute. In fact the CNC only gives an integral measure of the aerosol. It does not make it possible to discover the initial diameter of the nucleus. *Nolan* (1972) and *Liu et al.* (1975) review the different types of CNC and their development.

Scientific research in this field of metrology (optical detection after enlargement), branched out into two main directions following the development of the CNC. In the United States it focused on the detection of finer and finer diameters. In France work was concentrated mainly on increasing the rate of sampling. Research into diameter led to devices becoming available on the market. For industrial reasons, commercial or otherwise, high yield devices are not currently available, although a patent has been filed and sold by the CEA.

2.1.1 Ultrafine CNC

The so-called "ultrafine" CNC (TSI Model 3025) has a counting efficacy of more than 50% for 3 nm aerosols. It can detect even smaller diameters but with an efficiency which is much less reliable. This device is based on the work on heterogeneous condensation in a refrigerated pipe by *Metayer et al.* (1982), *Metayer* (1982), *Liu et al.* (1982), *Wilson et al.* (1983), *Stolzenburg* and *McMurry* (1984), *Bartz et al.* (1985), *Keady et al.* (1988) and *Stolzenberg* (1988). The operating principle is practically the same (*Fig. 1*), apart from the condensation. The aerosol is injected into the centre of the condenser which is,

in this case, coaxial. Filtered air saturated with butanol vapor accompanies the aerosol through the condenser to reduce losses by diffusion through the walls. Also to reduce losses, the aerosol does not go right through the saturator. The main part of the sampling (1.4 liter per minute of air-borne aerosols) is filtered, dried and then passed into the saturator before joining the rest of the sampling (0.3 liter per minute) which has not been filtered and which has passed through a capillary tube only a few centimeters in length. In this way the losses by diffusion of the ultrafine aerosol is reduced to a minimum and finer and finer nuclei are activated. This CNC (TSI Model 3025), known as the ultrafine CNC (UCNC), is limited to concentrations of 10^5 particles per cubic centimeter to avoid coagulation.

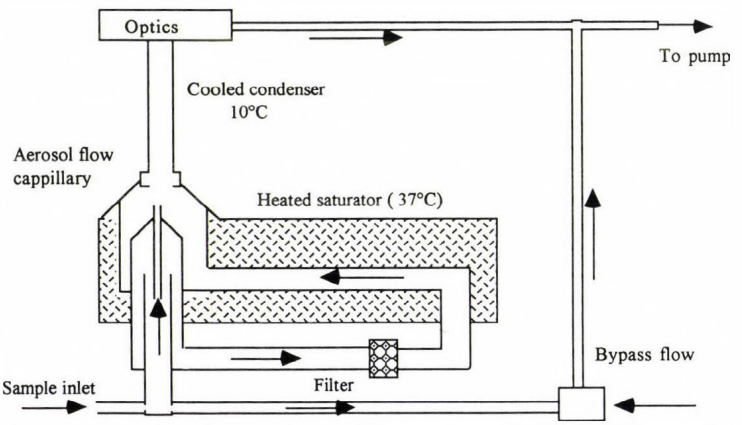


Fig. 1. Schematic cross section of the continuous-flow ultrafine CNC (UCNC).

2.1.2 High-flow CNC

French research on the CNC (CEA - Paris XII University) has investigated the problem of yield. Continuing the work of *Metayer* (1982), *Assa Achy* (1987) developed a high-flow CNC. He designed a new CNC (still based on the principle of continuous flow) capable of detecting aerosols of 10 nm with an efficacy higher than 50% and extracting 28.4 liters per minute! It should be recalled that standard TSI CNCs do not exceed 1.4 liter per minute. The CEA received a patent for this device bought by an American company which has never put it on the market. On the other hand, the possibility of distinguishing nuclei from the produced droplets has been investigated. Indeed, according to *Stolzenburg* (1988) and *Ahn* (1990) the size of the nucleus depends on the size of the final droplet. *Rebours et al.* (1996), *Rebours* (1994) continued the work

of *Metayer* (1982) and *Assa Achy* (1987) and found, using a system of high yield particle enlargement, an indisputable relation of dependence between the size of the drops and the size of the nuclei when the latter was between 4 and 20 nm, making it feasible to use a high flow nanometric granulometer. This device has an activation efficiency of 100% for nuclei of 4 nm.

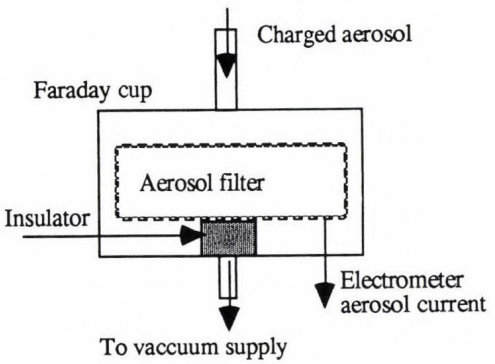


Fig. 2. Schematic cross section of the aerosol electrometer.

3.3 Electrometer

The electrometer, (*Fig. 2*) is a device composed of a Faraday cup and an ammeter capable of measuring very weak currents (in the order of 10^{-15} A). Thanks to this device the volume concentration of an aerosol can be measured if the electric charge carried by each particle is known (*Lui* and *Pui*, 1974; *Sem,* 1975). Moreover the capture or detection efficacy of the electrometer does not depend on the size of the aerosol particles (*Winklmeyer et al.,* 1991; *Wiedensohler et al.,* 1994). In this device the air containing the electrically charged particles enters the Faraday cage to cross a filter with a very high efficiency. The filter placed under a wire mesh traps the particles whilst their electrical charges are recuperated by the mesh and sent to an electrical detector (electrometer). As in case of the CNC, the electrometer gives the concentration of the aerosol but not the size of the particles. The measured current i is a function of the volume output Qv, of the aerosol entering the cup, of the number of charges n carried by each particle, of the elementary charge e and of the concentration of particles C. The current is given by the following relation:

$$i = C\, n\, e\, Q_v.$$

3. Generation of nanometric aerosols

There are several types of generators for polydisperse or quasi-monodisperse nanometric aerosols, for instance the heated wire generator (*O'Connor* and *Roddy*, 1966), exploded wire generator (*Phalen*, 1972), electric arc generator (*Boffa* and *Pfender*, 1973) and electron torch generator (*Tarroni et al.*, 1974). They all, however, have a major flaw: unstable generation over time. Only generators using homogeneous condensation and "electrospraying" can give a stable nanometric aerosol (*Fig. 3*).

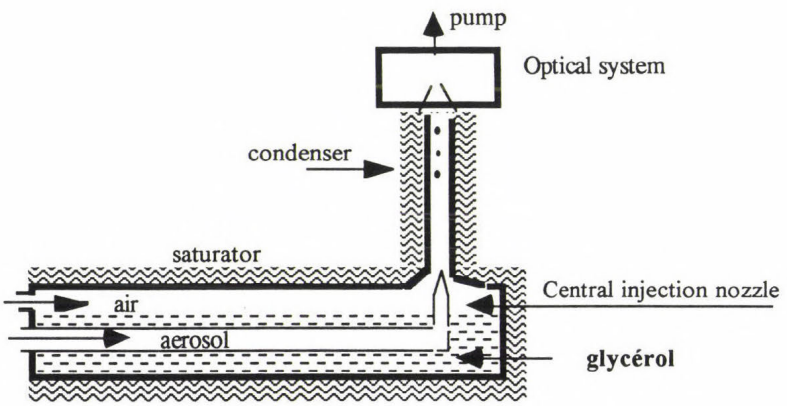


Fig. 3. Schematic cross section of the particle growth system.

3.1 Generation by evaporation condensation

Various publications deal with different examples of this type of generator. They are described by *Spurny* and *Hampl* (1965), *Spurny* and *Lodge* (1968, 1972, 1973), *Spurny et al.* (1980), *Sutugin et al.* (1970/1971) and *Dousaka et al.* (1982). They all use a furnace to vaporize a substance which, by homogeneous condensation, produces a highly concentrated polydisperse aerosol, ($\sigma_g \approx 2.5$). We will consider the devices giving the best σ_g.

To obtain high concentrations of nanometric aerosols, *Scheible* and *Porstendorfer* (1983) suggest to vaporize a substance (NaCl or Ag) in a tube furnace under a nitrogen flow. The vapor-nitrogen mixture is cooled to ambient temperature when leaving the furnace into the extension of the tube (*Fig. 4*). The aerosol obtained is not very polydisperse ($\sigma_g \approx 1.3$ to 2.0). The aerosol volume concentration exceeds 10^6 particles per cm^3. The largest diameters are obtained for the high temperatures (12 nm at 1300°C for Ag, 825°C for NaCl

and 2 nm at 1065°C for Ag, 600°C for NaCl). This generator produces an aerosol where the sizes of the particles become unstable after roughly an hour and the user has no way of influencing its concentration.

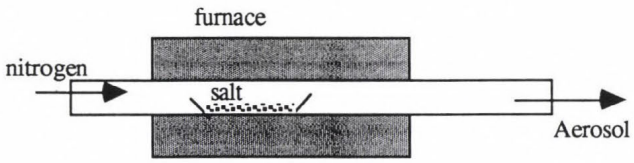


Fig. 4. Nanometric generator.

Bartz et al. (1987) suggest a slightly different nanometric generator based on the one used by *Liu* and *Levy* (1980) to produce sulfuric acid aerosols (*Fig. 5*). They pulverize an aerosol dissolved in a liquid (NaCl in water for instance) in a standard Collison atomizer before drying it and sending it into a tube furnace to be vaporized. The vapors leaving the furnace enter a mixing section through a nozzle where they are accelerated and mixed with cold, filtered, compressed air. The produced aerosol passes into a cooling section where it is brought down to the ambient temperature. The cooling and mixing sections allow improved control of the condensation of the vapors leaving the furnace. The supply of filtered compressed air prevents, by dilution, the coagulation of the aerosol. This generator thus allows an aerosol stability for more than 12 hours. The final size of the nanometric aerosol depends on how the temperature of the furnace, the rate of dilution and the concentration of the atomized solution are regulated. This makes it possible to reduce even more the σ_g of the nanometric aerosol obtained.

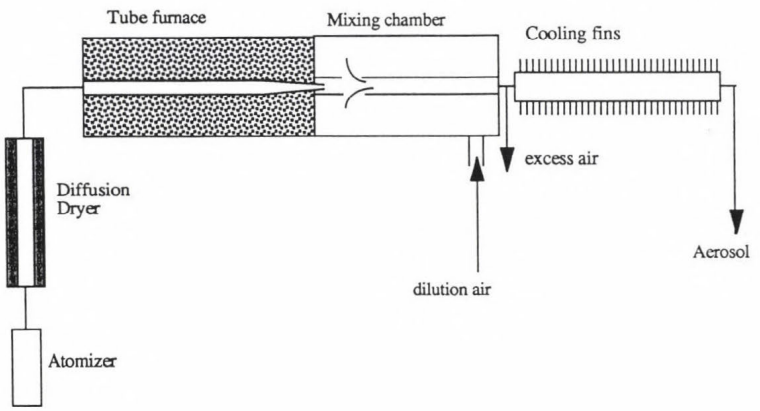


Fig. 5. Nanometric generator of Bartz.

3.2 Electrohydrodynamic pulverization (electrospraying)

Electrospraying, a phenomenon observed by *Zeleny* (1915, 1917) is based on the capacity of an electric field in the ambient air to disperse a liquid. Droplets charged by electrospraying are produced by applying a difference in potential of several kilovolts between a plate and the exit of a capillary tube containing the liquid, as we can see in *Fig. 6* below. The droplets formed are electrically charged. Their size varies, according to different conditions, from several millimetres to several tenths of microns (*Cloupeau*, 1994).

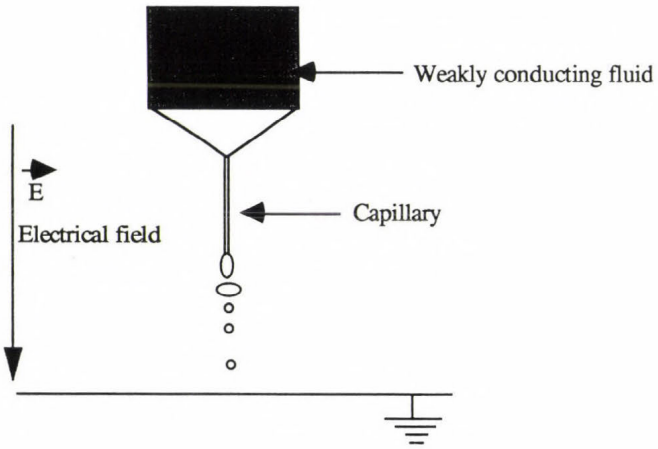


Fig. 6. Schematic cross section of electrospraying.

Chen et al. (1995) suggest a generator (*Fig. 7*) for particles from 4 nm to 1.2 μm using the electrospraying of liquid conductors. This work initiated by *Vonnegut* and *Neubauer* (1952) has produced a fine aerosol generator (*Grace* and *Mareijnissen*, 1994; *Loscertals* and *Fernandez de la Morra*, 1994). The electrospraying phenomenon is explained in detail by *Cloupeau* and *Prunet-Foch* (1989, 1990, 1994). In addition, the magazine, "Journal of Aerosol Science", devoted a special issue (Vol. 25, No. 6, 1994) to this phenomenon.

Finally, photochemical generators (*Dubstov* and *Baklanov*, 1996; *Dubstov et al.*, 1996, 1995) produce relatively monodisperse aerosols ranging in size from several angstroms to a dozen nanometers. Clusters are obtained in a reactor by exposure to radiation from a high pressure mercury lamp before enlargement in vapors supersaturated with dibutyl phthalate.

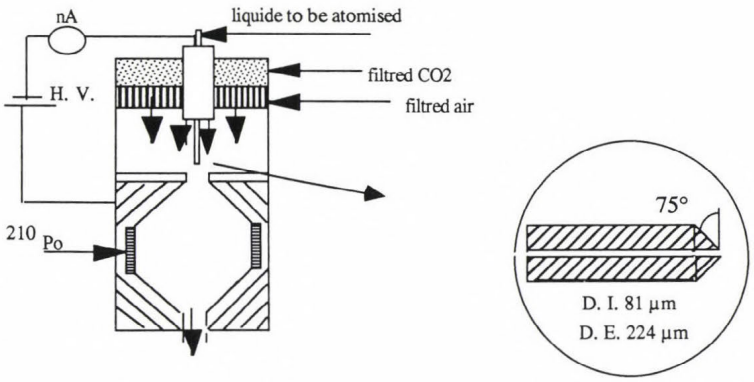


Fig. 7. Chen's generator (electrospraying).

4. Generation of monodisperse nanometric aerosols

None of the above mentioned generators produce a rigorously monodisperse aerosol. A calibrated nanometric aerosol is achieved by electrostatic classification using the electrical mobility of electrically charged aerosols placed in an electric field. There are two devices which make this possible, the DMA and the Radial DMA.

4.1 The differential mobility analyzer (DMA)

To obtain nanometric monodisperse aerosols with the generators described above the electrical properties of the aerosols are exploited in the differential mobility analyzer (*Zeleny*, 1915; *Hewit*, 1957; *Knutson* and *Whitby*, 1975; *Knutson*, 1976). This device generates particles of up to 7 nm. Size accuracy can be as much as 2% (*Liu* and *Pui*, 1974).

The DMA is composed of two coaxial cylinders between which a horizontal electric field is set up. The polydisperse aerosol, previously brought to a state of Boltzmann electrical equilibrium, is introduced along the internal walls of the external cylinder. Dried, filtered air circulates along the internal cylinder where an extractor valve is placed. A variable high voltage, between 0 and 12 kV, makes it possible to vary the electric force ($F = qE$) to which the particle penetrating into the selection area is subjected. A second force is exerted on this same particle due to the downward flow of the filtered air along the walls of the internal electrode.

The aerosol should be electrically charged beforehand and the total charge law should be known if we are to have a precise idea of the quantity of atomic charge carried by each of the particles. The Boltzmann charge law is usually

applied in this device (*Bricard*, 1977; *Liu* and *Pui*, 1974b). The global charge is nil in that case, which means that we have as many particles charged positively as negatively. As a result, within the selection area, the neutral particles pass without being diverted, the negative particles are deposited on the internal walls of the external tube, thus only the positively charged particles are trapped or selected. Consequently, there is a drop in the concentration of the selected particles in relation to that of the injected particles. This generator rarely exceeds 10^5 cm^{-3}.

The "short DMA" is used to reduce losses by diffusion in the DMA. It has been specially developed for nanometric aerosols and works on the same principle as the DMA, but is much shorter in length (11.11 cm instead of 44.44 cm) (*Winklmayr et al.*, 1991). It allows nanometric aerosols down to 3 nm to be generated (*Hummes et al.*, 1996).

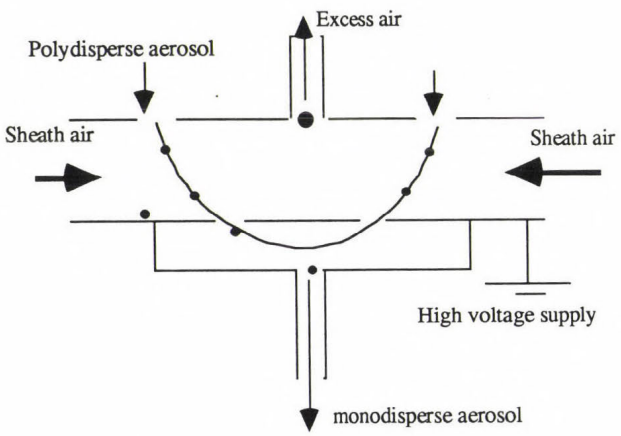


Fig. 8. Schematic cross section of a S.M.E.C.

4.2 The S.M.E.C. (Radial DMA)

This other device has been developed at CEA (*Pourprix et al.*, 1990; *Pourprix* and *Daval,* 1990) and studied by *Mesbah* (1994), *Zhang et al.* (1995), *Boulaud et al.* (1996), *Fissan et al.* (1996), *Lebronnec et al.* (1996). The electromobility spectrometer (named Radial DMA) is composed of two circular plates between which filtered air is introduced. This device with a circular geometry functions, of course, according to the same principle as the DMA (cylindrical geometry). An electric field is set up between the plates by electrifying the lower plate (*Fig. 8*). The polydisperse aerosol, injected through a valve on the upper plate, is extracted with the excess air by a valve on the same plate. The monodisperse aerosol selected is extracted by a nozzle in the lower part of the device on the

second plate. The geometry of this device, the distance between the plates and the length of the selection area all make it much easier to construct, more compact and less cumbersome than the DMA. In addition, there should be fewer losses by diffusion than with other devices of its generation.

5. Granulometry of the nanometric aerosol

After passing in the review the methods for generating and classifying nanometric aerosols, we will now consider the different methods and techniques allowing to determine the size of this aerosol. The electric, aerodynamic (or mechanical) or optical (after enlargement) properties of these aerosols are exploited for this purpose.

5.1 The DMPS/SMPS system

With the introduction of the ultrafine CNC we can now carry out a granulometry in the 5 nm to 1 μm range in an acceptable time thanks to a DMPS system (*D*ifferential *M*obility *P*article *S*ize), composed of an electro-mobility analyzer (DMA), a condensation nuclei counter, a microcomputer (PC) and a program for data inversion (*Keady et al.* 1983). *Knutson* had the idea as early as 1976. The development of the DMPS became possible with the arrival of the microprocessor and especially thanks to a program for data inversion provided by *Fissan et al.* 1983, *Kousaka et al.* (1985), *Hopel* (1978), *Haaf* (1980), *Reischl* (1981), *Alofs* and *Balakumar* (1982) and *ten Brink et al.*(1983) who solved the problems posed by the charge in a bipolar environment and multiple charges. The reaction time was from 2 to 3 min. according to the precision requested. A more powerful and more rapid software in the Windows environment forms, with the previous system (DMA-CNC-PC), what is called the SMPS (*S*canning *M*obility *P*article *S*izer). The SMPS produced by the firm TSI gives the granulometry in 60 or 30 seconds from 5 nm (*Wang* and *Flagan*, 1990). With the arrival of the short DMA we can hope to see, very soon, the detection limit of this system dropping to the 3 nm limit currently imposed by the CNC.

5.2 The diffusion battery

As pointed out in the introduction, nanometric aerosol is subject to Brownian movement. Consequently, the aerosol does not follow the lines of flow of the carrier fluid. In a pipe it tends to diffuse towards the walls and remains fixed there since the adhesion forces are very strong. Thus, there is particle loss on the walls and the aerosol concentration decreases as the air goes along the pipe. This, often undesirable effect is used in the diffusion battery to obtain infor-

mation on the size of an aerosol particle. There are several types of diffusion batteries: parallel plate batteries, cylindric tube batteries, grid batteries, etc. However, only the wire mesh battery is commercially available since it is the only one for which data processing has been developed. In fact, in this type of granulometric measuring the practically unsolvable problem is the restitution of the granulometric spectra which are numerical solutions of a standard unstable Fredholm integral equation. Two types of measures can be made with diffusion batteries: the measure of the penetration or the measure of the deposit on the internal wall of the battery in terms of the capture abscissa. Indeed, using a condensation nucleus counter, the upstream and downstream concentration (and thus the penetration) or the concentration at a given point can be determined.

The grid battery has led to a granulometric measuring device which can measure particles to 2 nm (*Dubstov et al.*, 1995, 1996; *Eremenko et al.*, 1995). Only the multiple-stage wire mesh diffusion battery developed by *Sinclair* and *Hoopes* (1975), *Sinclair et al.* (1976), is produced commercially (TSI). It has ten stages, each one with n grid, n varies according to i, the number of stages, according to the relation: $n = i(i + 1)/2$.

The theory on which this battery is based is explained by *Cheny* and *Yeh* (1980). *Sinclair* (1986) provides an excellent review of the subject. This is the battery which has been studied most over the last years (*Yeh*, 1982; *Yamada et al.*, 1988; *Cheng et al.*, 1980; *Cheng et al.*, 1988; *Schiebel*, 1984; *Wang*, 1993; *Skaptsov*, 1996). At the present time it is the only battery which gives the granulometry of the nanometric aerosol in real time when connected to an ultrafine CNC as shown in *Fig. 9*.

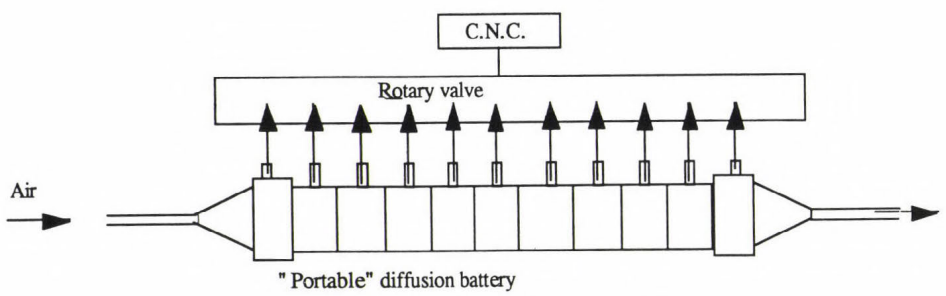


Fig. 9. Schematic diagram of the automatic diffusion battery TSI.

However specific diffusion batteries, for example for radioactive aerosols, developed in laboratories can detect aerosols down to 0.5 nm (*Ramarmurthi et al.*, 1993; *Kerouanton et al.*, 1996).

5.3 The hypersonic impactor

The hypersonic impactor (*Fernandez de la Morra*, 1990a, 1990b) is based on the principle of a shock wave separating the aerosol from its carrier gas. The shock wave is produced sequentially by sending a jet of air at atmospheric pressure through an orifice in a plane plate (infinite) located in a cavity where the pressure is only a few torrs (*Fig. 10*). The distance between the plate and the orifice determines the size of the aerosol to be separated or deposited (*Fig. 11*). The aerosol is detected thanks to an electrometer or, in the case of radon daughter products, by techniques for detecting radioactivity (*Olawoyin et al.*, 1995).

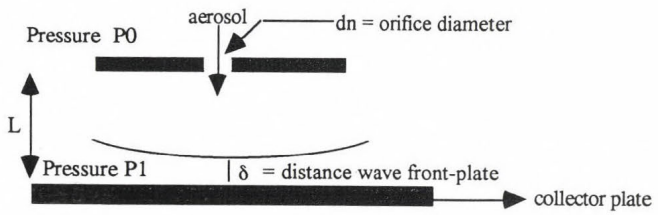


Fig. 10. Shock-wave formation.

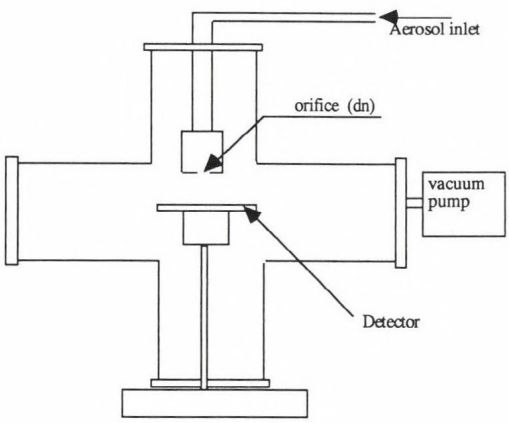


Fig. 11. Sketch of the hypersonic impactor.

For $P_o/P_1 >> 1$ and $d_n << 1$, *Fernandez de la Morra* (1990a) showed that the number of critical Stokes S_o of the flow is given by the relation

$S_o = 0.1983\ \rho_p\ d_p\ C_o^2/(d_n\ P_o)$, where ρ_p is the density of the particle, d_p is its diameter and C_o is the speed of sound in the air.

The capture efficacy of the plate and the diameter of the captured particles depends on L/d_n and P_o/P_1.

6. Conclusion

As explained above, nanometric aerosols first began to interest Aerosol Physicists more than a century ago. However, over recent years, there has been steadily increasing interest, especially in the field of metrology with more and more teams worldwide working on this type of aerosol.

Indeed, to estimate the planet's radiative balance, it is primordial to understand the effect of homogeneous nucleation on the concentration of ultrafine aerosols which act as condensation nuclei for the formation of clouds and precipitation.

For the nanostructure industry, which is developing considerably, the need for information on this aerosol is growing and this can only continue. The importance of this aerosol for the nuclear industry, in the case of an accident in a nuclear reactor, should also be remembered. As we have seen, there is still room for progress in all fields (detection, generation, granulometry, etc.)

Nanometric aerosol physics has not had its day yet — it is far from it. It is faced with new needs and new challenges from fields such as environmental protection, climate evolution, metrology, nanostructure industry, microelectronics, etc.

References

Agarwal, J.K., Sem, J. and *Pourprix, M.*, 1977: A continuous flow CNC capable of counting single particles. *Proc. of Ninth International Conference on Atmospheric Aerosols Condensation and Ice Nuclei*. Galway, Ireland.

Agarwal, J.K. and *Sem, G.J.*, 1980: Continuous flow, single-particle counting condensation nucleus counter. *J. Aerosol Sc. 11*, 343-

Ahn, K., 1988: Ultrafine aerosol measurement and particle activation and droplet growth processes in condensation nucleus counter. Thèse Université du Minnesota.

Aitken, J., 1888/89: On the number of dust particles in the atmosphere. *Proc. R. Soc. Edinburgh 16*, 135-

Aitken, J., 1890/91: On a simple pocket dust-counter. *Proc. R. Soc. Edinburgh 18*, 39-

Alofs, D.J. and *Balakumar, P.*, 1982: Inversion to obtain aerosol size distribution from measurements with a differential mobility analyser. *J. Aerosol Sci. 13*, 513-

Assa Achy, M., 1987: *Mise au point et étalonnage d'un compteur de noyaux de condensation à grand débit*. Thèse de l'Université Paris XII.

Balkanov, A.M., Dubstov, S.N., Caldow, R., Havlicek, M. and *Sem, G.*, 1995: Application of photochemical ultrafine aerosol particles generators for SMPS and DB size resolution. *J. Aerosol Sci. 26*, S751

Bartz, H., Fissan, H. and *Liu, B.Y.H.*, 1987: A new generator for ultrafine aerosols below 10 nm. *Aerosol Sci. Technol. 6,* 163-

Bartz, H., Fissan, H., Helsper, C., Kousaka, Y. Okuyama, K., Fukushima, N., Keady, P.B., Kerrigan,S., Fruin, S.A., McMurry, P.H., Pui, D.Y.H. and *Stolzenburg, M.R.*, 1985: Response characteristics for four different condensation nucleus counters to particles in the 3–50 nm diameter range. *J. Aerosol Sci. 16,* 443-

Boffa, C.V. and *Pfender, E.*, 1973: Controlled generation of monodisperse aerosols in the submicron range. *J. Aerosol Sci. 4,* 103-

Boulaud, D., Pourprix, M., Gougeon, R., Lebronec, E. and *Renoux, A.*, 1996: On the radial flow differential mobility analysers for the determination of aerosol particle density and mass. *J. Aerosol Sci. 27,* S307-

Bricard, J. Madelaine, G., Reiss, P. and *Turpin, P.V.*, 1972: Compteur de noyaux de condensation à flux continu. *C.R. Acad. Sc.* Paris t. 275, 837-

Bricard, J. Delattre, P. and *Madelaine, J.*, 1974: Counting of condensation nuclei at low pressure: its application to photolysis of gaseous impurities in the stratosphere. In *Proc. of the Third Conference on the Climatic Impact Assessment Program. Rep.* DOT-TCS-OST-74-15, US Dept. Transp. Washington, D.C.

Bricard, J. Delattre, P., Madelaine, J. and *Pourprix, M.*, 1976: Detection of the ultra fine particle particles by means of a continuous flux CNC. In *Fine particles* (ed.: *B.Y.H. Liu*). Academic Press, New York.

Bricard, J., 1977: Physique des aérosols. *Rapport CEA R 4831-2.*

Cheng, Y.G., Keating, J.A. and *Kanapilly, G.M.*, 1980: Theory and calibration of a screen-type diffusion battery. *J. Aerosol Sci. 11,* 549-

Cheng, Y.G., Yamada, Y. and *Yeah, H.C.*, 1990: Diffusion deposition on model fibrous filters with intermediate porosity. *Aerosol Sci. Technol. 12,* 286-

Chen, D. R., Pui, D.Y.H. and *Kaufman, S.L.*, 1995: Electrospraying of conducting liquids for monodisperse aerosol generation in the 4 nm to 1,8 μm. *J. Aerosol Sci. 6,* 963-

Cloupeau, M. and *Prunet-Foch, B.*, 1989: Electrostatic spraying of liquids in cone jet mode. *J. Electrostatics 22,* 135-

Cloupeau, M. and *Prunet-Foch, B.*, 1990: Electrostatic spraying of liquids: main functioning modes. *J. Electrostatics 25,* 165-

Cloupeau, M. and *Prunet-Foch, B.*, 1994: Electrohydrodynamic spraying functioning modes: a critical review. *J. Aerosol Sci. 25,* 1021-

Coulier, M., 1875: Note sur une nouvelle propriete de l'air. *J. Pharm. Chimie 22,* 165-

Dubstov, S.N., Balkanov, A.M., Makela, J.M. and *Augustin, J.*, 1996: Performance of the photochemical aerosol generator at 2-10 nm. *J. Aerosol Sci. 27,* S401-

Dubstov, S.N., Koutzenogii, K.P., Levykin, A.N. and *Skubneskaya, G.I.*, 1995: Photochemical aerosol formation of haloidbenzenes — comparaison between theory and experiment. *J. Aerosol Sci. 26,* 705-

Eremenko, S.I., Caldow, R., Baklanov, R. A.M., Havlicek, H. and *Sem, G.*, 1995: Diffusion battery particle sizing system based on the MSA data conversion algorithm: experimental examination 3. *J. Aerosol Sci. 26,* S747-

Fernandez de la Morra, J., Rao, N. and *McMurry, P.H.*, 1990a: Hypersonic impaction of ultrafine particles. *J. Aerosol Sci. 21,* 169-

Fernandez de la Morra, J., Hering, S.V., Rao, N. and *McMurry, P.H.*, 1990b: Inertial impaction of fine particles at moderate Reynolds numbers and the transonic regime with a thin plate orifice nozzle. *J. Aerosol Sci. 21,* 889-

Fissan, H., Hummes, D., Stratmann, F., Buscher, P., Neumann, S., Pui, D.Y.H. and *Chen, D.*, 1996: Experimental comparaison of four differential mobility analysers for nanometer aerosol measurements. *Aerosol Sci. Technol. 24,* 1-

Fissan, H., Helpser, C. and *Thielen, J.H.*, 1983: Determination of particle size distribution by means of an electrostatic classifier. *J. Aerosol Sci. 14,* 354-

Grace, J.M. and *Marijnissen, J.C.M.*, 1994: A review of liquid atomisation by electrical means. *J. Aerosol Sci. 25,* 1005-

Haaf, W., 1980: Construction of a new plate condenser electric mobility analyser and first results. *J. Aerosol Sci. 11*, 201-

Hewitt, G.W., 1957: The charging of small particles for electrostatic precipitation. *Trans. Am. Inst. Elect. Engrs 76,* 300-

Hopke P., 1990: A critical review of measurements of the unattached fraction of radon decay products. *U.S. Department of Energy Office of Energy Research,* No. DE-FGO2-89ER60878.

Hopel, W.A., 1978: Determination of the aerosol size distribution from the mobility distribution of the charged fraction of aerosols. *J. Aerosol Sci. 9*, 41-

Hummes, D., Neumann, S., Fissan, H., Chen, D.R., Pui, D.Y.H., Quant, F.R. and *Sem, G.J.*, 1996: Nanometer differential mobility analyser (nano DMA): experimental evaluation and performance verification. *J. Aerosol Sci. 27,* S135-

Keady, P.B, Denler, V.L., Sem, G.J, Stolzenburg, M.J. and *McMurry, P.H.*, 1988: A condensation nucleus counter designed for ultrafine particle detection above 3 nm. *12th International Conference on Atmospheric Aerosols and Nucleation,* Vienna, Austria.

Keady, P.B., Quant, F.R. and *Sem, G.J.*, 1983: Differential mobility particle sizer: a new instrument for high resolution aerosol size distribution measurement below 1μm. *TSI Quarterly 9*, 3-

Kerouanton, D., Tymen, G. and *Boulaud, D.*, 1996: Small particle diffusion penetration of an annular duct compared to other geometries. *J. Aerosol Sci. 27,* 345-

Knutson, E.O. and *Whitby, K.T.*, 1975: Accurate measurement of aerosol electric mobility moments. J. *Aerosol Sci. 6,* 443-

Knutson, E. O., 1976: Extended electric mobility method for measuring aerosol particle size and concentration. In *Fine Particles* (ed.: *B. Y. H. Liu*). Academic Press, New York.

Kousaka, V., Niida, T., Okuyama, K. and *Tanaka, H.*, 1982: Development of a mixing type condensation nucleus counter. *J. Aerosol Sci. 13,* 231-

Kousaka, V., Okuyama, K. and *Adachi, M.*, 1985: Determination of particle size distribution of ultrafine aerosols using a differential mobility analyser. *Aerosol Sci. Technol. 4,* 209-

Kousaka, V., Okuyama, K., Adachi, M. and *Mimura, T.*, 1986: Effect of Brownian diffusion on electrical classification of ultrafine aerosol particles in differential mobility analyser. *J. Chem. Engng. Jpn. 19,* 401-

Lebronec, E., Boulaud, D., Gougeon, R., M. Pourprix, M. and *Renoux, A.*, 1996: Les analyseurs differentiels de mobilité électrique, une nouvelle méthode pour la détermination de la densité et de la masse des aérosols. 1, 161 *ASFERA*, Paris.

Liu, B.Y.H. and *Pui, D.Y.H.*, 1974a: A submicron aerosol standard and the primary, absolute calibration of the condensation nuclei counter. *J. Colloid and Interface Sci. 47,* 155-

Liu, B.Y.H. and *Pui, D.Y.H.*, 1974b: Equilibrium bipolar charge distribution of aerosols. *J. Colloid and Interface Sci. 49,* 305-

Liu, B.Y.H. Pui, D.Y.H., Hogan, A.W. and *Rich, T.A.*, 1975: Calibration of the Pollak counter with monodisperse aerosols. *J. Appl. Met. 14,* 46-

Liu, B.Y.H., Pui, D.Y.H., McKenzie, R.L., Agarwal, J.K., Jaenicke, R., Pohl, F.G., Preining, O., Reichel, G., Szymansky, W. and *Wagner, P.E.*, 1982: Intercomparison of different absolute instruments for measurment of aerosol number concentration. *J Aerosol Sci. 13*, 429-

Liu, B.Y.H. and *Pui, D.Y.H.*, 1975: On the performance of the electrical aerosols analyzer. *J. Aerosol Sci. 6,* 249-264.

Liu, B.Y.H. and *Levy, J.*, 1980: Generation of submicron sulfuric acid aerosol by vaporisation and condensation. *In Generation of aerosols and Facilities for Exposure Measurements* (ed.: *K. Willeke*). Ann Arbor Science Publishing, Ann Arbor, MI.

Loscertals, I.G. and *Fernandez de la Morra, J.*, 1993: Characterisation of electrospray-generated *of Ultrafine Particles*. Delft, 28-29 May. (ed.: *J.C.M. Marijnissen* and *S. Pratssinis*). Delft University Press, The Netherland.

Malet, J., Montassier, N., Boulaud, D. and *Renoux, A.*, 1996a: Diffusion of aerosols with in flight formation in laminar tube flow. *J. Aerosol Sci. 27,* S261-

Malet, J., Montassier, N., Boulaud, D. and *Renoux, A.*, 1996b: Détermination expérimentale du coefficient de diffusion du ^{218}Po. *Actes ASFERA*, Paris.

Mesbah, B., 1994: *Le spéctromètre de mobilité électrique circulaire; performances et applications.* Thèse de l'Université Paris XII, France.

Mesbah, B., Fitzgerald, B., Hopke, P. and *Pourprix, M.*, 1996: Mobilité des particules ultrafines radioactives; nouvelle technique de mesure. *Actes de l'ASFERA*, Paris.

Metayer, Y., 1982: *Techniques de mesure de l'aérosol fin.* Thèse de l'Université Paris XII.

Metayer, Y., Perrin, M.L. and *Madelaine, G.J.*, 1982: Analysis of a continuous flow condensation nuclei counters. *J. Aerosol Sci. 13*, 170-

Nolan, P.J. and *Pollak, L.W.*, 1946: *Proc. R. Irish. Acad.*, Dublin 51 A2, 9-

Nolan, P.J., 1972: The photoelectric nucleus counter. *Sci. Proc. R. Dublin Soc. A4*, 161-

O'Connor, T.C. and *Roddy, A.F.*, 1966: *J. Rech. Atmos. 2*, 239-

O'Connor, T.C., Sharkey, W.P. and *Flanagan, V.P.*, 1961: Observations on the Aitken nuclei in Atlantic air. *Quart. J. Roy. Meteorol. Soc. 87*, 105-

Olawoyin, O.O., Raunemaa, T.M. and *Hopke, P.K.*, 1995: A system for aerodynamically sizing ultrafine radioactive particles. *Aerosol Sci. Technol. 23*, 121-

Paugam, 1978: *Etude des principaux paramètres électriques de l'air, sur un site côtier, en atmosphère peu polluée.* Thèse de l'Université de Bretagne Occidentale de Brest.

Phalen, R.F., 1972: Evaluation of an exploded-wire aerosol generator for use in inhalation studies. *J. Aerosol Sci. 3*, 395-

Pourprix, M., Daval, J. and *Berne, P.*, 1990: Device permitting the controlled deposit of fine particles on surfaces. *AAAR 90 Meeting*, Philadelphia, PA, June 18-22.

Pourprix, M. and *Daval, P.*, 1990: *Electrostatic Precipitation of Aerosols on Wafers, a Mobility Spectrometer* (eds.: *S. Masuda* and *K. Takahashi*). Pergamon Press, Vol. 2.

Rebours, A., 1994: *Etude de la détection et de la granulométrie en temps réel de l'aérosol ultrafin par un système de grossissement de particules.* Thèse de l'Université Paris XII.

Rebours, A., Boulaud, D. and *Renoux, A.*, 1996: Recent advances in nanoparticle size measurement with a particle growth system combined with an optical counter- a feasibility study. *J. Aerosol Sci. 27*, 1227-

Reischl, G.P., 1981: On the equivalence of low pressure impactor and differential mobility analyser data for submicron particle size distributions. *Proc. of 9th Annual Conference of GESELLSCHAFT Fur Aerosolforschung.* Duisburg, West Germany.

Renoux, A., 1996: Radon and natural indoor radioactive aerosol. *Időjárás 100,* 89-105.

Scheibel, H.G. and *Porstendorfer, J.*, 1983: Generation of monodisperse Ag- NaCl- aerosols with particle diameters betwen 2 and 300 nm. *J. Aerosol Sci. 14*, 113-

Sem, G.J., 1975: Design an application of an electrical size analyser for submicron aerosol particles. Analysis Instrumentation Vol. 13. *Proc. of the 21st Annual ISA Analysis instrumentation Symposium.* Instrument Society of America, 400 Stanwix St., Pittsburgh, Pa. 15222.

Skapsov, AZ.S., Baklanov, A.M., Dubtsov, S.N., Laulainen, N.S., Sem, G. and *Kaufman, S.*, 1996: Measurement of ultrafine aerosol particle size below 15 nm by a diffusion battery. *J. Aerosol Sci. 27 S,* 167-

Sinclair D. and *Hoops, G.S.*, 1975: A continuous flow condensation nucleus counter. *J. Aerosol Sci.* 6, 1-

Sinclair, D., 1986: Measurement of nanometer aerosols. *Aerosol Sci. Technol. 5*, 187-

Spurny, K. and *Hample, V.*, 1965: Preparation of radioactive labeled condensation aerosols: I. aerosols of sodium chloride, silver iodide and sulhuric acid. *Coll. Czech. Chem. Communs 30,* 507-

Spurny, K. and *Lodge, J.P.*, 1968: Aerosol filtration by means of nuclepore filters. Atmos. Environ. 2, 429-

Spurny, K. and *Lodge, J.P.*, 1972: Preparation of radioactive labeled condensation aerosols -V-. *A.J.H.A.J. 33*, 431-

Spurny, K. and *Lodge, J.P.*, 1973: Herstellugen hochdisperser modellaerosole fur storbfurschung and filter-prufung. *Staub-Reinalt. Luft. 33*, 166-

Spurny, K., Opiela, H., Weiss, G. and *Lodge, J. P.*, 1980: On the preparation of highly dispersed radioactively labeled condensation aerosols of silver and silver compounds. *Atmos. Environ. 14,* 871-

Stolzenburg, M.R. and *McMurry, P.H.*, 1984: A theoretical model for ultra-fine aerosol condensation nucleus counter. In *Aerosol* (eds.: *Liu, B.Y.H., Pui, D.Y.H.* and *Fissan, H.J.*). Elseiver New York.

Stolzenburg, M.R., 1988: *An Ultrafine Aerosol Size Distribution Measuring System.* Ph.D. Thesis University of Minnesota.

Sutugin, A.G. and *Fuchs, N.A.*, 1970: Formation of condensation aerosols under rapidly changing environmental conditions — theory and method of calculation. *J. Aerosol Sci. 1,* 287-

Sutugin, A.G., Fuchs, N.A. and *Kostev, E.I.*, 1971: Formation of condensation aerosols under rapidly changing environmental conditions - non coagulated, high dispersed aerosols. *J. Aerosol Sci. 2,* 361-

Taroni, G., Prodi,V., Melandri, L., Bompane,G.F., Zaiacomo, T. and *Formignani, M.*, 1974: Production of ultrafine monodisperse aerosols by condensation. *Proc. of the GAef-meeting*, Bad Sooden, Germany.

ten Brink, H.M., Plomp, A., Spoelstra, H. and *de Vate J.F.*, 1983: A high resolution electrical mobility aerosol spectrometer. *J. Aerosol Sci. 14,* 589-

Tymen, G., Le Bihan, O. and *Droal C.*, 1995: Evolution dynamique de la distribution en taille des particules atmosph;riques ultrafines à la station cotière de Porspoder. 20, *Actes ASFERA,* Paris.

Vonnegut, B. and *Neubauer, R.L.*, 1952: Production of monodisperse liquid particles by electrical atomisation. *J. Colloid Sci. 7,* 616-

Wang, S.C. and *Flagan, R.C.*, 1990: Scanning electrical mobility spectrometer. *Aerosol Sci. Technol. 13,* 230-

Wang, H.C., 1993: Thermal rebound of nanometer particles in a diffusion battery. *Aerosol Sci. Technol. 18,* 180-

Wiedensohler, A., Aalto, P., Covert, D., Heintzenberg, J. and *McMurry, P.H.*, 1994: Intercomparison of four methods to determine size distributions of low concentration ultafine aerosols with illustrative data from arctic. *Aerosol Sci. Technol. 21,* 95-

Wilson, J.C., Hyun, J.H. and *Blackshear E.D.*, 1983: The function and response of an improved stratospheric condensation nucleus counter. *J. Geophys. Res. 88,* 6781-

Winklmayer, W., Reischeil, A.O., Lindner, A.O. and *Berner, A.*, 1991: A new electromobility spectrometer for the measurement of aerosol size distributions in the size range from 1 to 1 000 nm. *J. Aerosol Sci. 22,* 289-

Yamada, Y., Cheng, Y.S. and *Orman, M.M.*, 1988: Evaluation of a coarse screens a diffusion cell material for ultrafine aerosols below 0,02 μm. *J. Aerosol Sci. 19,* 733-

Yeah, H.C., Cheng, Y.S. and *Orman, M.M.*, 1982: Evaluation of various types of wire screens as diffusion battery cells. *J. Colloid Interface Sci. 86,* 12-

Zhang, S.H., Akutsu, Y., Russell, L.M., Flagan, R.C. and *Seinfeld, J.H.*, 1995: Radial differential mobility analyser. *Aerosol Sci. Technol. 23,* 357-

Zeleny, J., 1915: On the conditions of instability of electrified drops, with applications to the electrical discharge from liquid points. *Proc. Camb. Phil. Soc. 18,* 71-

Zeleny, J., 1917: Instability of electrified surfaces. *Phys. Rev. 10,* 1-

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 102, No. 1, January–March 1998, pp. 19–41

On the origin of model errors. Part I. Effects of the temporal discretization for Hamiltonian systems

Barbara Kádár, István Szunyogh[1]

Department of Meteorology, Eötvös Loránd University,
H-1083 Budapest, Ludovika tér 2, Hungary
E-mail: kadar@nimbus.elte.hu

and **Dezső Dévényi**

NOAA/ERL/Forecast Systems Laboratory,
325 Broadway, Boulder CO, USA; E-mail: devenyi@fsl.noaa.gov
also affiliated with
Cooperative Institute for Research in Environmental Sciences,
University of Colorado at Boulder, Boulder, CO, USA

(Manuscript received 23 September 1997; in final form 21 November 1997)

Abstract—There are two main sources of errors in numerical modeling of the atmosphere: the errors of initial conditions and that of the models themselves. The algebraic structure of equations is strongly related to their integrability, thus the problem of sensitivity to initial data cannot be handled separately from that of model formulation (i.e. of model errors). In a nonintegrable system the utilization of a numerical solution algorithm is a must. The present review paper (in two parts) deals with model errors originating from the inevitable discretization of the continuous equations in space and time. The effects of these errors are investigated in case of Hamiltonian systems in terms of phase space structure. It is pointed out in Part I that (1) even if the discretized ordinary differential equations were perfect the time integration schemes unavoidably introduce errors, (2) the time discretization errors have a strong influence on the time evolution of the probability distribution function in the phase space, thus an ensemble of numerical runs is always distorted by model errors, (3) the initial spin-up process is unavoidable, even if the true state of the system can be observed. In a Hamiltonian system the accuracy of a numerical solution can always be improved by increasing the order of the integration scheme and decreasing the length of the time step. The concepts introduced are illustrated with simple examples and with numerical experiments carried out with the 2D vorticity equation and the two-layer quasi-geostrophic model.

Key-words: model errors, integrability, Hamiltonian structure.

---

[1] Present affiliation: UCAR scientific visitor at the National Centers for Environmental Prediction, USA; E-mail: Istvan.Szunyogh@noaa.gov

1. Introduction

The quality of numerical weather prediction is limited by two factors: the limited accuracy of initial conditions and the imperfection of the models. The main attention has in recent years been focused on the effects of uncertainties in the initial conditions. Lots of effort has been devoted to reduce the analysis errors through the development of enhanced data assimilation (*Pires et al.*, 1996 and references therein) and observation techniques (*Lorenz* and *Emmanuel*, 1997; *Bishop* and *Toth*, 1996). One might have the impression that the only requirement for producing better forecasts is to reduce the initial errors as much as possible. One should not forget, however, that even if the governing laws of the atmosphere were known exactly, and the numerical models based on these knowledge were perfect, furthermore, if the initial fields were exactly known, it is very unlikely that the state of the atmosphere at any future time could be predicted within the error limit of the initial data. This is the consequence of the fact that, to our present knowledge, the atmosphere is a nonintegrable nonlinear system, and as such it shows generic chaotic behaviour.

Since the nonhydrostatic adiabatic primitive equation and its consistent simplifications possess Hamiltonian structure and most of our knowledge about the connection between chaos and integrability is related to conservative Hamiltonian systems we limit our discussions to Hamiltonian systems of equations. Though it can be argued that the atmosphere is not conservative, rather a forced dissipative system, still a number of phenomena can be described within the conservative framework. The sceptic reader can easily check in standard textbooks (e.g. *Holton*, 1992 and *Pedlosky*, 1987) that almost all rigorous results of atmospheric dynamics were achieved with conservative systems of equations. Inspection of the contraction speed of the Liouville volume validates the non-forced non-dissipative approximation for processes with characteristic time much less than the characteristic time of the contraction. Usually it is the case when physical processes are neglected in simplified atmospheric models.

As a recent example, *Camassa* and *Tin* (1996) demonstrated that the dynamics of the dissipative 5-variable Lorenz model (L5) could be completely explored assuming that it was a perturbed version of the conservative one. In other words, the careful examination of the conservative subsystem gave the key to the understanding of the full forced-dissipative system. We believe that the L5 system is not an exception and dynamics of the more complex high dimensional forced-dissipative systems cannot be captured without a solid base of knowledge on their conservative counterparts.

The authors of the present review paper would like to point out that the problem of sensitivity to initial uncertainties and that of integrability are strongly connected, and none of them should be overlooked in favour of the other. We are not dealing here with model errors resulting from the incomplete

and/or incorrect parameterizations of physical and subgrid processes. We focus on the difference between the numerical model and the system modeled in the sense of algebraic structure and integrability. Particularly, we examine the effect of discretizations that, in many cases, destroy the structure of the equations. In the followings we do not attempt to give a complete overview of numerical Hamiltonian problems, this can be found in excellent papers and books like *Mackay* (1992), *Sanz-Serna* and *Calvo* (1994), *McLachlan* and *Scovel* (1993). The main goal here was to collect and illustrate those results of this quickly developing discipline which may substantially deepen our understanding on numerous problems of atmospheric dynamics and numerical weather prediction.

Eulerian equations of atmospheric dynamics are partial differential equations (PDEs) and their numerical approximation usually involves two steps. First, a spatial discretization is performed which transforms the evolution equations into a finite set of ordinary differential equations (ODEs), then a temporal discretization scheme is applied to facilitate the numerical integration of the equations. Motivated by the contemporary results of mathematics on numerical PDEs and ODEs (*Sanz-Serna*, 1992) we follow the dynamical systems approach, i.e. our main focus is rather on the qualitative (global, geometric) than on the quantitative (local) deficiencies caused by the truncations. The first part of the paper deals with the errors introduced in the numerical temporal integration, while the second part discusses the problem of spatial truncation. The reason for reversing the logical order is that almost all results on the chaotic nature of large scale atmospheric motions are achieved by using the hypothesis that the spatially discretized equations perfectly simulates the real atmosphere. While in Part II we will demonstrate that the spatial discretization inevitably alters the qualitative, as well as the quantitative behaviour of geophysical fluid dynamical system, hereafter we will illustrate that the model solutions are inevitably imperfect even for a perfect system of ODEs and initial conditions.

At the beginning two simple examples are presented in Section 2 and then we introduce the bases of Hamiltonian formalism in Section 3. In Section 4 the effects of temporal discretizations on the algebraic structure of the discretized system is discussed and a short account of structure preserving schemes are given. We illustrate the concepts introduced in the paper with simple examples and, in Section 5, with results of numerical experiments that were carried out with a structure preserving spatial truncation of the two-dimensional vorticity equation and the two-layer quasi-geostrophic model. Section 6 summarizes our considerations.

Throughout the paper we tried to introduce every concept that is necessary for understanding the main points, but we did not intend to give a complete introduction to dynamical systems theory. For further reading and references see, for example, *Götz* (1994, 1995).

2. *Simple examples*

Pendulum. With a suitable choice of units (i.e. taking the length of the weightless rod as unit of length and $\frac{1}{g}kg$ as the unit of mass), the system of equations describing the motion of the mathematical pendulum of unit mass takes the form

$$\begin{aligned} \dot{p} &= -\sin q, \\ \dot{q} &= p, \end{aligned} \tag{1}$$

where q is the angle of the rod to the vertical and p is the angular velocity. Since the order of the equation can be reduced by one with the help of a *first integral*, the energy, this system is totally *integrable*. The possible states of the system on the two-dimensional (p,q) phase plane form smooth curves, each of which corresponds to a given energy level (*Fig. 1*). Starting the integration from an initial condition that differs from the "true state" with a small error, the erroneous trajectory we follow will remain close to the true curve except if the two trajectories are on different sides of the separatrix. In this latter case,

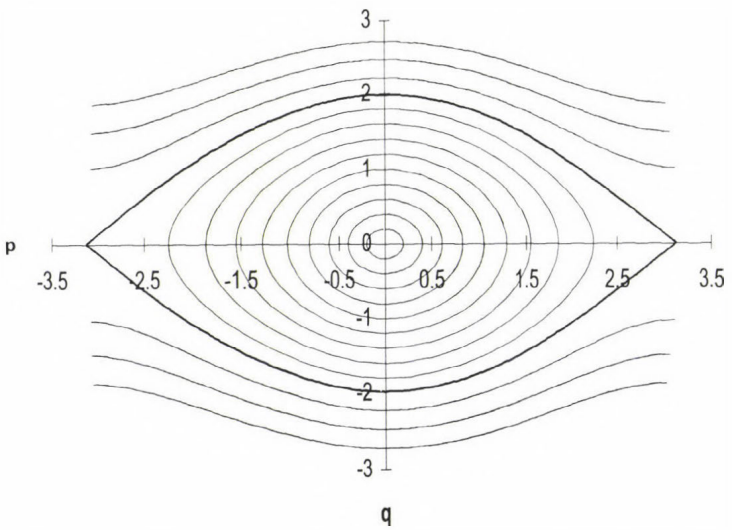


Fig. 1. Phase portrait of the pendulum. The solid lines are phase curves associated with given energy levels. Bold line indicates the separatrix.

however, if several observations of the system's trajectory are available over a given period of time, a 4D data assimilation algorithm may find iteratively an initial condition that corresponds to a qualitatively better trajectory, i.e. one that

follows the right direction of the pendulum. The highly unlikely case when either the true or the analyzed state falls on the separatrix could be handled in a similar way. *Lorenz* (1960) showed that a maximum simplification of the barotrop vorticity equation that is capable to simulate nonlinear barotrop phenomena is a vorticity triad. A vorticity triad consists of only three Fourier modes, with the sum of their wave vectors being equal to zero and the interaction of any pair of them altering the value of the third one. With a suitable transformation of variables and exploiting the invariance of enstrophy the real vorticity triad can be written in the form of Eq. (1) (see e.g. *Bokhove* and *Shepherd*, 1996).

Nondissipative 5-variable Lorenz model. The nondissipative five-component model (L5) of *Lorenz* (1986) was derived from a spectral shallow water model through extreme truncations. After further reduction with the use of an invariant quantity the equations can be written in the form (*Bokhove* and *Shepherd*, 1996)

$$\begin{aligned} \dot{q}_1 &= \epsilon p_1 , \\ \dot{p}_1 &= -\epsilon\, C \sin 2\left(q_1 + b\sqrt{\epsilon}\, q_2\right), \\ \dot{q}_2 &= p_2 , \\ \dot{p}_2 &= -q_2 - \epsilon^{3/2}\, b\, C \sin 2\left(q_1 + b\sqrt{\epsilon}\, q_2\right). \end{aligned} \tag{2}$$

As it was first observed in *Camassa* (1995), this is a nonlinearly coupled system of two integrable subsystems, a nonlinear pendulum and a linear harmonic oscillator: if $b = 0$ the last two equations give the linear differential equation of the harmonic oscillator $\ddot{q}_2 = -q_2$, while the first pair of equations reduces to $\ddot{q}_1 = -\epsilon^2 C \sin 2\, q_1$, which corresponds to the mathematical pendulum (Eq. (1)) with the length of the rod and $\frac{2\epsilon^2 C}{g}\, kg$ taken as the units of length and mass, respectively, and with q_1 measuring the half angle of the rod to the vertical. As it was mentioned previously, the pendulum is equivalent with a vorticity triad, while the solutions of the harmonic oscillator model gravity wave oscillations. The ratio of the frequencies of motions in the subsystems can be controlled with the parameter ϵ. At the limit $\epsilon \to 0$ a formal separation of timescales is present, and one can speak of p_1 and q_1 as *slow* and of p_2 and q_2 as *fast* variables. Accordingly, when $b \neq 0$ and ϵ is small Eq. (2) describes the nonlinear interactions of slow vortical and fast gravity wave-like motions. The system of Eq. (2) has the appealing feature of being the lowest order truncation of the atmospheric primitive equations. *Lynch* (1996) pointed out that structurally similar equations govern the motion of the elastic pendulum. Though the coupling terms between the pendulum and the linear oscillator are somewhat

different in the two systems the basic conclusions of the important papers *Camassa* (1995), *Camassa* and *Tin* (1996) and *Bokhove* and *Shepherd* (1996) do not depend on this detail. The reader, therefore, can always visualize the L5 model as an elastic pendulum in order to gain a better heuristic understanding of the underlying physics.

Bokhove and *Shepherd* (1996) demonstrated, that as the nonlinear coupling term b increases, huge portions of the phase space sections become chaotic. The system does not follow smooth trajectories when started from initial condition, albeit correct, in these chaotic regimes. The key process here is the increasing chaotic exchange of energy between the two subsystems: at a given instant the observer-forecaster may erroneously conclude that the pendulum will have enough energy to complete a rotation, simply because the linear oscillator charges more energy from the other subsystem than that was predicted. Once the inevitable uncertainty in the prediction of the energy exchange is larger than the error in the initial estimate of the energy, model errors start to dominate over the effects of initial uncertainties.

Remarkable theoretical results have been achieved using the geometric approach of dynamical systems theory. With these results (*Camassa*, 1995; *Camassa* and *Tin*, 1996) the vigorously debated existence/nonexistence of an integrable slowest manifold (*Lorenz*, 1980, 1986; *Leith*, 1980; *Errico*, 1984; *Lorenz* and *Krishnamurthy*, 1987) in the L5 has been proved in a local/global sense. While *Camassa* (1995), *Camassa* and *Tin* (1996) utilized the fact that the L5 model is a nonlinearly coupled system of two simple subsystems with well-known phase space geometry, *Bokhove* and *Shepherd* (1996) showed that if the formal separation of timescales is valid the model can be regarded as a slightly perturbed integrable Hamiltonian system, and the theorems of Hamiltonian perturbation theory apply. Thus, if the energy of the gravitational modes is sufficiently small at the beginning of the integration, it remains bounded, and the nonlinear system preserves (although slightly modified) its "nice" trajectories in most of the phase space. It suggests that if one can keep the energy of the linear oscillator at an infinitesimally small level at the beginning of model integration the qualitative dynamics of the pendulum remains intact, the initial and the integration errors have no catastrophic effects on the prediction. The practical realization of this idea is the normal mode initialization (*Lynch*, 1996).

The above example shows that chaos and nonintegrability occur even in very simple models. Indeed, this is generic in nonlinear dynamics. The classification of nonlinear systems as integrable or nonintegrable, however, is a problem that has no general solution yet. Some nonlinear PDEs encountered frequently in mathematical physics exhibit perplexing integrable behaviour. One illustrative example is the Korteweg-de Vries (KdV) equation. In an early experiment (in 1955) Fermi, Ulam and Pasta (FUP) were investigating the behaviour of a nonlinearly coupled chain of harmonic oscillators. They were

expecting sharing of energy between the modes (a kind of statistical equipartitioning of energy), but it turned out that the energy was cycling periodically among the modes initially generated. This astonishing integrable behaviour was explained later when the KdV equation, which is the continuum limit of the chain of oscillators used in the FUP experiment, was proved to have Hamiltonian structure and to be integrable. Its now famous, stable, nonlinear solutions are the so called *solitons*. It is interesting to mention that soliton like solutions are subjects to a non-linear superposition principle deeply related to the integrability property. For more about integrability in nonlinear dynamics see *Tabor* (1989), or other standard texts on dynamical system theory (e.g. *Ott*, 1994; *Arnold*, 1989).

The primitive equations are nonintegrable, at least, no one proved the contrary so far. From this point of view the goal of numerical modeling is to find approximate integrable models, so that in spite of the inevitable initial errors the state of the flow could be estimated over a given period of time. This is not impossible as it has been demonstrated over the history of numerical modeling and is manifested in the indisputable improvement in the skill of numerical weather prediction models.

3. Hamiltonian formulation

In the following subsections we briefly introduce the concept of Hamiltonian structure for both finite dimensional and continuous systems. The finite dimensional Hamiltonian formulation with the so called canonical variables (*generalized coordinates*, *generalized momenta*) has been widely used in classical mechanics. The pursue for canonical representation, however, restricted attention to even-dimensional systems. In the next subsections we present the more general description of finite dimensional Hamiltonian systems that led to the extension of Hamiltonian mechanics to odd-dimensional and also to infinite dimensional problems. A good introduction to the Hamiltonian methods can be found e.g. in *Olver* (1989) and in *Shepherd* (1990).

3.1 Finite dimensional Hamiltonian systems

Let $\mathbf{x}(t) = (x_1(t),\ldots,x_n(t))^T$ denote the vector of the state variables of an n-dimensional system. This system is called *Hamiltonian* if the differential equation describing its evolution in the phase space takes the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{D}(\mathbf{x})\nabla H(\mathbf{x}), \tag{3}$$

where the entries of the *structure matrix* **D** satisfy the equalities

(i) skew-symmetry

$$D_{ij}(\mathbf{x}) = -D_{ji}(\mathbf{x}), \qquad i,j = 1,...,n, \tag{4}$$

(ii) Jacobi identity

$$\sum_{l=1}^{n} \left(D_{pl}\frac{\partial D_{qr}}{\partial x_l} + D_{rl}\frac{\partial D_{pq}}{\partial x_l} + D_{ql}\frac{\partial D_{rp}}{\partial x_l} \right) = 0, \qquad p,q,r = 1,...,n \tag{5}$$

and H, the so called *Hamiltonian function*, is a real valued function, usually the total energy of the system.

For the familiarization with the Hamiltonian notation, it might be instructive to see the examples of Section 1 cast in the above form.

Nonlinear pendulum. If the vector of state variables is $\mathbf{x} = (p,q)^T$, and the Hamiltonian is defined as $H = \frac{p^2}{2} - \cos q$, which is the total (kinetic plus potential) energy of the pendulum, while the skew-symmetric structure matrix is $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, Eq. (3) readily gives Eq. (1). The structure matrix being a constant trivially obeys condition (5).

L5 model. To see the equivalence of Eq. (2) and Eq. (3), let $\mathbf{x} = (p_1,p_2,q_1,q_2)^T$, define the Hamiltonian as $H = \epsilon\frac{1}{2}p_1^2 - \frac{1}{2}\epsilon\, C\cos 2\left(q_1 + b\sqrt{\epsilon}\, q_2\right) + \frac{1}{2}\left(p_2^2 + q_2^2\right)$ and use the structure matrix $\begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}$, where $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Note that both of the above models are examples for *canonical* Hamiltonian systems, in which $\mathbf{D} = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}$ with I being the $n \times n$ identity matrix, and n is the number of degrees of freedom in the system. The q-s are called *generalized coordinates* and the p-s are the *generalized momenta*. The definition Eq. (3), however, is not restricted to this special type of structure matrices, not even to even-dimensional cases.

The reason why canonical Hamiltonian systems have got particular attention is that according to Darboux' theorem (see e.g. in *Olver*, 1989) for any finite dimensional Hamiltonian system defined on an m dimensional manifold there exist local coordinates $(p_1,\ldots,p_n,q_1,\ldots,q_n,z_1,\ldots,z_l)$, $m = 2n + l$ in which the structure matrix has the form

$$\mathbf{D} = \begin{bmatrix} 0 & -I & 0 \\ I & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{6}$$

where I is the $n \times n$ identity matrix. The coordinates $z_1,\ldots,z_l$ are the so called *distinguished coordinates*, that are constants along the Hamiltonian flow and any quantity which is function of the distinguished coordinates alone is an invariant of the system. These are the so called *distinguished functions* or *Casimir invariants*.

3.2 Infinite dimensional Hamiltonian systems

The counterpart of Eq. (3) for infinite dimensional evolution equations is

$$\frac{\partial \boldsymbol{u}}{\partial t} = \mathcal{D}\,\delta\,\mathcal{H}(\mathbf{u}), \tag{7}$$

where the prognostic variable u(x(t),t) is a vector valued function of space and time, whose partial derivative with respect to time is taken, and the Hamiltonian $\mathcal{H}(\mathbf{u})$ is a functional (real valued function of a function) whose functional derivative plays the role of the gradient in Eq. (3). The skew-symmetric structure matrix of Eq. (3) is replaced by a skew-adjoint differential operator $\mathcal{D}(\mathbf{u})$ that must satisfy the continuous equivalent of the Jacobi identity (5) as well.

Let the two-dimensional vorticity equation stand here as an example. Take the two-dimensional vorticity $\zeta(\mathrm{x},t)$ as the prognostic variable, the two-dimensional Jacobian $\partial(\zeta,.)$ (with a negative sign) as $\mathcal{D}$ and $\mathcal{H}(\psi) = \frac{1}{2}\int_\Omega \nabla\psi^2 d\mathrm{x}$ as the Hamiltonian, where $\psi(\mathrm{x},t) = \Delta^{-1}\zeta(\mathrm{x},t)$ is the stream function. Upon substitution to Eq. (7) one obtains the familiar governing equation for an inviscid, nondivergent fluid flow over the model domain Ω

$$\frac{\partial \zeta}{\partial t} = \partial(\zeta,\psi). \tag{8}$$

For the Hamiltonian formulation of the shallow water equation, the baroclinic quasi-geostrophic flow over topography and that of the nonhydrostatic primitive equation see *Shepherd* (1990) and references therein.

4. *Temporal discretization*

The problem of integration in time is that in case of most integration schemes a number of important geometric properties of the flow defined by the spatially discretized system are lost under the transformation of timestepping. Such property can be, for example, energy conservation, volume preservation (*Quispel*, 1995), reversibility (*McLachlan et al.*, 1996) or, in case of Hamiltonian systems, preservation of area or symplectic structure (*Yoshida*, 1990). Lots of activity has been focused on devising temporal integration schemes (integrators) that preserve certain geometric properties of the flow, and it seems, that essentially, any properties of the ODE can be retained with an appropriate integrator. But no integration scheme can preserve "everything". As we outline below conservation of the finite dimensional Hamiltonian structure (i.e. *symplectic structure*) and exact conservation of energy are conflicting requirements of a nonintegrable system.

4.1 Preservation of structure

As it was mentioned in the previous section, some equations of fluid dynamics have *structure preserving* spatial truncation, that is the discretization in space of the continuous Hamiltonian system results in a finite dimensional Hamiltonian system of ordinary differential equations. In this subsection we briefly describe what symplectic structure means, and introduce integration schemes that preserve this property, the so called *symplectic integrators*.

For one-degree-of-freedom systems symplectic structure is just a different expression for the preservation of oriented phase plane area. Since the phase space is two-dimensional in this case, it is equivalent with phase space volume preservation which follows from Liouville's theorem. For n-degree-of-freedom (that is 2n-dimensional) systems the symplectic map is a generalization of area preservation stricter than volume preservation. In fact, if the model domain is simple enough, it is equivalent with the system being Hamiltonian. It requires that the sum of certain phase space areas should be conserved along the flow. More specifically, if an m-(not necessarily even-)dimensional system is given with the canonical coordinates $(p_1,\ldots,p_n, q_1,\ldots,q_n, z_1,\ldots,q_l)$, $m = 2n + l$, with the z_i-s being distinguished coordinates, the condition of symplectic area-preservation reads

$$\oint_{\Gamma(t=0)} \sum_{i=1}^{n} p_i \, dq_i = \oint_{\Gamma(t=T)} \sum_{i=1}^{n} p_i \, dq_i, \tag{9}$$

where Γ is a closed curve surrounding a tube of phase space trajectories and

evolving with the flow. That is here the sum of the projections of the area encircled by Γ on the (p_i, q_i) phase planes is conserved.

In order to preserve the Hamiltonian structure the integration scheme should be a discrete mapping of the phase space that also preserves the area in the above sense. Such mappings are called *canonical* or *symplectic* transformations. One of the simplest examples for symplectic integrator is the *midpoint rule*

$$\mathbf{x}^{n+1} = \mathbf{x}^n + h_{n+1}\,\mathbf{F}\left(\frac{1}{2}(\mathbf{x}^n + \mathbf{x}^{n+1}); t_n + \frac{1}{2}h_{n+1}\right), \tag{10}$$

where $\mathbf{F}$ is the right-hand side of the ODE and the superscript n denotes the approximated value of the prognostic variable $\mathbf{x}$ at the n-th time step with h being the length of a time step. Using the notation of Section 2, for a Hamiltonian system, $\mathbf{F}$ is the right-hand side of Eq. (3), i.e. $\mathbf{F} = \mathbf{D}\nabla H$. A number of Runge-Kutta methods are also symplectic, provided that their coefficients satisfy certain conditions. A detailed account of symplectic numerical methods can be found in *Sanz-Serna* and *Calvo* (1994). The interested reader can find further surveys of the progress made in the field of symplectic integrators e.g. in *Yoshida* (1993) or in *McLachlan* and *Scovel* (1993).

The difference between "normal" and symplectic methods is demonstrated on *Fig. 2*. The pendulum equation (Eq. (1)) was integrated with the explicit Euler method (Fig. 2a, d), with the implicit Euler scheme (Fig. 2b, e) and with the midpoint rule (Fig. 2c, f). The points of a circle of radius 0.3, centered at $p = 1$, $q = 0$ in the (p,q) phase plane were chosen as initial conditions. $h = \frac{\pi}{12}$ was used as time step. On Fig. 2a-c the computed points are plotted after every third time steps up to the *27*th step, while on the panels Fig. 2d-f the trajectories of the point $p = 1$, $q = 0.3$ are followed, plotting the points after each time step up to the *60*th. Although the closed curve (initially a circle) should keep its area along the phase space flow generated by Eq. (1), the Euler methods fail to preserve this property. In addition they turn the fix point $(p,q) = (0,0)$ into stable and unstable spiral points (see Fig. 2e, Fig. 2d, respectively) unknown in Hamiltonian systems. Note that the circle of initial points can be regarded as an ensemble describing the uncertainty in the position of the initial point. It is clear, that as the ensemble is propagated along the flow it is not only reflecting the varying nature of the probability density function of the initial uncertainty, but is highly influenced by model errors, as well. The main problem with these errors that they are systematic and thus cannot be simulated as a result of random processes. This fact suggests that handling of model errors in an ensemble prediction system is a highly nontrivial problem.

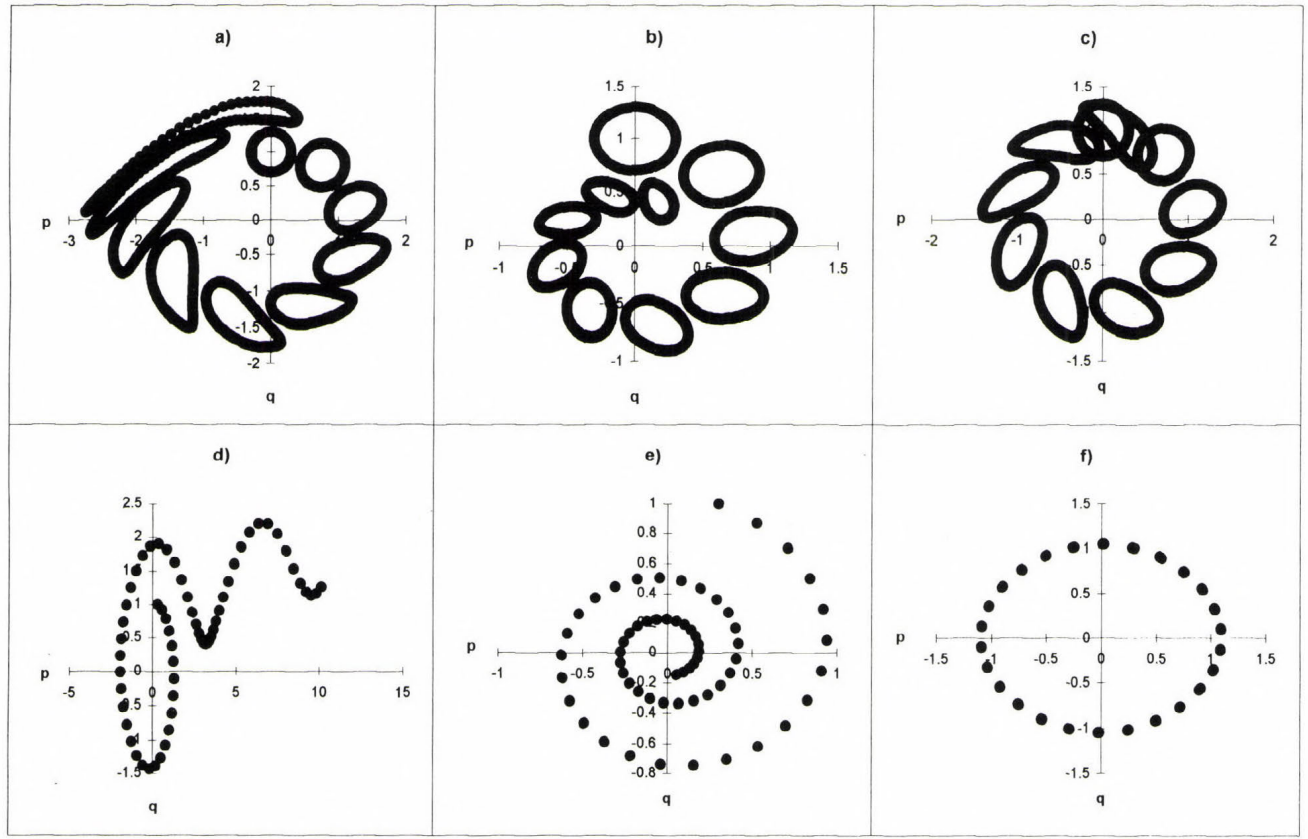


Fig. 2. The pendulum equation integrated by the explicit (*a, d*), implicit (*b, e*) Euler methods and with the midpoint rule (*c, f*). See also text in Section 4.1.

4.2 Conservation of energy

Since the invariants of motion constrain the flow into a smaller dimensional subspace of the phase space, it is assumed that the more invariants an integrator can preserve the better it is. Numerous schemes have been invented that conserve one or more invariants, e.g. energy, potential enstrophy, etc. These schemes, however, usually destroy the Hamiltonian structure of the flow. The question naturally arises, whether symplectic integrators can preserve the energy and/or other invariants. According to the theorem of *Ge* and *Marsden* (1988) *a symplectic method cannot conserve energy exactly for a general nonintegrable Hamiltonian system*. Even for an integrable system the preservation of the Hamiltonian is possible only if the time map of the scheme is identical with that of the true flow. Since, in general, the preservation of energy excludes the preservation of structure, and vice versa, a necessary compromise should be reached in case of every concrete problem at hand. In case of nonlinear wave equation (see *McLachlan*, 1994), for instance: Either one applies a symplectic scheme that conserves the structure (i.e. the shape) of the travelling waves, but then its speed will be in error, or one insists on the conservation of energy and uses an energy conserving scheme that models the speed of the wave correctly, but gives erroneous wave shape. Conservation of energy reduces the dimensionality of the problem only by one. It seems plausible, that the symplecticness, which distinguishes the class of Hamiltonian systems within that more general class of energy conserving ones, is a somewhat more distinctive characteristic of the Hamiltonian flows.

The fact that energy conservation and symplecticness are conflicting requirements of a scheme does not imply that the symplectic methods are completely wrong in terms of energy errors. In fact, it can be proved, that *for any symplectic integrator applied to an autonomous Hamiltonian problem, a modified autonomous Hamiltonian system can be found so that the numerically computed points are either exactly or "very nearly" on the trajectories of this modified system* (see e.g. *Sanz-Serna* and *Calvo*, 1994, 133 p.). Since the structure of a symplectic integrator is perfect within a negligible roundoff error this modification means a perturbation of the Hamiltonian, i.e. the model errors are encapsulated into the energy error. This perturbation leads to an initial jump in the energy of the model, but the energy error cannot grow without limit over any number of time steps. Indeed, the energy error can be reduced to the roundoff level increasing the order of the integrator and decreasing the time step. On the other hand the initial jump in the energy is an obvious model bias, the forecast will always be burdened with errors, even if we know the true governing equations and the initial conditions. The initial jump of the energy is a well known phenomenon in NWP and is generally described as a result of the equilibration process between forcing and dissipative terms. While the contribution of these terms is obvious one should keep in mind that the above

adjustment process is a must even if the source and the sink terms are all switched off.

Consider again the simple example of the pendulum. *Fig. 3* shows the "true" trajectory of the pendulum emanating from the initial point $p = 1.41615$, $q = 0.0$ (solid line) and points (small diamonds) that were computed with a symplectic Runge-Kutta scheme ($p^{n+1} = p^n - h \sin q^{n+1}$, $q^{n+1} = q^n + hp^n$). It can be shown (*Sanz-Serna* and *Calvo*, 1994) that the scheme preserves to the second order the modified Hamiltonian $H_2^h = \frac{p^2}{2} - \cos q + \frac{h}{2} p \sin q$. Since the trajectory corresponding to the approximately conserved modified Hamiltonian is also a closed curve the energy error cannot grow without limit. The facts that the energy errors remain bounded even over extremely long integrations and that it can be further reduced to arbitrarily small amount using higher order schemes and shorter time steps, are great benefits of symplectic integrators. On the other hand, the perfect conservation of structure and a good preservation of energy cannot guarantee any error bound for the individual variables of a nonintegrable chaotic system.

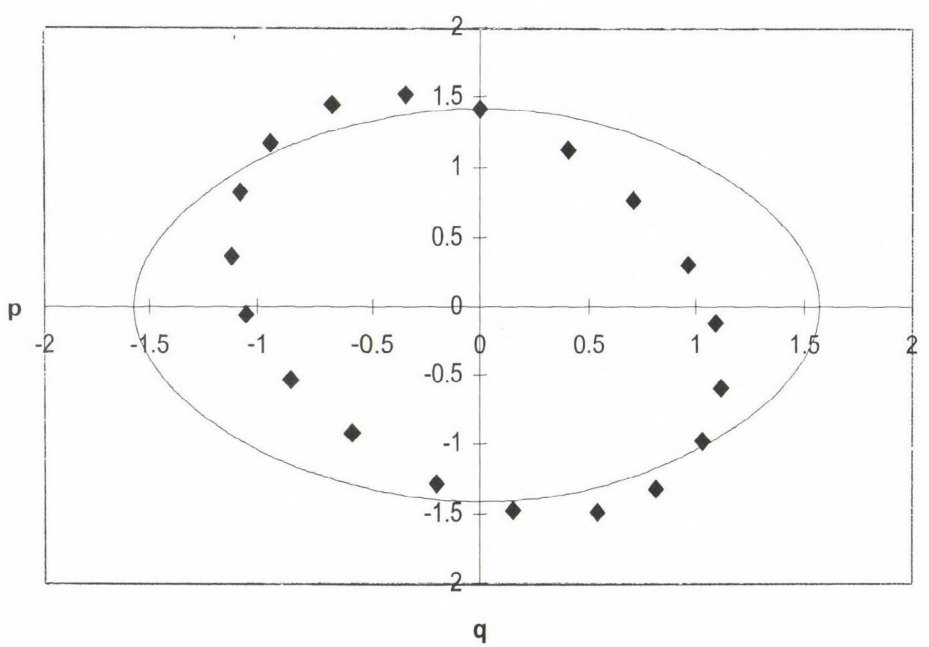


Fig. 3. The true trajectory (solid line) of the pendulum from initial point $p = 1.41615$, $q = 0.0$ and points (diamonds) computed with a symplectic Runge-Kutta method from the same initial condition.

5. *Numerical experiments*

In this section some numerical results will be presented to illustrate the above discussed features of symplectic integrators. In the experiments summarized below the structure preserving spatial truncation of the two-dimensional vorticity equation (hereafter 2DV) (*Zeitlin*, 1991) and that of the two-layer quasi-geostrophic equations (denoted by L2 QG) (*McLachlan et al.*, 1997) were used. A comparative study of the structure preserving truncation and a traditional spatial discretization applied to these equations will be detailed in the second part of this paper. For the time integration a slightly modified version of McLachlan's explicit symplectic integrator was used. The integration scheme and its modification is briefly described in the Appendix. A detailed description can be found in *McLachlan* (1993) or in *McLachlan et al.* (1997).

In the description of the experiments we refer to the resolution with TM, where M is a positive integer, and it implies that a cutoff wavenumber of M is used, that is, spectral modes associated with wave vectors $\mathbf{k} = (k_1, k_2)$, $|k_1|$, $|k_2| \leq M$ are considered in the truncated system.

Since the spatially truncated system possesses a Hamiltonian structure and the time integration scheme is symplectic, the preservation of structure is guaranteed in the integration. As a consequence, the Casimir invariants related to the structure matrix (e.g. the enstrophy) can be shown to be conserved within the error limit of number representation. On the other hand, if the system is nonintegrable, the conservation of the Hamiltonian (i.e. the energy) and of those invariants that are associated with its symmetries are inevitably burdened with errors. Thus a straightforward method to test the skill of a symplectic scheme is to examine the behavior of the energy error.

The relative energy error, $(H(t) - H(t=0))/H(t=0)$, was calculated for model runs started from randomly generated initial conditions at different resolutions. The integrations were carried out for 10^6 time steps at T1 and T2, and for 10^5 time steps at T13. In the first series of experiments first order integration scheme was used and the time step was uniformly set to 1.0. Taking the reciprocal of the average (potential) vorticity $1/\bar{\zeta}$ sec as time unit, the unit time step corresponds to $\bar{\zeta}$ nondimensional time unit. A similar time unit for the atmosphere is 10^5 sec, thus $\bar{\zeta} \times 10^5$ sec is a fair estimate of the time steps used in the different size systems. In the numerical experiments the initial vorticity field for the 2DV model were generated randomly with averages 0.09796 1/sec, 0.06091 1/sec and 0.00801 1/sec at T1, T2 and T13, respectively. With these values the unit time step corresponds to an atmospheric equivalent time step of about 160 min, 102 min and 13 min at the given resolutions. The average potential vorticity in the initial conditions generated for the L2 QG model was of the magnitude 10^{-2} 1/sec at all three resolutions, thus the unit time step in these cases agrees to an atmospheric equivalent of about 15 min.

In the 2DV runs the maximum energy error was in the magnitude of 10^{-1} at T1, 10^{-2} at T2, and 10^{-1} at T13. These significant errors are due to the fact that the time step is too large for randomly generated fields. Reduction of the time step to its hundredth resulted in relative errors two magnitudes smaller. Another way of reducing the energy errors is to apply a higher order integration scheme. Thus in the next row of experiments the second order version of McLachlan's time integrator was applied. The relative energy errors obtained with 1.0 length time steps are plotted on *Fig. 4* at T1, T2 and T13. The magnitude of the relative errors are of 10^{-2} at T1, 10^{-4} at T2 and 10^{-4} at T13. Reduction of the time steps to 0.01 results in further decrease in the errors, the magnitude of maximum relative energy errors being 10^{-7} at T1, 10^{-8} at T2 and 10^{-7} at T13.

The relative energy errors in the L2 QG runs with the first order scheme and 1.0 time steps were 10^{-3} at T1, T2 and also at T13. Using hundred times smaller time steps gave $O(10^{-5})$ relative energy errors. Second order integration in time gave relative errors in the magnitude of 10^{-6} and 10^{-5} at T1 and T2, and of 10^{-4} at T13 when the time step was set to 1.0, and of 10^{-8}, 10^{-9} and 10^{-8} at T1, T2 and T13, respectively, when the time steps were 0.01.

Using the second order scheme implied only doubling of the number of operations and was as effective as reducing the time step which required significantly more CPU time to perform the integration over the same period of time (hundred times more time steps had to be performed). The advantage of using higher and higher order schemes, however, is not a general rule since further increase of order leads to a quadratic drop in the computational efficiency, while the same factor for reducing the time step is only linear.

An important feature of the relative energy errors plotted on the panels of Fig. 4 is that in all three cases the relative error of energy oscillates around a nonzero value, but remains bounded. This is also true for the relative energy errors in the L2 QG model (not shown). This is the manifestation of the fact that the symplectic scheme conserves the energy of an approximating or perturbed Hamiltonian system. The perturbed Hamiltonian of a higher order scheme is closer to the "true" energy level than that of a lower order scheme. This is demonstrated on *Fig. 5a-b*, where the relative error of energy for the 2DV model is plotted at T2, calculated with first (Fig. 5a) and second order (Fig. 5b) schemes from the same random initial conditions. Note that the average relative perturbation of the Hamiltonian (the value around which the errors oscillate) are of opposite sign on these plots.

Fig. 5c-d show a similar pair of energy error curves. These plots were obtained with the 2DV model at T13 resolution, with first and second order scheme. The initial conditions were the discrete approximation of a rotating cone placed in the vorticity field. That is the initial field in this experiment was a discrete Fourier transform of a rotating cone of radius ρ defined in the physical space as

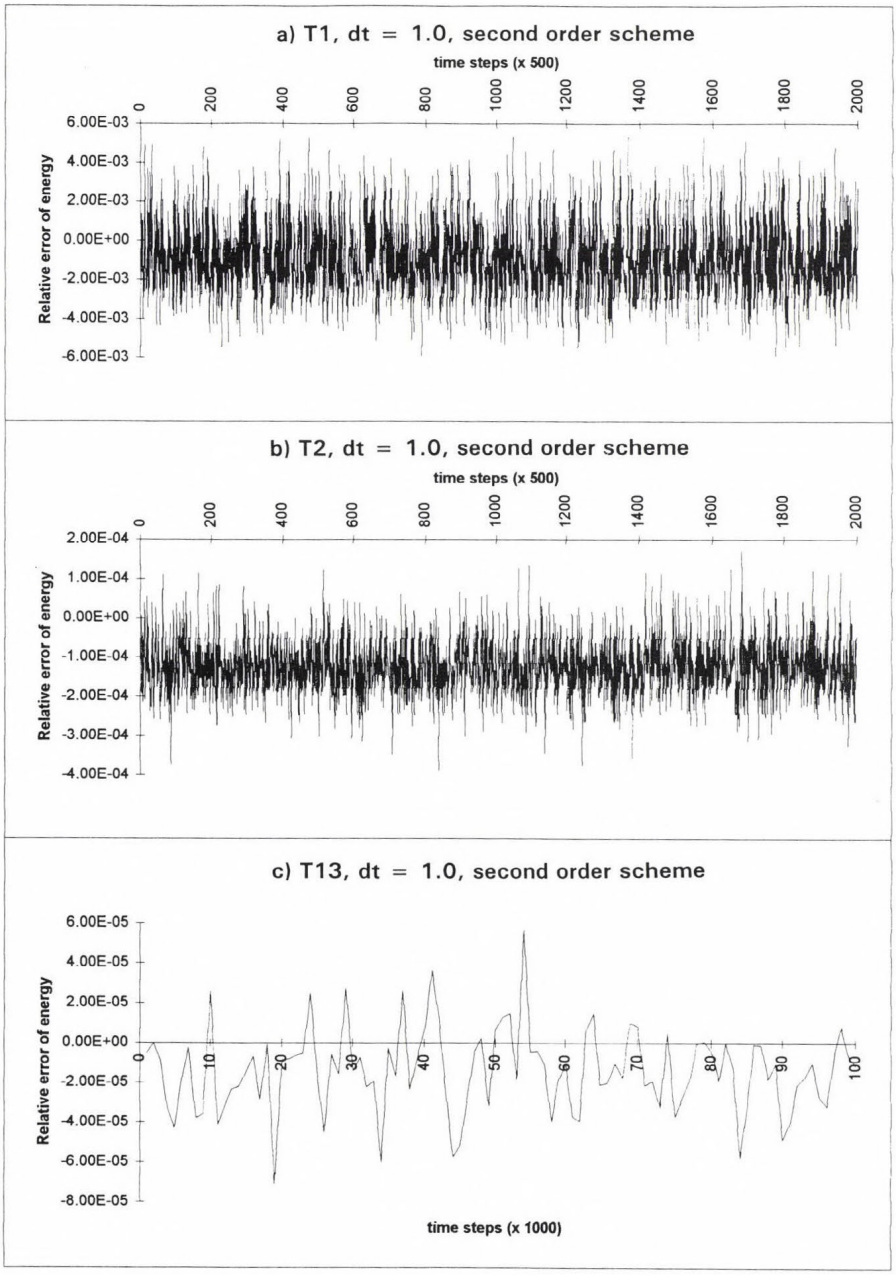


Fig. 4. Relative energy errors at T1 (*a*), T2 (*b*) and T13 (*c*) obtained by integrating the 2DV model with second order symplectic scheme and time step 1.0.

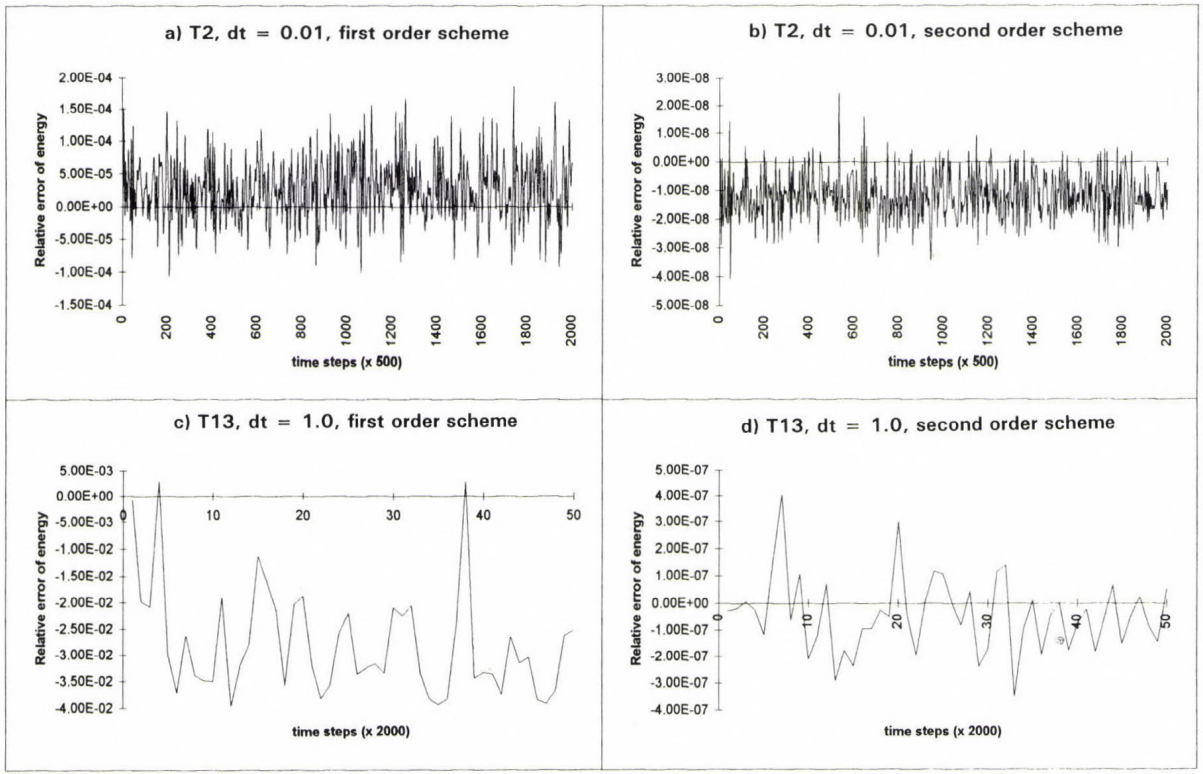


Fig. 5. Panels *a, b* show the relative energy errors obtained by integrating the 2DV model at T2 with the first (*a*) and second (*b*) order scheme started from random initial conditions. Panels *c, d* present the relative energy at T13 with first (*c*) and second (*d*) order scheme when the initial condition corresponds to the spectral transform of a cone placed in the vorticity field.

$$\zeta(\mathbf{r},t=0) = \begin{cases} 1 - \frac{|\mathbf{r}|}{\rho} & \text{for } |\mathbf{r}| \le \rho \\ 0 & \text{for } |\mathbf{r}| > \rho \end{cases}, \tag{11}$$

where $|\mathbf{r}| = ((r_1 - r_{01})^2 + (r_2 - r_{02})^2)^{\frac{1}{2}}$ and $\mathbf{r}_0 = (r_{01}, r_{02})$ is the place of the rotation axis of the cone. (Here $\mathbf{r}$ denotes vector of place in the physical domain.) This is a stationary solution of the continuous equation, thus a proper test also for the accuracy of the models (see Part II of this paper). In the numerical experiment $r_{01} = 3$, $r_{02} = 7$ and $\rho = 3$ were chosen and unit time steps were used. The relative error of energy is plotted at every 2,000th time steps up to the 100,000th time step on Fig. 5c-d. An interesting feature of numerical models can be observed on the relative energy error plots of the first order run. There are some "jumps" in the relative energy error at the beginning of the integration interval before a steady oscillation begins. Presumably, because the system "needs some time" to find that energy shell of the approximating Hamiltonian system, on which it remains afterwards (*McLachlan*, 1996). Evidently, the use of a higher order scheme resulted in the improvement of the energy conservation with orders of magnitude, and at the same time the initial "jumps" almost vanished indicating that the approximating Hamiltonian is much "closer" to that of the truncated system.

6. Summary

The "perfect model" hypothesis is a widely used presumption in atmospheric dynamics and in numerical weather prediction that makes possible to deal with the problem of initialization with the tools of dynamical systems theory. On longer terms, however, model errors tend to dominate. Their sources and effects are more complex than that of the errors in the initial conditions since they include e.g. the parameterizations of physical and subgrid processes. There are no general methods to approach these heterogeneous sources of errors. The present review article (along with an accompanying paper) focuses on the systematic errors caused only by the temporal and spatial discretizations. We examined the effect of truncations on the algebraic — Hamiltonian — structure of conservative continuous equations. This restriction of attention to conservative systems is justified by the facts that Hamiltonian formalism has been payed more and more attention recently and a number of exciting theoretical results have been obtained in this framework (see Section 1 and 2). We gave a brief introduction to the Hamiltonian formalism in Section 3.

Given a spatially discretized system of equations with Hamiltonian structure, the temporal integration plays a crucial role in determining the phase space

behaviour of the flow. The timestepping algorithm inevitably alter the structure and/or the energy level of the flow. In Section 4 the reader can get acquainted with the concept of symplecticness, a characteristic feature of finite dimensional Hamiltonian flows, and with symplectic time integrators that preserve this property. The conservation properties of these schemes are briefly reviewed. The numerical results presented in Section 5 confirm the theory that the energy conservation of symplectic integrators is remarkable, even though it is impossible to conserve the exact energy of a nonintegrable system. Despite of the attractive results in terms of energy conservation, we abstain ourselves from drawing final conclusions about the applicability of these schemes until the end of the second part of this paper.

Acknowledgements—This work was partly supported by a grant from State Scientific Research Fund (OTKA F 019268). Encouraged by the informal discussions great progress was achieved during the authors' stay at the Isaac Newton Institute for Mathematical Sciences, Cambridge, England, within the program "Mathematics of the Atmosphere and Ocean Dynamics" in the summer and autumn of 1996. The critical comments of the anonymous referees on the previous version of the manuscript are greatly appreciated.

References

Arnold, V.I., 1989: *Mathematical Methods in Classical Mechanics*. Springer-Verlag, New York, 508 pp.

Bishop, C.H. and *Toth, Z.*, 1996: Using ensembles to identify observations likely to improve forecasts. *Contributions to the 11th AMS Conference on Numerical Weather Prediction*, Aug. 19-23, Norfolk, Virginia, 72-74.

Bokhove, O. and *Shepherd, T.*, 1996: On Hamiltonian balanced dynamics and the slowest invariant manifold *J. Atmos. Sci. 53*, 276-297.

Camassa, R., 1995: On the geometry of the atmospheric slow manifold. *Physica D 84*, 357-397.

Camassa, R. and *Tin, S.*, 1996: The global geometry of the slow manifold in the Lorenz-Krishnamurthy model. *J. Atmos. Sci. 53*, 3251-3264.

Errico, R.M., 1984: The statistical equilibrium solution of a primitive-equation model. *Tellus 36A*, 42-51.

Ge, Z. and *Marsden, J.E.*, 1988: Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators. *Phys. Lett. A 133*, 134-139.

Götz, G., 1994: Application of nonlinear dynamics in atmospheric sciences. *Időjárás 98*, 65-86.

Götz, G., 1995: Predictability of nonlinear dynamical systems. *Időjárás 99*, 1-32.

Holton, J.R., 1992: *An Introduction to Dynamic Meteorology*. 3rd edition. Academic Press, San Diego, 507 pp.

Leith, C.E., 1980: Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci. 37*, 958-968.

Lorenz, E.N., 1960: Maximum simplification of dynamic equations. *Quart. J. Roy. Meteorol. Soc. 12*, 243-254.

Lorenz, E.N., 1980: Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci. 37*, 1685-1699.

Lorenz, E.N., 1986: On the existence of a slow manifold. *J. Atmos. Sci. 43*, 1547-1557.

Lorenz, E.N. and *Emmanuel, K.A.*, 1997: Optimal sites for supplementary weather observations: Simulation with a small model. Submitted to *J. Atmos. Sci.*

Lorenz, E.N. and *Krishnamurthy, V.*, 1987: On the nonexistence of a slow-manifold. *J. Atm. Sci. 44*, 2940-2950.

Lynch, P., 1996: *The Elastic Pendulum: A Simple Mechanical Model of Atmospheric Balance*. The Irish Meteorological Service Technical Note, No. 54. The Irish Meteorological Service.

Mackay, R.S., 1992: Some aspects of the dynamics and numerics of Hamiltonian systems. In *The Dynamics of Numerics and the Numerics of Dynamics* (eds.: *D.S. Broomhead* and *A. Iserles*). Clarendon Press, Oxford, pp. 137-194.

McLachlan, R., 1993: Explicit Lie-Poisson integration and the Euler equations. *Phys. Rev. Lett. 71*, 3043-3046.

McLachlan, R., 1994: Symplectic integration of Hamiltonian wave equations. *Numer. Math. 66*, 465-492.

McLachlan, R.I. and *Scovel, C.*, 1993: A survey of open problems in symplectic integration. *Proc. Conf. Integration Algorithms in Classical Mechanics* (ed.: *G. Patrick*). Under review.

McLachlan, R.I., *Szunyogh, I.* and *Zeitlin, V.*, 1997: Hamiltonian finite-dimensional models of baroclinic instability. *Phys. Lett. 229A*, 299-305.

McLachlan, R.I., *Quispel, G.R.W.* and *Turner, G.S.*, 1996: Numerical integrators that preserve symmetries and reversing symmetries. Under review.

Olver, P.J., 1989: *Applications of Lie Groups to Differential Equations*. Springer-Verlag, New York, 513 pp.

Ott, E., 1994: *Chaos in Nonlinear Dynamical Systems*. Cambridge University Press, Cambridge, 385 pp.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics*. Second edition. Springer-Verlag, New York, 710 pp.

Pires, C., *Vautard, R.* and *Talagrand, O.*, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus 48A*, 96-121.

Quispel, G.R.W., 1995: Volume-preserving integrators. *Phys. Lett. A 206*, 26-30.

Sanz-Serna, J.M., 1992: Numerical ordinary differential equations *vs.* Dynamical systems. In *The Dynamics of Numerics and the Numerics of Dynamics* (eds.: *D.S. Broomhead* and *A. Iserles*). Clarendon Press, Oxford, pp. 81-106.

Sanz-Serna, J.M. and *Calvo, M.P.*, 1994: *Numerical Hamiltonian Problems*. Chapman & Hall, London, 207 pp.

Shepherd, T.G., 1990: Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Adv. Geophys. 32*, 287-337.

Tabor, M., 1989: *Chaos and Integrability in Nonlinear Dynamics*. John Wiley & Sons, New York, 364 pp.

Yoshida, H., 1990: Construction of higher order symplectic integrators. *Phys. Lett. A 150*, 262-268.

Yoshida, H., 1993: Recent progress in the theory and application of symplectic integrators. *Cel. Mech. Dyn. Astr. 56*, 27-43.

Zeitlin, V., 1991: Finite-mode analogs of 2D ideal hydrodynamics: Coadjoint orbits and local canonical structure. *Physica D 49*, 353-362.

Appendix

1. McLachlan's algorithm for the integration of the 2D vorticity equation

The algorithm, whose detailed description can be found in *McLachlan* (1993), is a realization of the so called *composition method*, in which the vector field defined by the right-hand side of Eq. (3) is decomposed into vector fields, that are (more) easily integrated by suitable explicit methods.

Hereafter the vector of prognostical variables will be denoted with ζ, whose components $\zeta_{\mathbf{i}}$ are the spectral coefficients of vorticity associated with the wave vector $\mathbf{i}$. The index vectors are from the finite lattice $\{\mathbf{i} \,|\, |i_1|, |i_2| \leq M\}$, where M is the cutoff wavenumber. N denotes $2 \cdot M + 1$. The algorithm described in *McLachlan* (1993) is applicable when N is prime.

The algorithm is based on the splitting of the Hamiltonian $H = \sum_{\mathbf{k}} H_{\mathbf{k}}$, where $H_{\mathbf{k}} = \frac{1}{2} \sum_{n=0}^{N-1} \frac{\zeta_{n\mathbf{k}} \zeta_{-n\mathbf{k}}}{|n\mathbf{k}|^2}$, with $\mathbf{k} \in \mathbf{K} = \{(0,1)\} \bigcup \{(1,m) \,|\, 0 \leq m < N\}$ and the indices are taken modulo N.

McLachlan's observation is that this splitting of the Hamiltonian reduces the problem of integrating the truncated spectral vorticity equation to the solution of a set of linear differential equations with constant coefficients, each corresponding to a vector field $\mathbf{D}\nabla H_{\mathbf{k}}$, with $\mathbf{k} \in \mathbf{K}$. Let denote $\exp(dt \cdot \mathbf{D}\nabla H_{\mathbf{k}})$ the resolvent of these linear differential equations. A first order approximation in the time step dt of the solution of Eq. (3), i.e. that of the equation $\dot{\zeta} = \mathbf{D}\nabla H = \sum_{\mathbf{k}} \mathbf{D}\nabla H_{\mathbf{k}}$ is $\zeta(t) = \prod_{\mathbf{k}} \exp(dt \cdot \mathbf{D}\nabla H_{\mathbf{k}}) \zeta(t=0)$ (see e.g. *McLachlan*, 1993). While

$$\zeta(t) = \exp\left(dt/2 \cdot \mathbf{D}\nabla H_{\mathbf{k}_1}\right) \ldots \exp\left(dt/2 \cdot \mathbf{D}\nabla H_{\mathbf{k}_{N+1}}\right)$$
$$\times \exp\left(dt/2 \cdot \mathbf{D}\nabla H_{\mathbf{k}_{N+1}}\right) \ldots \exp\left(dt/2 \cdot \mathbf{D}\nabla H_{\mathbf{k}_1}\right) \zeta(t=0),$$

where $\mathbf{k}_i \in \mathbf{K}$, gives a second order (in dt) approximation to the solution of Eq. (3) (*McLachlan*, 1996).

The vector field $\mathbf{D}\nabla H_{\mathbf{k}}$ defines an explicitly integrable system of equations because each component of the prognostic variable vector can be written as $\zeta_{\mathbf{j}+m\mathbf{k}}$, where $-M \leq m \leq M$ and $\mathbf{j}$ is an appropriate transition vector of the indices. Using the notation $z_m = \zeta_{\mathbf{j}+m\mathbf{k}}$ the evolution of the vector z can be given with the set of linear differential equations of $\dot{z}_m = \sum_{n=-M}^{M} a_n z_{m-n}$, the discrete Fourier transform of which gives an explicitly integrable system $d\tilde{z}/dt = \Lambda \tilde{z}$, where $\tilde{z} = \boldsymbol{F} z$, $\Lambda = diag(\boldsymbol{F}\mathbf{a})$, $\boldsymbol{F}$ being the matrix of the discrete Fourier transformation.

Thus the algorithm for integrating the 2D vorticity equation consists of the following steps:

(i) for $\mathbf{k} \in \mathbf{K}$ do,

(ii) for $\mathbf{j} = 1\text{st}, \ldots, M\text{th}$ translation of $\mathbf{k}$ do,

(iii) with $z_m = \zeta_{\mathbf{j}+m\mathbf{k}}$ set $z = \boldsymbol{F}^{-1} e^{\Lambda \cdot dt} \boldsymbol{F} z$, where $\Lambda = diag(\boldsymbol{F}\mathbf{a})$,

(iv) set $\zeta_{-(\mathbf{j}+m\mathbf{k})} = z_m^*$,

(v) end do,

(vi) end do.

2. Modifications to McLachlan's algorithm

In order to be able to carry out effective numerical integrations with the use of available FFT routines, we extended the algorithm to certain non-prime Ns. Our modifications were based on the following observations:

(i) If N is prime, the only common element in the sets $\sigma_{\mathbf{k}} = \{n\mathbf{k} \ mod N \mid 0 \leq n < N\}$ with $\mathbf{k} \in \mathbf{K}$ is the $\mathbf{0} = (0,0)$ index and $\bigcup_{\mathbf{k} \in \mathbf{K}} \sigma_{\mathbf{k}}$ covers the entire lattice.

(ii) If N is non-prime, then the $\sigma_{\mathbf{k}}$-s have more common elements — consequently they do not cover the entire lattice —, because the equation $s = nr \ mod N$ can have more than one solution for a given gridpoint (s,r).

(iii) If $N = p^l$, with p prime, the $\sigma_{\mathbf{k}}$-s cover the entire lattice if $\mathbf{k} \in \mathbf{K}'$, where $\mathbf{K}' = \mathbf{K} \bigcup \{(n,1) \mid 1 < n < N \text{ and } [n,N] > 1\}$ ([.,.] denotes the largest common factor).

It follows from the above statements that McLachlan's algorithm can be applied in $N = p^l$ cases as well, only the splitting of the Hamiltonian must be modified in the following way: $H = \sum_{\mathbf{k} \in \mathbf{K}'} H_{\mathbf{k}}$, where $H_{\mathbf{k}}(\mathbf{m}) = \dfrac{H(\mathbf{m})}{\chi(\mathbf{m})}$, where $\chi(\mathbf{m})$ gives the number of $\sigma_{\mathbf{k}}$ sets that include the index $\mathbf{m}$.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 102, No. 1, January–March 1998, pp. 43–52

Interpolation of bivariate functions in connection with isoline construction problem

B. G. Vager[1] and **N. K. Serkov**[2]

[1]*Department of Calculating Mathematics,*
St.Petersburg State University of Architecture and Civil Engineering,
198005, 2nd Krasnoarmeiskaya, 4, St. Petersburg, Russia
E-mail: BVager@compmath.abu.spb.ru

[2]*State Hydrological Institute,*
199053, 2nd liniya, V.O., 23, St. Petersburg, Russia

(Manuscript received 2 January 1997; in final form 6 October 1997)

Abstract—The description and critical comparative analysis of the main methods of two-dimensional interpolation for scattered data are given. The new method of interpolation from regular rectangular grid founded on the quadratic B-splines is suggested. The results of numerical experiments are announced.

Key-words: interactive surface tensioning, interpolation, isoline, knot, kriging, scattered data, spline, smoothing.

1. Formulation of the problem

One of the actual problems in applied mathematics which is important in geophysics is the construction of isoline maps. From mathematical point of view the problem is reduced to interpolation of functions $f(x,y)$ whose values f_i are known in given scattered points P_i $(i = 1,\ldots,N)$.

The interpolant S from some given set W is usually required to be smooth, converging to f if the distance among grid points is decreasing to zero. Moreover the physical reasons involve that S should possess "locality", i.e. its value $S(x,y)$ in every point (x,y) should depend on a few number of values f_i in some neighboring grid points only.

The interpolation may be exact if

$$S(P_i) = f_i (i = 1,\ldots,N), \tag{1}$$

or with smoothing if condition (1) does not hold. The latter case usually occurs when the values f_i are the results of measurements which were done with some remarkable errors.

The well-known way to construct isoline maps is to interpolate given values to the grid points of regular rectangular grid first and then to interpolate in an arbitrary point (x,y) by means of one simple method of interpolation, by bicubic splines for example. This is a possibility to achieve a calculative efficacy.

In this paper we consider some methods of two-dimensional interpolation from scattered points in the (x,y)-plane and one new method of interpolation from a regular rectangular grid to an arbitrary point (x,y).

2. Optimal interpolation

In this method the values $f_i(i = 1,\ldots,N)$ are considered as the realization of some stochastic process f. Let stochastic process f be stationary in wide sense, i.e. a mean value $\mathbf{E}f(\mathbf{u})$ is independent from $\mathbf{u}$ and the variance

$$Var[f(\boldsymbol{u}) - f(\boldsymbol{v})] = 2g(\boldsymbol{h}) = E[(f(\boldsymbol{u}) - f(\boldsymbol{v}))^2] \tag{2}$$

depends on the difference vector $\mathbf{h} = \mathbf{u} - \mathbf{v}$ only. If the variance depends on the length $h = |\mathbf{h}|$ only, the stochastic field is called to be isotropic.

Let us formulate the problem of optimal interpolation, or the best linear unbiased estimation (*Kolmogorov*, 1941):

for $P_0 \in R^2$ we have to find weights $p_i \in R(i = 1,\ldots,N)$ minimizing

$$\sigma^2 = \boldsymbol{E}[f(P_0) - S(P_0)]^2, \tag{3}$$

where $S(P_0)$ is the linear combination of known values $f_i = f(P_i)$:

$$S(P_0) = \sum_1^N p_i f_i. \tag{4}$$

A more general problem can also be considered. One may suggest that values f_i have a random errors Δf which are not correlated and with f, having the zero mean value and the measure of error

$$\eta = \mathbf{D}(\Delta f)/\mathbf{D}f, \tag{5}$$

where $\mathbf{D}$ is the sign of dispersion; $\eta = 0$ in the errorfree case.

This method was first used by the French geologist *Krige* (1951) (as referred in the name "kriging") in statistical geophysics, and by the Russian scientist *Gandin* (1963) in climatology. The latter used an autocorrelative function (ACF) $r(x)$ instead of the variogram $g(x)$:

$$r(x) = 1 - g(x)/\mathbf{D}f. \tag{6}$$

The unknown weights p_i in Eq. (4) satisfies the system of linear equations

$$\eta p_i + \sum_{j=1}^{N} \mu_{ij} p_j = \mu_{i0}, \tag{7}$$

where $\mu_{ij} = r(x_i - x_j)$, $(i = 1,\ldots,N;\ j = 0,\ldots,N)$.

The exactness of optimal interpolation may be described by the measure of its error (*Gandin*, 1963)

$$\epsilon = S^2/\mathbf{D}f, \qquad (0 \leq \epsilon \leq 1). \tag{8}$$

The value ϵ may be written as

$$\epsilon = 1 - \sum_{1}^{N} p_i \mu_{i0}. \tag{9}$$

It can easily be shown that

$$\epsilon + \delta = 1, \tag{10}$$

where $\delta = \mathbf{D}f/\mathbf{D}S$ and $\mathbf{D}S$ is the dispersion of the random variable $S(x_0)$.

The equation (10) is equivalent to

$$\mathbf{D}f = \sigma^2 + \mathbf{D}S. \tag{11}$$

This estimation of the smoothing effect of optimal interpolation should be taken into account in the cases where individual values of the interpolant are of interest and their decreasing is not desirable (see also *Kostyukov*, 1982).

Moreover it is suggested that the ACF (or the variogram) is known i.e. there is a sufficiently large number of measurements or/and some a'priori information.

3. *The method of inverse distance raised to a power*

In this method the values of the interpolated function are weighted in some grid points P_i nearest to the point (x,y) of interpolation with the weights which are inverse to a distance raised to the power a:

$$S(x) = \sum_{i=1}^{N} p_i f_i, \tag{12}$$

$$p_i = 1/\left(r_i^a \sum_{j=1}^{N} (1/r_j^a)\right), \quad (0 \leq a \leq 2), \tag{13}$$

where

$$r_j = ||(x,y) - P_j||. \tag{14}$$

Formula (12) involves that $lim\, p_j = \delta_{ij}$ (Kronecker's delta) if $||(x,y) - P_i|| \to 0$.

The value a regulates the influence of distances: the more is a the more quickly the influence of values of the function in the grid points is decreasing as the point of interpolation is being farer and farer from the grid points.

This method can be used as smoothing. In this case the weights are choosen as

$$p_i \approx 1/(r_i^a + b), \tag{15}$$

where $b > 0$ is an empirical parameter (*Kostyukov*, 1982).

The described method has one undesirable property: isolines near the grid points are often distorted ("bull's eye" effect).

Moreover, if there are some grid points which are situated close to each other on one side of the point of interpolation and only one grid point on the contrary side with the same distance, it is evident that "the heap overcomes". This fact is not in agreement with the main physical ideas.

Next method suggested by *Babaliev* (1973) is free from the above mentioned defect.

4. *The weighted anisotropic interpolation (WAI)*

In this manner the method was named by *Kostyukov* (1982). The weights in formula (4) satisfy the system of linear equations:

$$\sum_{i=1}^{N} \pi_j r_{ij} = r_{i0} \qquad (i = 1,...,N) \tag{16}$$

where $r_{ij} = ||P_i - P_j||$, $r_{i0} = ||P_i - (x,y)||$, $p_i = \pi_i/\Sigma\pi_i$.

So, condition (1) holds, and the reciprocal arrangement of data points is taken into account. When $N = 2$ the method is equivalent to the method of inverse distance with $a = 1$ and $b = 0$ in formula (15).

5. Shepard's method

The method is based on ideas of the least squares and inverse distance methods. The interpolant $S(x,y)$ is given in any point $(x,y) \in D$ by the formula

$$S(x,y) = \sum_{k=1}^{N} W_k(x,y) S_k(x,y) / \sum_{k=1}^{N} W_k(x,y), \tag{17}$$

where the nodal functions S_k given by

$$\begin{aligned} S_k(x,y) = {} & A_{1,k}(x - x_k)^2 + A_{2,k}(x - x_k)(y - y_k) + A_{3,k}(y - y_k)^2 \\ & + A_{4,k}(x - x_k) + A_{5,k}(y - y_k) + f_k, \qquad (k = 1,...,N), \end{aligned} \tag{18}$$

are the second degree polynomials of (x,y), which interpolate the data values f_k at grid points (x_k,y_k). The coefficients $A_{1,k}$, ..., $A_{5,k}$ are obtained by a weighted least squares fit to the closest NQ $(4 < NQ < \min\{41, N\})$ data points with weights similar to W_k which are taken to be

$$W_k(x,y) = [(R_k - D_k)_+ / (R_k D_k)]^2, \tag{19}$$

where $a_+ = \max(0,a)$, $D_k(x,y) = ||(x,y) - (x_k,y_k)||$ and R_k is the radius of the influence, which varies with k and is chosen so that the NW $(0 < NW < \min\{41, N\})$ data points are within the radius. Parameters NQ and NW are taken by the user. For a sufficiently large N the recommended values are $NW = 19$ and $NQ \leq 13$.

The value $W_k(x,y)$ is not defined by formula (19) at the grid point (x_k,y_k) but $S(x,y) \to f_k$ as $(x,y) \to (x_k,y_k)$, $(k = 1,...,N)$. It can be shown that $S \in C^1(D)$. A detailed description of the method is given in the paper *Algorithm 660* (1988).

6. *Interpolation by splines*

Let's formulate the problem of seeking for a spline $S \in W_2^2(D)$ which satisfies Eq. (1) and minimizing condition

$$\int_D \left[S_{xx}^2 + 2S_{xy}^2 + S_{yy}^2\right] dx\, dy = \min_{W_2^2(D)}. \tag{20}$$

It may be shown (see *Vasilenko*, 1983, p. 49–50), that this spline is given by

$$S(x,y) = 0{,}5\sum_{i=1}^{N} \lambda_i[(x - x_i)^2 + (y - y_i)^2] \ln[(x - x_i)^2 + (y - y_i)^2] + \nu_{00} + \nu_{10}x + \nu_{01}y, \tag{21}$$

where the unknown coefficients $\lambda_1,\ldots,\lambda_N$, ν_{00}, ν_{10}, ν_{01} satisfy the system of linear equations. This system is given in the monograph (*Vasilenko*, 1983).

The system has the only solution if there are at least three data points not lying on the same straight line. The more the value of N the more the condition number of the system is.

7. *Interpolation by multiquadrics*

Hardy (1971) suggested quadratic functions, or quadrics for modeling of smoothing surfaces. These functions are given by the equation

$$f_i(x,y) = [(x - x_i)^2 + (y - y_i)^2 + b]^{1/2}, \tag{22}$$

where b is the parameter taken by the user. If $b = 0$ the Eq. (22) defines a cone, else it defines a hyperboloid with vertical axes.

The general equation of the surface is the sum

$$H(x,y) = \sum_{i=1}^{N} c_i[(x - x_i)^2 + (y - y_i)^2 + b]^{1/2}. \tag{23}$$

Coefficients c_i are obtained from the system, which arises from interpolation conditions.

The surface constructed by this way is too smooth. The results are good only if data functions are varying slowly.

To obtain the coefficients c_i it is necessary to solve the system with dense matrix. For $N > 1000$ the problem is the storage of the matrix as well as the time of calculations.

8. The Italian method

In this method (*Montefusco* and *Casciola*, 1989) the global interpolant is constructed by smooth "pasting together" of partial interpolants, given in triangles which are defined by chosen triangulation of data point set. Interpolants are constructed provided minimum of some (energetic) functional involved interactive given parameter ("tension parameter"). It gives a strong tool to users to have influence on the approximative surface behavior and allows to smooth undesired oscillations near the points with high gradients.

The algorithm consists of three consecutive steps:

(1) Triangulation of the domain D — the convex hull of the set of data points with the tops of triangulation in these points (*Lawson*, 1977);

(2) Approximation of a first partial derivatives;

(3) Construction of the global interpolant (*Nielson*, 1980).

In step (2) there is possibility for interactive influence on the interpolant by means of the tension parameter to control the form of the interpolation surface and avoid undesired oscillations. The tension parameter is recommended to vary between sufficiently wide limits to achieve a marked effect.

An example for the use of the Italian method for different values of tension parameter, when data are natural, is given below (see *Fig. 1*).

9. Numerical experiments

The objective analysis of wind, geopotential, temperature and air pollution fields was carried out by *Kostyukov* (1982) on the base of extensive natural data. The author made use of least squares polynomial approximation, inverse distance raised to a power, optimal interpolation and WAI methods. Numerical experiments using independent data showed that the best results were received when the two latter methods were applied. WAI was preferred as it does not require a supplementary knowledge of statistical structure of meteorological fields to be interpolated.

We made use of the above mentioned methods and multiquadrics, Shepard, spline functions and Italian methods also. The numerical experiments were carried out on lake depth data. The analysis of results showed that in the case of sufficiently slowly varying functions and approximately regularly placed grid ponts, all of the methods in question led to similar results which were quite good. The analogous conclusion is fair for tests when the grid points are modeled randomly and the function to be interpolated is taken as smooth

analytic function. In real situations when the function to be interpolated has high gradients the interpolant has oscillations which are uncharacteristic of the initial data. The Italian method allows to avoid this defect by a suitable choosing of the tension parameter.

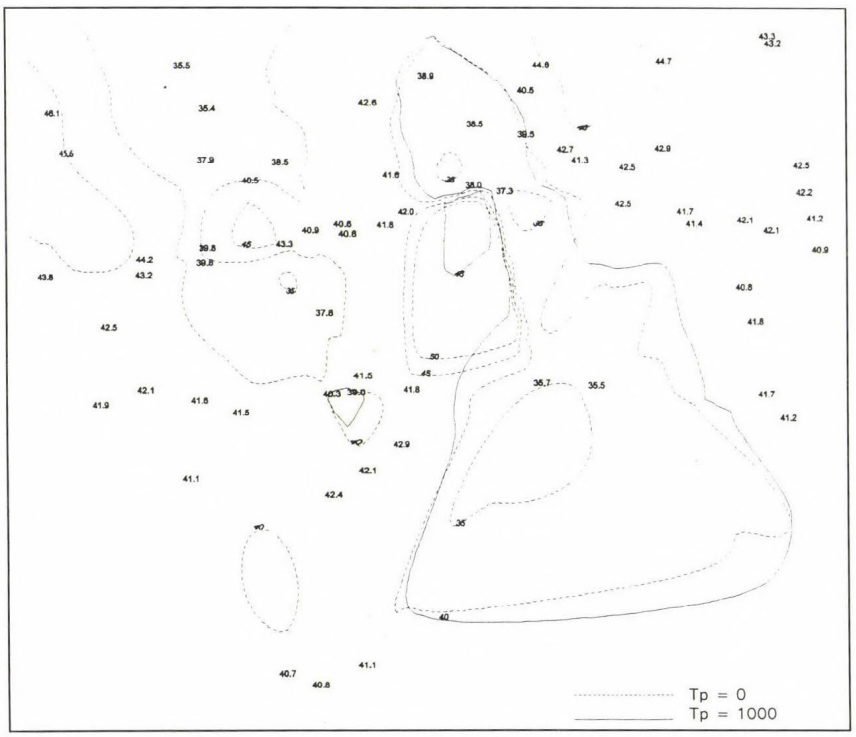


Fig. 1. Isolines of water depth field on the Kama river
(the numbers refer to the measurement points).

10. Interpolation when a grid is regular

In this case bicubic splines may be used for instance (*Vasilenko*, 1983; *Vager* and *Serkov*, 1987). The multiple solutions of the linear equation systems with high dimensional three-diagonal matrices are required to realize the above mentioned methods.

Powell (1977) formulated the interpolation problem of given function f when the data points are situated on a regular rectangular grid. Let the steps in axes x,y are d_x, d_y, respectively. We seek for bivariate function $S(x,y)$ such that

(1) $S(P_i) = f_i \qquad (i = 1,\ldots,N)$;
(2) $S \in C^1(D)$;
(3) $|S(x) - f(x)| = O(d^3)$, where $d^2 = d_x^2 + d_y^2$, if $f \in C^3(D)$;
(4) The property of locality (see the point **1** above) holds;
(5) Isolines $S(x,y) = const$ are easy to construct.

Powell (1977) suggested to construct the interpolant as a spline. His approach comes to a numerical solution of the system of the linear equations which is analogous to that arising from the finite differences solution of the Laplase equation.

We worked out a modification of the Powell's method. The value of the interpolant in each point (x,y) is given as a linear combination of the values f_i in 16 grid points nearest to the point (x,y). Coefficients of this linear combination are the quadratic splines given by **explicit** formulae, so there is no need to solve any systems. This approach allows to decrease the computing time essentially. The formula for $S(x,y)$ has the form

$$S(x,y) = \sum_{1=1}^{4} \sum_{k=1}^{4} \sigma_{i-2+1}(x)\, \sigma_{j-2+k}(y) f_{i-2+1,\, j-2+k}, \tag{24}$$

where if $x \in [x_i, x_{i+1}]$, for instance

$$\sigma_{i-1}(x) = 3\lambda^2/4 - \lambda/2\lambda, \quad \text{if} \quad 0 \le \lambda \le 1/2,$$
$$\sigma_{i-1}(x) = -\lambda^2/4 + \lambda/2\lambda - 1/4, \quad \text{if} \quad 1/2 \le \lambda \le 1,$$

where $\lambda = (x - x_i)/d_x$.

It can be showed that all conditions (1)–(5) are held.

11. Aims of further research

(1) To find the more expedient domains of using of each above mentioned interpolation method on the base of extensive numerical experiments. To set in accordance the various literatural recommendations which are often contradictional. For instance, *Kravchenko* (1984) informs that the using of splines is effective when $N < 150$. The authors of the well-known package of applied programs SURFER recommend $N > 1000$ on the contrary.
(2) To work out an effective method for approximation of bivariate functions in the case of the domain D is not one-connective (it means that there are a number of "holes" within D i.e. a lake with islands). The known methods usually work as there is no difference between "holes" and domains with a sparse arrangement of data points.

References

Algorithm 660, 1988: Collected algorithms from ACM. *ACM Trans. on Mathematical Software 14*, No. 2, 149.

Babaliev, A.M., 1973: About one method of functions of many independent variables (in Russian). *Computer Graphics and Its Application*. Novosibirsk, 118-121.

Gandin, L.S., 1963: *Objective Analysis of Meteorological Fields* (in Russian). Gidrometeoizdat, Leningrad.

Hardy, R., 1971: Multiquadric equation of topography and other irregular surfaces. *J. Geophisical Res. 78*, 1905-1915.

Kolmogorov, A.N., 1941: Interpolation and extrapolation of stationary random sequences (in Russian). *Izv. AN SSSR, Ser. Mathem. 5*, No. 1, 3-14.

Kostyukov, V.V., 1982: *Objective analysis and adjusting of meteorological fields* (in Russian). MO Gidrometeoizdata, Moscow.

Kravchenko, Yu.A., 1984: Methods of modeling of topography surfaces (in Russian). *Review information CNIIGAiK*, Moscow.

Krige, D.G., 1951: A statistical approach to some basic mine problems on the Witwatersrend. *J. Cem. Metall Min. Soc. S. Afr. 52*, 119-139.

Lawson, C.L., 1977: Software for C1 surface interpolation. In *"Mathematical Software III"* (ed.: *J.R. Rice*). Academic Press, Orlando, Fla., pp. 161-194.

Montefusco, L.B. and *Casciola, G.*, 1989: Algorithm 677. C^1 Surface interpolation. *ACM Trans. Math. Soft. 15*, No. 4, 365-377.

Nielson, G.M., 1980: Minimum norm interpolation in triangles. *SIAM J. Numer. Anal. 17*, No. 1, 44-62.

Powell, M.J.D, 1977: Numerical methods for fitting functions of two variables. *State Art Numer. Anal. Proc. Conf.*, New York, 1976, 563-604.

Vasilenko, V.A., 1983: *Spline Functions: Theory, Algorithms, Programs* (in Russian). Nauka, Novosibirsk.

Vager, B.G. and *Serkov, N.K.*, 1987: *Spline Methods in Applied Meteorology and Hydrology* (in Russian). Gidrometeoizdat, Leningrad.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 102, No. 1, January–March 1998, pp. 53–66

Estimation of solar radiation components over Qena/A.R. Egypt at cloudless sky conditions (model verification)

Sayed M. El-Shazly

Department of Physics, Faculty of Science,
South Valley University, Qena/A.R. Egypt; E-mail: svalleyu@frcu.eun.eg

(Manuscript received 27 January 1997; final form 16 June 1997)

Abstract—In this paper three simple, widely applicable models were used to estimate the various components of solar radiation under clear sky conditions at Qena /Egypt. The first model is an empirical one and estimates the global solar radiation as a function of the solar elevation h as: $G = A \sin(h) + B$. The second and third models were developed by *Kasten* (1983) and *Davies* and *McKay* (1982) representing semi empirical and physical models, respectively. The results produced by the models have been compared with measured hourly and daily values. Model performance was assessed from the mean bias error (MBE), the mean absolute error (MAE) and the root mean square error (RMSE). The comparison shows that the global and direct solar radiation are more accurately estimated than the diffuse radiation. With respect to global radiation (G) all the three models perform well, but the first and second ones estimate better. The mean bias error in estimating hourly values of G is smaller using the first model (= zero in all months), while for daily values the Kasten model is better.

With respect to direct solar radiation (I), Kasten model provided better daily and hourly estimates than the Davies and McKay one. In general, the superiority of an appropriate statistical approach against a physical one is obvious. The monthly variation of the model performance indicates somewhat general consistencies between the months, which are characterized by high aerosol vibration and that show high error indices.

Key-words: solar radiation components estimation, statistical models, physical models, aerosol effect.

1. Introduction

As long as precise measurements of solar radiation are absent, one still needs models to estimate various components of it, which are necessary for solar energy utilization and studying of atmospheric conditions. This need is more significant in the developing countries, where these measurements are scarce,

their spatial density is inadequate. The meteorological and engineering literature is replete with such procedures. There are mainly two approaches. The first approach is represented by several empirical and semi empirical models, these are commonly called statistical models. They were developed to satisfy local needs and the users in other locations have to verify or revise the numerical values of the constants and coefficients before applying them. However the semi empirical models have some claim to generality. The second approach is represented by the models based on radiative transfer calculations that simulate the physical processes in the atmosphere, e.g. the attenuation due to the absorption and scattering of water vapor, ozone, Rayleigh and Mie scattering by gases and aerosols. These models are usually referred to as physical models. They require a lot of information about the constituents of the atmosphere, but they do not need solar radiation data and may be applied anywhere.

The aim of this paper is the analysis of the accuracy of two statistical and one physical model in estimating the radiative fluxes on a horizontal surface in Qena/Egypt at cloudless days. A selection was made from models and model forms which may be suitable for general application.

2. *Selected models*

In this paper we use some models, which are considered relatively simple and well suited to the available meteorological information.

The *first model* (*Kasten* and *Czeplak*, 1980; *Holstlag* and *Van Ulden*, 1983; *Van Ulden* and *Holstlag*, 1985; *Sahsamanoglou*, 1991), also the simplest one, estimates the global solar radiation (G) as a function of the solar elevation h, according to the following known relation:

$$G = \mathrm{A}\ \sin(h) + \mathrm{B}, \tag{1}$$

where A and B are coefficients estimated on the basis of hourly global solar radiation using least square method.

The *second model* (*Kasten*, 1983) estimates G in the case of cloudless sky as:

$$G = G_0\, \mathrm{C}\, \exp(-\mathrm{D}\, T_L\, m), \tag{2}$$

where G_0 is the extraterrestrial radiation, T_L is the Linke turbidity factor, m is the relative optical air mass (see Eq. (12)) and C and D have the values C = 0.84 and D = 0.027, based on data over West Germany. T_L is calculated as:

$$T_L = (1 + [(\tau_0 + \tau_w + \tau_a)/\tau_r)], \tag{3}$$

in which τ_0, τ_w, τ_a and τ_r are the spectrally integrated optical depths for ozone

absorption, water vapor absorption, aerosol attenuation and Rayleigh scattering, respectively. Kasten did not include the calculation of direct beam radiation in his model. This modification was made using :

$$I = I_0 \exp(-T_L \tau_r m). \tag{4}$$

The diffuse component is given as:

$$D = I \sin h - G. \tag{5}$$

The *third model* (*Davies* and *McKay*, 1982) calculates the direct (I) and diffuse (D) radiation, both on horizontal surface, as:

$$I = G_0 (T_0 T_r - a_w) T_A, \tag{6}$$

$$D = G_0 [(1 - T_r)/2 + (T_0 T_r - a_w)(1 - T_A) \omega g], \tag{7}$$

where T_0 is the transmissivity after absorption by ozone, T_r is the transmissivity after Rayleigh scattering, a_w is the absorptivity of water vapor, T_A is the transmissivity after extinction by aerosol, ω is the spectrally-averaged single scattering albedo for aerosol and g is the ratio of forward to total scattering by aerosol. Global radiation is expressed as the sum of I and D. The following formulas (*Iqbal*, 1983) have been used for the parameterization of the different models:

$$G_0 = I_0 \cos Z \tag{8}$$

in which I_0 is the corrected value of the solar constant and Z is the solar zenith angle given as:

$$I_0 = I_{sc} [1 + 0.033 \cos(360\, d_n/365)] \tag{9}$$

and

$$Z = \cos^{-1} (\sin \phi \sin \delta + \cos \phi \cos \delta \cos H), \tag{10}$$

where φ is the latitude of the station, δ and H are the declination and hour angles of the sun, respectively.

$$T_0 = 1 - \left[0.1611\, x_1\, (1.0 + 139.48\, x_1)^{-0.3035}\right.$$

$$\left. - 0.002715\, x_1\, (1 + 0.044\, x_1 + 0.0003\, x_1^2)^{-1}\right], \quad (11)$$

in which $x_1 = m\, U_0$, where U_0 is the ozone layer thickness in cm (NTP) taken from *Robinson* (1966) and m is the relative optical air mass calculated using the *Kasten* formula (1966):

$$m = \left[\cos Z + 0.15/(93.885 - Z)^{1.253}\right]^{-1}. \quad (12)$$

$T_r = \exp(-\tau_r\, m)$, in which τ_r is calculated according to *Louche et al.* (1986) as:

$$\tau_r = (6.5567 + 1.7513\, m - 0.1202\, m^2 + 0.0065\, m^3 - 0.00013\, m^4)^{-1}. \quad (13)$$

$$a_w = 2.4959\, x_2 \left[(1 + 79.034\, x_2)^{0.6828} + 6.385\, x_2\right]^{-1}, \quad (14)$$

in which $x_2 = wm$, where w is the perceptible water thickness in cm, calculated with the aid of the *Leckner* (1978) formula:

$$w = 0.493\, (\varphi_r/T) \exp (26.23 - 5416/T). \quad (15)$$

φ_r is the relative humidity in fraction of one and T is the ambient temperature in Kelvin, taken from the Meteorological Department of A.R. Egypt.

$T_A = \exp(-\tau_a\, m)$, in which τ_a is calculated using the following formula (*Freund*, 1983):

$$\tau_a = (-1/m) \ln\left[(I/I_0)/(T_0\, T_r - a_w)\right]. \quad (16)$$

$g = 0.93 - 0.21 \ln (m)$. A constant value of 0.75 was used for ω. τ_0 and τ_w in the Kasten model are calculated with the aid of the above given equations of T_0 and T_w as:

$$\tau_{0,w} = (-1/m) \ln T_{0,w}. \quad (17)$$

3. *General climate of Qena/Egypt*

Qena is located in the south part of Egypt at latitude 26°10'N, longitude 32°43'E and elevation 78 m above sea level. Climatically, Qena lies within the sub-tropical region characterized by warm sunny days and rather cool nights in winter (Dec–Feb) and hot, very dry summers (Jun–Aug). In spring (March-

May) the khamasin depressions are the main features which are associated with high and medium clouds. They can be vigorous and cause severe sand storms raised by strongly southern winds. The study region has almost calm weather, low cloudiness (80% of the days of the year are cloudless) and nearly no precipitation. Analysis of trend values of the most important weather elements, which were taken from the Meteorological Authority of A.R. Egypt, shows the following:

(i) Average temperature ranges from 14.5°C in January to 34°C in July.

(ii) Average relative humidity varies from 21% in May and June to 48% in December.

(iii) Significant percentage of winds is calm ($\cong$52%). The prevailing winds are W, NW, SW, and N, with percentages of occurrence of 15.9%, 11.8%, 11.7% and 4.5%, respectively. The majority of winds range from 2 to 3.1 m s^{-1} and the least occurrence of speed intervals ranges from 8.8 to 10.8 m s^{-1}.

4. Experimental data

Either for empirical and semi empirical models or for quality control of the physical ones, measured values of solar irradiance and other parameters have been used. The hourly values of global (G_h) and diffuse (D_h) solar radiation, which are used in this study were recorded through a program for measuring the solar radiation components, supervised by the author, over Qena/Egypt. A precision pyranometer (Kipp and Zonen Model CM 6B) is used to measure G, another similar one, fitted with a shadow band (radius 620 mm and width 60 mm), constructed by the author following Kipp and Zonen rules, has been used to measure D. Both pyranometers were conjuncted with a two channels solar integrator (Kipp and Zonen Model CC12) to record the values of hourly G_h and D_h in kJ/m^2 as well as daily values of G_d and D_d in MJ/m^2 for a period of 2 1/2 years (Feb 1992–Sep 1994). More details about the experimental techniques and the specification and exposure of the instruments are given in a previous contribution (*El-Shazly*, 1994). From these data, those under clear sky conditions have been selected for this study. The criterion for clear sky condition is the cloudiness to be less than one okta. From G and D data, the corresponding values of I were calculated. The measurements were performed on the roof of a building of the Literature Faculty, South Valley University, Qena/Egypt. There were no evidence of errors in the measured radiation records except for a few cases when D_h exceeds G_h near sunrise and sunset. In these instances the observations were rejected.

5. *Results and discussion*

5.1 Estimation of the coefficients A, B, C, *and* D *in Eqs. (1) and (2)*

The values of A, B, C, and D were estimated on the basis of hourly values of *G, D* and *h* with the aid of the least square method. Because the quantities of water vapor and aerosols in the atmosphere vary from month to month, so different pairs of values of A and B were estimated for each month. *Table 1* summarize the monthly values of A and B in Qena/Egypt. Also, its average values in some other areas are given in *Table 2* for comparison. With respect to C and D in the Kasten formula, they were found to be 0.801 and 0.011, respectively. These values are in a very good agreement with the values estimated by Kasten and in other locations in Egypt as given in *Table 3*. This table shows that the values of C and D are nearly the same for both Egypt and Germany. This result indicates that the considering of the turbidity factor in the empirical relations to estimate the solar radiation components under cloudless sky conditions, may make these realtions quasi independent of the climatological conditions of the stations and have some claim to generality.

Table 1. Monthly values of coefficients A and B (W/m^2) in Qena/Egypt

Month	A	B	R^2
January	1135.1	-69.5	0.920
February	1183.7	-84.0	0.900
March	1087.7	-46.4	0.940
April	1152.7	-97.1	0.990
May	1175.9	-132.9	0.980
June	1162.7	-128.2	0.950
July	1143.9	-112.9	0.960
August	1125.7	-105.9	0.950
September	1076.2	-48.6	0.930
October	1172.6	-106.0	0.920
November	1131.6	-78.0	0.940
December	1108.5	-62.2	0.990

Table 2. Average values of coefficients A and B (W/m^2) at different areas (*Sahsamanoglou* and *Makrogiannis*, 1991)

Area	A	B
Boston (42°13'N/71°07'W)	1098	-65
N. Atlantic (52°30'N/20°W)	1100	-50
Harrogate (54°00'N/1°30'W)	990	-30
Hamburg (53°38'N/9°50'E)	910	-30
De Bilt (52°06'N/5°11'E)	1041	-69
Qena (26°10'N/32°43'E)	1137	-88

Table 3. Values of coefficients C and D in the Kasten formula at different locations

Location	C	D
Hamburg (Germany)	0.840	0.027
Barani (A.R. Egypt)	0.933	0.034
Matruh (A.R. Egypt)	0.879	0.027
El Arich (A.R. Egypt)	0.887	0.032
Tahrir (A.R. Egypt)	0.835	0.027
Cairo (A.R. Egypt)	0.749	0.017
Aswan (A.R. Egypt)	0.864	0.030
Kharga (A.R. Egypt)	0.800	0.018
Qena (A.R. Egypt)	0.801	0.011

5.2 Model performance

Model performance was assessed primarily from the mean bias error (MBE), which characterizes systematic errors, and from the root mean square error (RMSE), characterizing nonsystematic errors. The mean absolute bias error (MAE) was also computed since the MBE may cancel significant positive and negative biases. These indices were calculated for each month for both hourly (kJ/m^2) and daily (MJ/m^2) totals of G, D and I:

$$RMSE = \left[\Sigma (Y_{cal} - Y_{\exp})^2 / N\right]^{0.5}, \tag{18}$$

$$MBE = \left[\Sigma (Y_{cal} - Y_{\exp}) / N\right], \tag{19}$$

$$MAE = \left[\Sigma \left|(Y_{cal} - X_{\exp})\right| / N\right]. \tag{20}$$

Figs. 1a, b and *c* summarize the performance statistics for each flux and model for daily and hourly irradiation in different months of the year. From these figures the following deductions can be made:

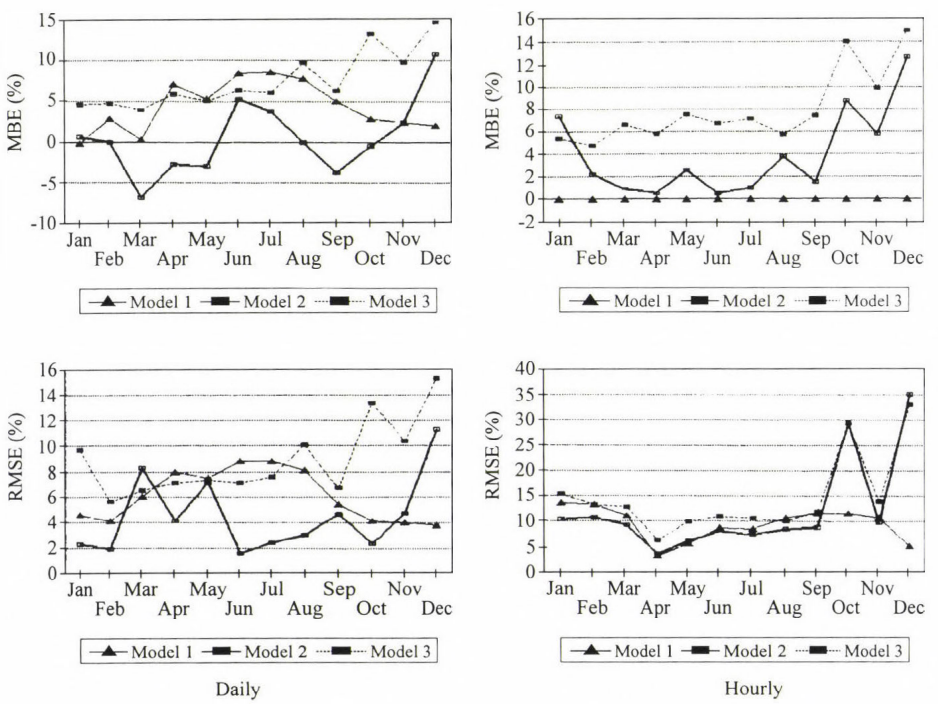
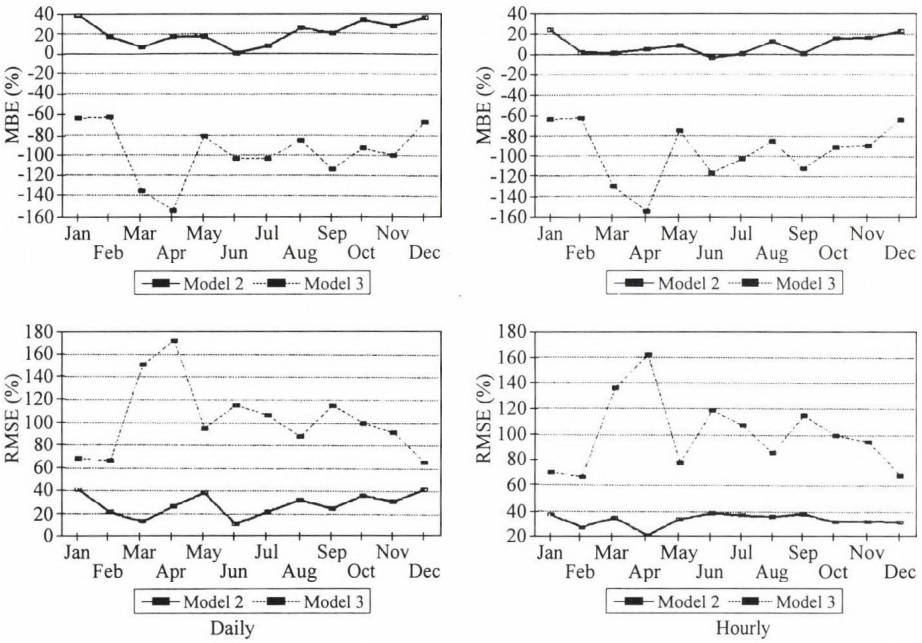
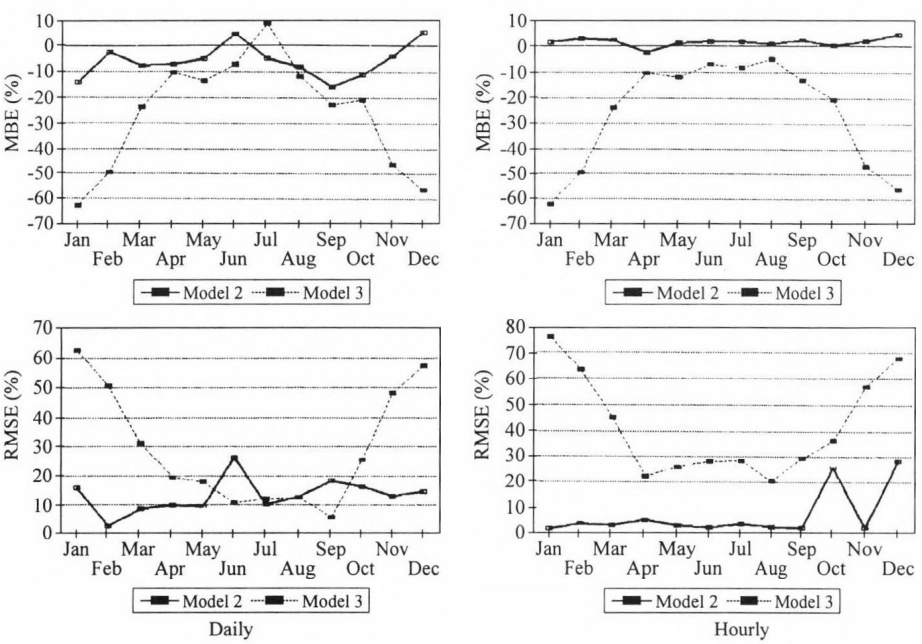


Fig. 1. Monthly variation of error indices in estimating (*a*) global solar radiation; (*b*) diffuse solar radiation; (*c*) direct solar radiation in Qena.

(b)


MBE (%)
RMSE (%)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Model 2
Model 3
Daily
Hourly

(c)


MBE (%)
RMSE (%)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Model 2
Model 3
Daily
Hourly

5.2.1 Global radiation

The three models performed similarly and estimated G with percentage of MBE and RMSE up to 14.7% and 15.3%, respectively, for daily values and 14.9% and 35.2% for hourly ones. However models 1 and 2 estimated the daily and hourly values with the least percentage of errors. The percentage of MBE in daily estimates lie in the ranges (-0.18–8.5), (-6.8–10.7) and (3.96–14.7), while they were (-0.03–0.0), (0.51–12.6) and (4.8–14.9) in hourly estimates, using models 1, 2 and 3, respectively. The corresponding percentages of RMSE were (3.8–8.8), (1.6–11.3) and (5.6–15.3) for G_d and (3.2–13.6), (3.7–35.2) and (6.3–33.2) for hourly values (G_h). As shown in Fig. 1, MBE using the first model is really zero for hourly values, which is a remarkably good result, making this model in favor to the other two models in estimating hourly values of G.

5.2.2 Diffuse and direct solar radiation

With respect to D and I, both daily and hourly, the Kasten model provided better estimates than the Davies-McKay model. However neither of the models performed well for estimating D. As it has been shown in a previous work (*El-Shazly*, 1994), models of Liu and Jordan type realize the best estimation of this radiation component in our region. This conclusion agrees very well with that found by other authors at other locations (*Davies et al.*, 1989).

For I, the Kasten model estimated I with percentages of RMSE in the range of 2.53–18.5 for daily values and 1.47–28.16 for hourly ones (with RMSE of $I_h < 4\%$ for most of the months of the year). RMSE in model 2 reaches 62.5% with most of the values in the range of 10.0–31% for daily estimates and 76% with most of the values in the range of 22.4–29.3 for hourly values. The corresponding MBE were between -15.4 and 5.4 using model 2 and between -62.5 and 8.86 using model 3 for daily estimates, while for hourly estimates it were between -2.4 and 4.56 and from -61.8 to 4.88 using the Kasten and Davies and McKay models, respectively.

For D, the percentage of RMSE for daily and hourly values, respectively, were (12.9–41.8) and (20.5–39) for the Kasten model and (65.9–171) and (67.3–162) for the Davies-McKay model. The MBE percentages were (-0.0–39.34) and (-154.5 – -62.5) and (-3.3–24.1) and (-154.4 – -62.94) for daily and hourly estimates using models 2 and 3, respectively.

As it has been shown above, the accuracy of the statistical approach is better than that of the physical one. This may be explained in view of the inadequacy of the necessary input parameters in the physical model (dust, water vapor and ozone) to estimate the various quantities related to the interaction of solar radiation with the atmosphere. Especial interest must be given to the aerosol effect, which is the major source of error in model estimation for the

real atmosphere as suggested by *Omran* (1989) and confirmed in this paper as will be shown in the next section. Measurements of these parameters with sufficient accuracy for each individual location will improve the accuracy of the physical model to be comparable or better than statistical ones.

5.2.3 Monthly variation of the performance of the selected models

The monthly performance of the three models are shown in the Fig. 1a, b, and c for G, D and I, respectively. From these figures, the following may be concluded:

(i) All models overestimate G, both daily and hourly, except in some months, in which model 2 underestimates its daily values (Mar, Apr, May, Sep and Oct). Model 1 shows no biases in hourly values of G.

(ii) Model 2 overestimated D_d, D_h (except in June) and I_h (except in April), while I_d was underestimated using this model (except in December and June). However both D and I were underestimated using model 3.

Considering the nature of the atmosphere of the study region, we can say that the fluctuation of the error indices through the months of the year may be generally due to the inadequate incorporating of aerosol in the calculations in the different models. In Qena city the main pollutants in the atmosphere are the aerosol dust particles dispersed from the surrounding east and west mountains as well as man activities (*El-Shazly*, 1989). So the attenuation by aerosols in the incoming solar radiation in this region is significant and more than that caused by other constituents (Water vapor, Ozone, and Rayleigh scattering) as it is shown in *Fig. 2*. Also its content varies considerably from day to day and month to month. A general consistency was found between the months, which are characterized by obvious aerosol variation and that show high error indices. Accordingly its effect can not be safely ignored in calculations with models attempting to imitate the physical processes, which attenuate radiation and appear in the different values of solar radiation components at the same Julian day. However there is little empirical information, which can be used in models, and aerosol effects can be incorporated crudely. The different optical characteristics of the aerosol in Qena will be the matter of an empirical study in the near future owing to its very important role in the radiation energy exchanges taking place within the earth-atmosphere system.

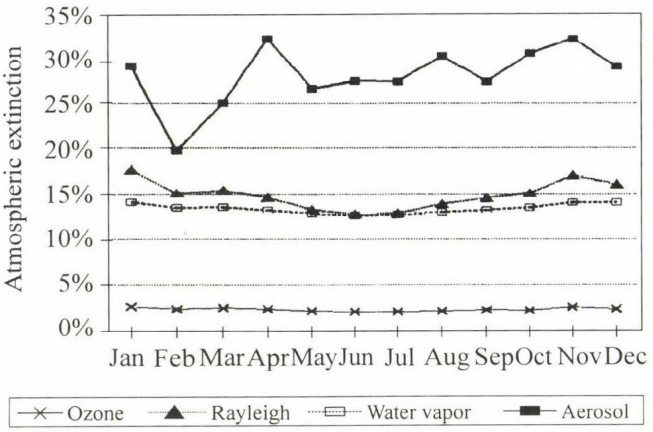


Fig. 2. Variation of the atmospheric extinction caused by different constituents in Qena/Egypt.

5.2.4 Comparison between the performances of the models at different stations

Comparing the model performances at different stations in Egypt and some other countries (*Davies*, 1989) shows the following:

(1) Model 1 provided similar high accuracy in estimating G_d in both Qena/Egypt (MBE = 4.33% and RMSE = 6.09%) and Thessaloniki/Greece (MBE = –0.01% and RMSE = 8.83%).

(2) With respect to the other two models, *Table 4* summarizes the percentage of MBE and RMSE, which were found estimating *G, D* and *I* both daily and hourly at different locations. The values of these indices show some claim to homogeneity in view of the different atmospheric conditions. The following general features may be concluded from this table:

   (a) Both models estimate *G* with smaller errors than *D* and *I* at all stations.

   (b) In Cairo/Egypt the two models perform better in estimating *I* than *D*, which is in contrast to the European stations, where *D* was estimated with slightly more accuracy than *I*.

   (c) In Qena/Egypt model 3 failed to match its performance in the other stations.

   (d) Higher values of both MBE and RMSE were found usually for hourly estimates of all solar radiation components compared to daily ones.

Table 4. Percentage of RMSE and MBE of solar radiation components at different locations

Location		MODEL 2						MODEL 3					
		G		*D*		*I*		*G*		*D*		*I*	
		MBE	RMSE	MBE	RMSE	MBE	RMSE	MBE	RMSE	MBE	RMSE	MBE	RMSE
Qena/Egypt	d	0.5	4.51	20.8	28.3	-5.8	13.2	7.5	8.99	-96.9	103	-26.25	29.5
	h	3.94	12.3	8.7	33.4	1.56	6.83	8	14.79	-95.6	100	-26.13	41.6
Matruh/Egypt	d	-0.5	6.6	-5	35.7			-7.2	10.1	-27	47.1		
El-Arich/Egypt	d	-1.1	12.3	4.7	36.1			-3.8	12.7	-5	37.5		
Cairo/Egypt	d	-0.2	9	0.5	26.5	0.36	18.8	0.32	8.1	-1.8	26	1.4	18.7
Aswan/Egypt	d	0.61	13	12.5	39			2.5	12.8	9.8	38.6		
Kharga/Egypt	d	-1.4	5.8	10	30.5			-3.9	6.4	-4.4	27.5		
De Bilt (The Netherlands)	d	3.8	19.3	-0.5	29.2	1.71	57.8	-1.5	17.7	-18.3	33.8	32	69.4
	h	3.8	39	-2.4	49.7	19.1	102	-1.5	38.7	-18.3	52.5	34.1	112.2
Hamburg (Germany)	d	5	21.8	10.8	38.6	1.8	57.4	1.2	18.8	-13.1	36.7	21.9	62.2
	h	5	35.5	9.2	56.9	1.9	80.3	1.2	33.7	-13.1	57.3	22	86
Kew/UK	d	4.7	20.5	24.9	38.8	-15	50.4	1.5	18.4	9.2	30.7	1.8	43.8
	h	4.7	34.3	24.9	52	-16	78.7	1.5	31.4	8.7	46.1	1.4	73.5

6. Conclusion

An attempt to estimate various components of solar radiation at Qena/Egypt using two main types of models, namely, statistical models (1 and 2) and a physical model (3) leads to the following conclusions:

(i) All selected models estimate the global radiation, both hourly and daily, with satisfactory accuracy, but model 1 and 2 usually better.

(ii) Model 2 estimates the diffuse and direct components better than model 3, which failed to estimate D, both hourly and daily.

(iii) Values of C and D coefficients in equation 2 agree very well with that found for other stations. This indicates the importance of introducing some meteorological parameters in the empirical formulas to improve its accuracy.

(iv) There is a general consistency between months with high variation of aerosol contents and that of large error indices.

(v) The accuracy of the statistical approach is better than that of the physical one, but they can not have general applicability.

(vi) Although model 2 produced generally the best estimates of solar radiation components in Qena, the differences between the statistical measures of errors for the best and the worst performing models may not be sufficient to be significant for solar energy or any other application.

References

Davies, J.A. and *McKay, D.C.*, 1982: Estimating solar irradiance and components. *Solar Energy 29*, 55-64.

Davies, J.A. and *McKay, D.C.* 1989: Evaluation of selected models for estimating solar radiation on horizontal surfaces. *Solar Energy 43*, 153-168.

El-Shazly, M.S., 1994: Solar radiation components at Qena/Egypt. *Időjárás 98*, 47-54.

El-Shazly, S.M., 1989: Studies of the number concentration and size distribution of the suspended dust particles in the atmosphere of Qena/Egypt. *Water, Air and Soil Pollution 45*,121-133.

Freund, J., 1983: Aerosol optical depth in the Canadian Arctic. *Atmosphere-Ocean 2*, 158-167.

Holstlag, A.A.M. and *van Ulden, A.P.*, 1983: A simple scheme for daytime estimation of the surface fluxes from routine weather data. *J. Climate Appl. Meteorol. 22*, 517-529.

Iqbal, M., 1983: *An Introduction to Solar Radiation*. Academic Press, New York.

Kasten, F., 1983: Parametrisierung der Global Strahlung durch Bedeckungsgrad und Trubungs Faktor. *Ann. Met. 20*, 49-50.

Kasten, F., 1966: A new table and approximation formula for the relative optical air mass. *Archive für Meteorol. Geophys. und Bioklim* B, 206-223.

Kasten, F. and *Czeplak, G.*, 1980: Solar and terrestrial radiation dependent on the amount and type of cloud. *Solar Energy 124*, 177-189.

Leckner, B., 1978: The spectral distribution of solar radiation at the earths surface. Elements of a model. *Solar Energy 20*, 143-150.

Louche, A., Peri, G. and *Iqbal, M.*, 1986: An analysis of Linke turbidity factor. *Solar Energy 7*, 393-396.

Omran, M.A.H., 1989: Modeling solar radiation and effects of aerosols over Egypt. *Ph.D. Thesis*, Cairo University (Egypt).

Robinson, N. (ed.), 1966: Solar Radiation. American Elsevier, New York.

Sahsamanoglou, H.S. and *Makrogiannis, T.J.*, 1991: Empirical estimation of daily sky solar radiation in Thessaloniki, Greece. *Theor. Appl. Climatol. 43*, 211-215.

Van Ulden, A.P. and *Holstlag, A.A.M.*, 1988: Estimation of atmospheric boundary layer parameters for diffusion applications. *J. Climat Appl. Meteor. 24*, 1196-1207.

BOOK REVIEW

Wilfried Schröder and *Hans-Jürgen Treder* (eds.): **The Earth and the Cosmos**. *Science Edition, IDCH IAGA/History commission DGG. 382 pages.*

This book is presented to the memory of *Professor Hans Ertel* (1907–1971). The object of this book is the introduction of Hans Ertel as scientist, person and leading manager in science.

This book uses many sources, e.g. correspondence with lot of scholars, his working life, some papers from his wide field of scientific work.

Reading this book we learn that Prof. Ertel made many important discoveries in theoretical meteorology, theoretical geophysics and geophysical hydrodynamics.

First of all Ertel's potential vorticity theorem has become more and more important in meteorology, geophysics, astrophysics, physics and geophysical hydrodynamics.

During the early 1930-ies Ertel addressed the general problem of the thermodynamics and friction of the atmosphere and ocean. His most important contributions were to show the importance of the variation of the Coriolis force with latitude and the theory of long waves in dynamical and synoptic meteorology. He studied the problems of interactions between the stratosphere and troposphere. He emphasized the importance of the advective-dynamics theory of air pressure variations and their periodicities.

In 1939 he published his famous book on theoretical meteorology.

After the Second World War *Rossby* invited Ertel to Stockholm and *Hilding Kueler* to Uppsala. Then and in the time of Rossby's visit in Berlin (1948) it was clear that Ertel's earlier studies had implied the existence of Rossby waves.

On the basis of "Ertel's vortex invariants" derived from the vorticity equations a symmetrical deformation tensor could be constructed which made it possible to make linear transformation of the vorticity components in their total-time derivatives.

One of his very important works: "A relationship between kinematical parameters of horizontal fields of flow in the atmosphere" was published in IDŐJÁRÁS (Vol. 74 (1970), 98-102).

Professor Ertel was a member of the Editorial Board of IDŐJÁRÁS. He was in correspondence with many Hungarian meteorologists, e.g. with *László Aujeszky, Frigyes Dési, György Kozma, Ferenc Rákóczi* and *Alfréd Zách*.

Hans Ertel was a member of the Deutsche Akademie der Wissenschaften since 1949, in 1952 he was chosen as one of its vice-presidents. He was at the

same time Director of the Institute of Meteorology and Geophysics in Berlin, Professor for Geophysics and Theoretical Mechanics at the Humboldt University and Director of the Institute of Physical Hydrography of the Academy of Sciences.

He helped to organize the commemoration of the work of Alexander Humboldt and Max Planck. He worked actively in the program of the Geophysical Year and in the work of the International Hydrological Decade. He was one of the leader scientists of the Carpathian Conference.

Unfortunately, Ertel published all his papers in German or Spanish, in journals which were not very well-known internationally, with the consequence that Ertel's work has not received the recognition it merits. Knowing this fact, we have to agree with *W. Schröder*, who told that Ertel's work "has been overlooked in the English scientific literature. Consequently many 'new' ideas in English books and journals are 'old', because they have been described already by H. Ertel in the years between 1930–1970".

The book consists copies of original articles, xeroxes of handwritings and letters.

We are in agreement with the subtitle of the book, it is really: "The Legacy of Hans Ertel".

Ferenc Rákóczi

ATMOSPHERIC ENVIRONMENT

an international journal

To promote the distribution of Atmospheric Environment *Időjárás* publishes regularly the contents of this important journal. For further information the interested reader is asked to contact *Prof. P. Brimblecombe*, School for Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.; E-mail: atmos_env@uea.ac.uk

Volume 32 Number 1 1998

Conference on the Benefits of the Urban Forest (sponsored by the) Sacramento Municipal Utility District, Sacramento, CA; 7 March 1995

Volume 32 Number 2 1998

NOTES TO CONTRIBUTORS OF *IDŐJÁRÁS*

The purpose of the journal is to publish papers in any field of meteorology and atmosphere related scientific areas. These may be

- reports on new results of scientific investigations,
- critical review articles summarizing current state of art of a certain topic,
- shorter contributions dealing with a particular question.

Each issue contains "News" and "Book review" sections.

Authors may be of any nationality, but the official language of the journal is English. Papers will be reviewed by unidentified referees.

Manuscripts should be sent to
Editor-in-Chief of IDŐJÁRÁS
P.O. Box 39
H-1675 Budapest, Hungary

in three copies including all illustrations. One set of illustrations has to be of camera ready quality, the other two might be lower quality.

Title part of the paper should contain the concise title, the name(s) of the author(s), the affiliation(s) including postal and E-mail address(es). In case of multiple authors, the cover letter should indicate the corresponding author.

Abstract should follow the title, it contains the purpose, the data and methods as well as the basic conclusion.

Key-words are necessary to help to classify the topic.

The text has to be typed in double spacing with wide margins. Word-processor printing is preferred. The use of SI units are expected. The negative exponent is preferred to solidus. Figures and tables should be consecutively numbered and referred to in the text.

Mathematical formulas are expected to be as simple as possible and numbered in parentheses at the right margin. Non-Latin letters and hand-written symbols should be indicated and explained by making marginal notes in pencil.

Tables should be marked by Arabic numbers and printed in separate sheets together with their captions. Avoid too lengthy or complicated tables.

Figures should be drawn or printed in black and white, without legends, on separate sheets. The legends of figures should be printed as separate list. Good quality laser printings are preferred as master copies.

References: The text citation should contain the name(s) of the author(s) in Italic letter and the year of publication. In case of one author: *Miller* (1989), or if the name of the author cannot be fitted into the text: *(Miller,* 1989); in the case of two authors: *Gamov* and *Cleveland* (1973); if there are more than two authors: *Smith et al.* (1990). When referring to several papers published in the same year by the same author, the year of publication should be followed by letters a,b etc. At the end of the paper the list of references should be arranged alphabetically. For an article: the name(s) of author(s) in Italics, year, title of article, name of journal, volume number (the latter two in Italics) and pages. E.g. *Nathan, K.K.,* 1986: A note on the relationship between photosynthetically active radiation and cloud amount. *Időjárás 90,* 10-13. For a book: the name(s) of author(s), year, title of the book (all in Italics except the year), publisher and place of publication. E.g. *Junge, C. E.,* 1963: *Air Chemistry and Radioactivity.* Academic Press, New York and London.

The final version should be submitted on diskette altogether with one hard copy. Use standard 3.5" or 5.25" DOS formatted diskettes. The preferred word-processors are WordPerfect 5.1 and MS Word 6.0.

Reprints: authors receive 30 reprints free of charge. Additional reprints may be ordered at the authors' expense when sending back the proofs to the Editorial Office.

More information: gmajor@met.hu
Information on the last issues:
http://www.met.hu/firat/ido-e.html

Published by the Hungarian Meteorological Service

Budapest, Hungary

INDEX: 26 361

HU ISSN 0324-6329