

Preliminary comparative assessment and elements of equivalence of air pollution measurement results of portable monitoring stations with using stochastic models

Tomasz Owczarek^{1*}, Mariusz Rogulski² and Artur Badyda²

¹Akademia Morska w Gdyni, Wydział Przedsiębiorczości i Towaroznawstwa, ul. Morska 83, 81-225 Gdynia, Polska

²Warsaw University of Technology, Faculty of Building Services, Hydro and Environmental Engineering, Nowowiejska 20, 00-653 Warsaw, Poland

Abstract. This article is an attempt to assess the suitability of new non-reference mobile measuring devices for the analysis of PM₁₀ concentrations. The aim is to compare the concentration daily values measured by these devices with the concentrations obtained by the reference device. It also examines the possibility of building models that correct these values to equivalent reference values. The analysis allows to conclude that the properties of the obtained measurements indicate a good chance of constructing an effective model of correcting the concentration values. Comparison of frequency, time variation and correlation of concentrations indicates the ability to demonstrate the equivalence of mobile devices with the reference method.

1 Introduction

The quality of the air and its impact on the environment, in particular on human health, is an issue that is driving more and more widespread groups of society. Therefore, the problem of measuring concentrations of pollutants in a fast, cheap and, above all, accurate and effective way is a problem for scientists and companies that produce measuring devices.

Among the air pollutions, particulate matters occupy a special place. They are a mixture of particles of different structure, origin and size. They may be of natural origin (arising from erosion, evaporation or natural physicochemical transformations) or anthropogenic (fuel combustion, communication emissions, industrial emissions) [1, 2]. The chemical composition of the particulate matter can vary greatly and depends largely on where it is located [3]. It may contain heavy metals (lead arsenic, nickel, cadmium), polycyclic aromatic hydrocarbons (especially benzo(a)pyrene) and anions (eg. SO₄²⁻, NO₃, Cl⁻) and cations (eg. Na⁺, K⁺, Ca²⁺, Mg²⁺, NH₄⁺) (for example [2, 4-6]).

* Corresponding author: t.owczarek@wpit.am.gdynia.pl

There are three fractions of particulate matter depending on the dust particle size: PM_{10} (dust containing particles up to 10 μm in diameter), $PM_{2.5}$ (up to 2.5 μm in diameter) and PM_1 (up to 1 μm in diameter). Studies have shown that the presence of dust in the air has a negative impact on human health [7]. This is especially true for smaller particulate matters diameter. Thicker fractions of dust can settle in the lungs and entire upper respiratory tract, while smaller particles can penetrate the lungs into the bloodstream and enter the brain, heart and other organs. They can cause upper respiratory tract diseases (asthma, COPD, dyspnea), but also cause heart disease, nervous system, neurosis and many others [8, 9].

The concentration of pollutants in the air is examined with using methods described in detail in the legislation. For the concentration of particulate matter, the gravimetric method is used as the reference method. This method is considered as a reference method and as such is considered to be the most accurate method [10, 11]. Unfortunately, due to the method of data collection, the results for this method are obtained once a day. In addition, the research is relatively expensive, there is a limited number of measuring devices (currently in Poland is about 200), and the measurements take place in very carefully selected places. As an alternative to measuring by reference method are measurements made by using automatic measuring devices.

Different measurement methods are used for their construction: with using a tapered element oscillating microbalance (TEOM), β -radiation attenuation or light scattering (optical method). The advantage of these devices is their relatively low cost and high flexibility of measurements. The disadvantage is unknown level of measurement accuracy. Therefore, in order to be able to use the measurement results from these devices, it is necessary to demonstrate their equivalence with the results obtained by the reference method. Equivalence procedures include a comparative study of the results obtained by both methods and the risk assessment of the results of severely deviated reference results [1, 12, 13].

The aim of the study is to compare the daily values of PM_{10} concentrations measured with using five new, mobile devices based on optical sensors with concentration values obtained by a measuring device owned by the Voivodeship Environmental Protection Inspectorate (VIEP), treated as reference values.

2 Methodology

The comparison of PM_{10} concentrations from the reference device and mobile devices was made both in the frequency domain and in the time domain. The purpose was to determine whether the values obtained by the test devices and the reference device are correlated, whether they have a similar frequency structure and whether the changes in time are similar. Confirmation of these facts will increase the likelihood of test measurements with the reference measurements and will allow further research to confirm the equivalence of new devices with the reference method, after the application of equations or correction factors.

The PM_{10} concentration measurements were conducted in Nowy Sącz. The VIEP reference device (Ref) and two mobile devices (U2 and U4) were located in the city center at a relatively short distance from each other. The other three units were located in the suburbs of the city, one to three kilometers from the center, in urbanized locations (U1 on the north, U3 on the south-west and U5 on the south-east of the center). The location of the equipment was dependent on the decision of the local authorities, which co-organized this tests. Because the devices were not close to each other or in the immediate vicinity of the reference device, it

was not possible to compare directly the concentration values. For comparison time series analysis was used to describe the behavior of the data and to detect its similarities.

The study used data from measurements from September 2016 to September 2017. Devices provide data on PM₁₀ concentrations hourly, but for comparison with the reference device and later with the reference method, the measurements were aggregated to the daily values by calculating the 24 hourly averages. 376 observations were thus obtained, which were used in further analysis.

The analysis of the measurement series used descriptive statistics methods such as graphs, means and standard deviations as well as Pearson correlation coefficients. The Principal Components Analysis (PCA) was used to verify the similarity of the concentration factors obtained on all devices. PCA's goal is to rotate a multidimensional coordinate system that describes factors affecting the magnitude of the concentration to maximize the variability (variance) of the particular coordinates. The results of this analysis allow us to assess whether each of the series describing the concentrations affects the same factors or whether they are formed by a different set of external factors.

One of the analyzed elements of the obtained PM₁₀ concentrations is their comparison with the frequency function. The similarity of periodic variations of the obtained series may indicate the convergence of the concentrations behavior obtained on the mobile devices with the measurement results of the reference device. In such a situation, to obtain convergent values with a device equivalent to the reference, only the correction using a linear model is sufficient.

To compare the time series in the frequency domain, used model based on Fourier series. A transform is used to divide a periodic function (a series of measurements) into a finite sum of trigonometric functions. This allows for a very precise description of the behavior and consequently a comparison of the observed PM₁₀ concentrations. Fourier series is a sum of trigonometric functions whose coefficients are determined for each frequency separately. The periodogram is the sum of the squares of the coefficients calculated for specified frequency:

$$P_k = a_k^2 + b_k^2 \quad (1)$$

Periodogram can be interpreted as a variance corresponding to fluctuations of a particular frequency or period. A detailed description of the Fourier series is in [14].

One of the important part of the analysis of the PM₁₀ concentration series is the question of whether the series is sufficiently informative to be able to describe it in terms of mathematical models and possible predictions of future values. The tool used for this purpose is an autocorrelation analysis, it means analysis of the relationship of series values with the same series delayed for a certain time.

In the study is also used the analysis of total and partial autocorrelation. Total autocorrelation of the order t is the occurrence of the correlation between the values of the series and the values of the series delayed by t periods. In the case of autocorrelation of the order t , the occurrence of autocorrelation is investigated, after the effect of all autocorrelation of rows 1 to $t-1$ is removed. Demonstration of the autocorrelation allows us to determine the relationships that occur in a series that can be interpreted as a charge of information or a memory of a series. This information can be used to model or predict a series value. Detailed information on the statistical methods about described statistical methods can be found in [15].

3 Results

In the analysis of the series of observed concentrations of PM₁₀, the behavior of the concentration values in the whole examined period was compared first. The observed concentrations are shown in figure 1.

Based on figure 1, it can be stated that in the cold months PM_{10} concentrations were high, reaching up to 200-300 $\mu g/m^3$ in January and February, with high variability of the obtained concentrations. It is also characteristic for this period of considerable divergence in the measurements obtained on different devices.

Lowest concentrations were obtained in warm months. The PM_{10} concentrations in June and July were between 5 and 25 $\mu g/m^3$. In this period, they also had the lowest variability and the highest similarity of values.

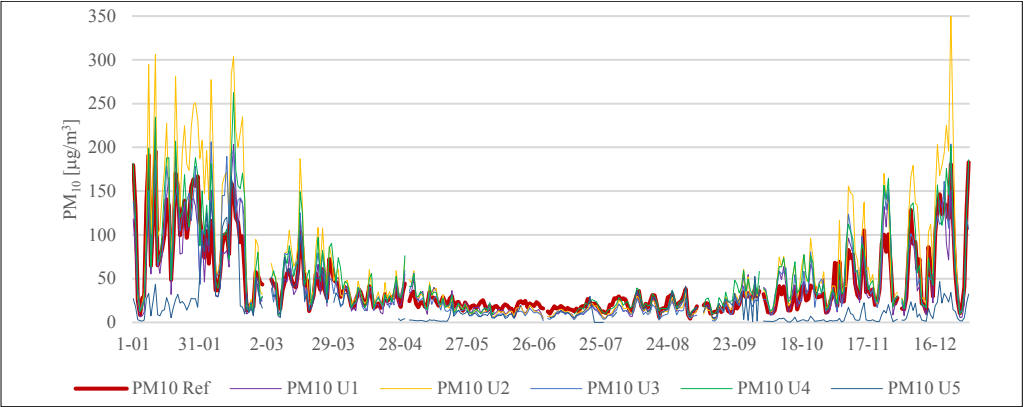


Figure 1. PM_{10} measurement results in Nowy Sącz (source – own study based on own research results)

A very low PM_{10} concentration is noted on the device number 5. This device is located in a part of the city, which is very few buildings and away from the busy roads. The location of the device was most likely to cause such low concentrations.

In order to eliminate influences in the same way on the tested devices as for the reference device, analysis was performed on the difference in concentrations. The comparison of the obtained differences is shown in figure 2.

It is noticeable high value of differences in the cold months disappearing in the warm months. In the case of three devices, however, these differences do not take up large values. In the case of two devices, especially in colder months, there are large differences of the systematic nature. In the case of the U2 device, the differences in most cases are negative (the values obtained on the device are higher than the value from the reference device), while in the case of the U5 meter the positive values are dominant (the device indicates significantly lower values than the reference device).

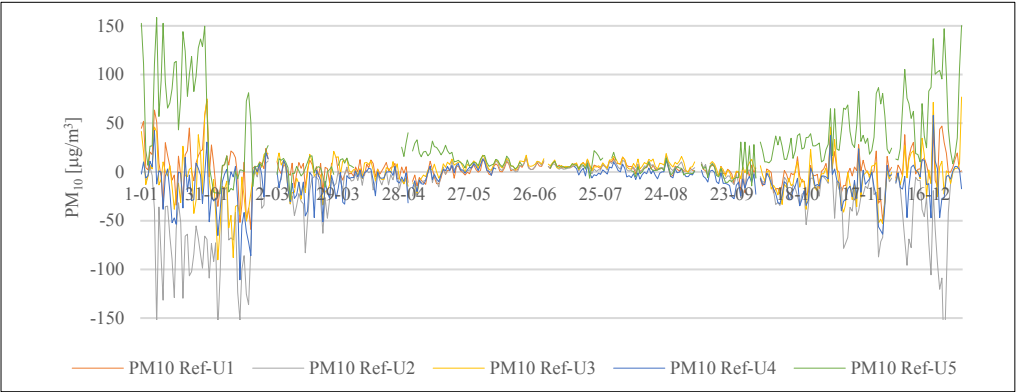


Figure 2. Differences in PM_{10} concentration in VIEP station and tested devices (source – own study based on own research results)

Hypothesis about the similarity of PM₁₀ concentration changes obtained on mobile devices and reference device can be confirmed by estimating and comparing the values of linear correlation coefficients between the measured values of these devices. Correlation coefficients are shown in figure 3.

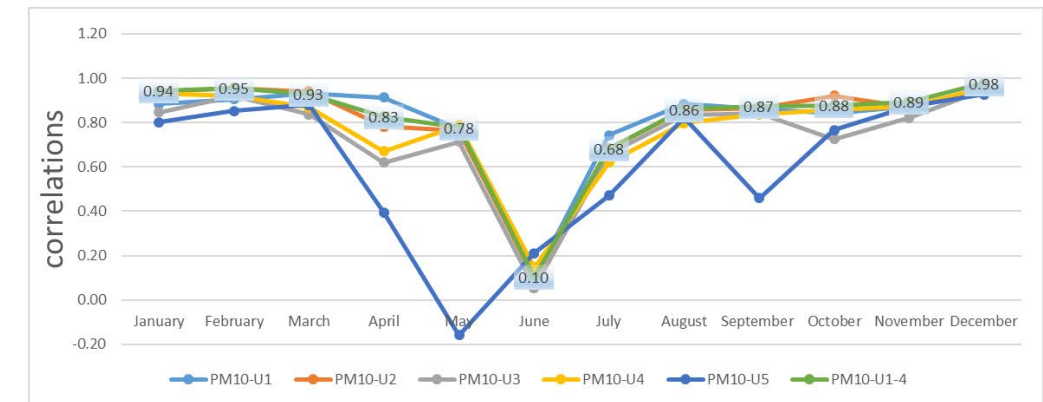


Figure 3. Correlation coefficients of measurements from mobile devices with measurements of reference device (source – own study based on own research results)

The values of correlation coefficients in the vast majority of the studied months are high, ranging from 0.8 to close 1. This is consistent with changes in PM₁₀ values measured on the reference device and on mobile devices. The correlation coefficients for the U5 is different ($r=0.1$). Due to the specific location, they were removed and average correlation coefficients for other devices (U1-U4) were calculated. Their values (placed in frames in Figure 3) indicate a very high compliance of observations from mobile devices and reference device.

The exception to the above conclusions is June, in which the values of all correlation coefficients have practically approached zero. Such an abnormal situation may be explained either by a very low volatility of data from this month (standard deviations from 3.21 to 4.24).

Further part of the analysis was devoted to the identification of factors influencing the measured values of PM₁₀ concentrations. The results of Principal Components Analysis is shown in figure 4 and table 1.

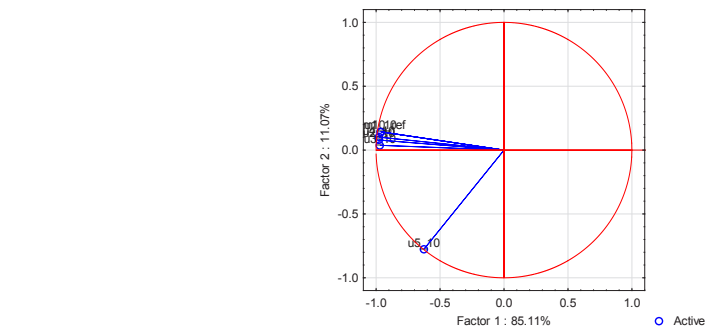


Figure 4. Projection of PM₁₀ concentrations on the two-dimensional plane of the factors in the principal components analysis (source – own study based on own research results).

The results of the main components analysis indicate that for all the mobile devices except the U5 device had the same impacts as for the reference device. The compatibility of the factors in the analysis is confirmed in table 1.

Principal component model results for all devices, except U5, have a similar factors (PC1 from -0.98 to -0.96, PC2 form 0.04 to 0.14), indicating that the same concentration factor values were used for the measurements of the concentrations from these devices. Even if the concentration values had different values, they were affected by the same factors and in the same way.

Table 1. Main component loads for PM₁₀ concentrations.

<i>Model PCA</i>	<i>Ref</i>	<i>U1</i>	<i>U2</i>	<i>U3</i>	<i>U4</i>	<i>U5</i>	<i>Own values</i>	<i>Percentage of the total</i>	<i>Total</i>
PC1	-0.96	-0.97	-0.98	-0.97	-0.98	-0.63	5.11	85.1%	96.2%
PC2	0.14	0.14	0.08	0.04	0.10	-0.78	0.66	11.1%	

Source. Own study based on own research results.

The obtained model of the main components effectively describes the variability of the analyzed feature. Two components with the highest charges (own values for PC1=5,11, and for PC2=0,66) explain the variation of concentrations in over 96%. However, it is not possible to give their interpretation, in other words – identified factors.

To eliminate the main factors affecting PM₁₀ concentrations during the investigated period, the main components were analyzed for the result of subtraction of the concentrations obtained on mobile devices and reference device. The results of PCA analysis are shown in table 2.

Table 2. Principal component loads for subtractions of PM₁₀ concentrations.

<i>Model PCA</i>	<i>Ref-U1</i>	<i>Ref-U2</i>	<i>Ref-U3</i>	<i>Ref-U4</i>	<i>Ref-U5</i>	<i>Own values</i>	<i>Percentage of the total</i>	<i>Total</i>
PC1	-0.72	-0.69	-0.82	-0.92	0.17	2.54	50.9%	82.2%
PC2	-0.49	0.65	-0.28	-0.02	-0.91	1.57	31.3%	

Source. Own study based on own research results

The results of the analysis indicate a greater divergence of factors describing differences in PM₁₀ concentrations. Two components PCA model (own values 2.54 and 1.57) describes variation in 82.2%. After eliminating the main aggregate at the concentrations obtained on the reference device, the remaining factors do not allow for clear conclusions. However, similar results of U1, U3 and U4 observations, as well as slightly different behavior of U2 and U5 devices, can be observed. In the whole study, the U5 behaves differently from the rest of the devices, while the behavior of the U2 is considerably more subtle, and only appears after the removal of the major concentration factors.

The periodogram values, obtained by Fourier spectral analysis, for PM₁₀ concentrations observed on all devices, together with the correlation coefficients for mobile devices and reference device, show a high likelihood of the series in terms of frequency distributions. Pearson correlation coefficients for periodograms range from 0.88 to 0.98. The PM₁₀ concentration components for all analyzed devices have the same structure and are very similar[†].

[†] Due to the limited size of the article, detailed Fourier analysis results are not presented

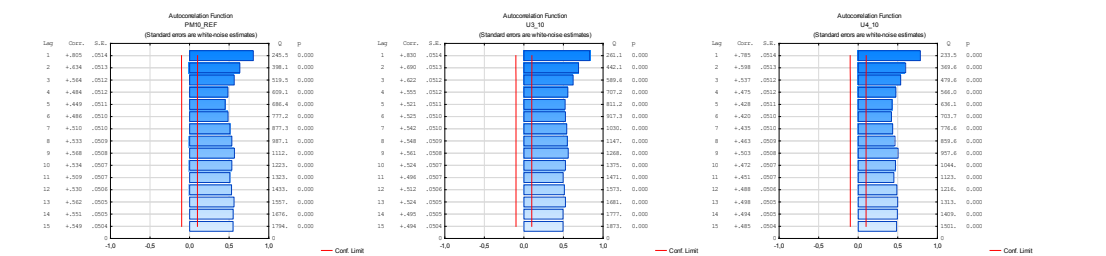
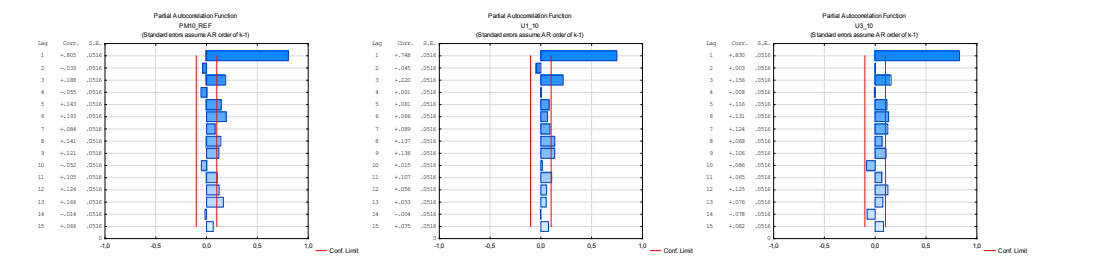


Figure 5. Selected autocorrelation charts of PM₁₀ concentration from mobile devices (source – own study based on own research results).

Results of autocorrelation analysis of PM₁₀ concentrations, for tree selected devices, are shown in figure 5.

As the result of the analysis observed strong correlations between the test series and the delayed series for at least 15 days. Values of one-day delay are from 0.75 to 0.83. With larger delays, autocorrelation values are smaller but decrease only to a certain level. With delays of 12 to 15 days, correlation coefficients take values close to 0.5.

The autocorrelation analysis is supplemented by partial autocorrelation analysis. Sample results of partial autocorrelation for reference device and devices U1 and U3 are shown in figure 6.



Graph 6. Partial autocorrelation selected charts of PM₁₀ concentrations from mobile devices (source – own study based on own research results).

Partial autocorrelation analysis showed strong autocorrelation with one-day delayed sequences and statistically significant autocorrelation with delayed series of 3, 7 or 8 and 12 or 13 days. It is therefore possible to assume that the information contained in the series of PM₁₀ concentrations can be used to construct a valid mathematical model describing these concentrations.

4 Conclusions

The PM₁₀ concentrations obtained on mobile devices differ noticeably from the results obtained at the reference station in the cold months (average results in January of references station 113.21 µg/m³, mobile devices – from 97.74 to 162.01 µg/m³). In warm months these differences are small (in July: Ref – 16.85 µg/m³, and mobile: from 9.45 to 17.10 µg/m³). As the concentration decreases, their variability decreases, so that in the warm months it is at a very low level (standard deviation of results in January from 12.33 to 81.66 µg/m³, in June – 3.51 to 4.24 µg/m³). In the warmer months, there were also greater divergences between the concentration indications for different measuring devices (average standard deviation of results in January 45.17 µg/m³, in June – 3.61 µg/m³).

There is a clear relationship between the concentration measurements obtained by both methods. The correlation coefficients of PM₁₀ concentrations obtained with using the

reference method and mobile devices calculated for individual months, except for June, have high values, from 0.68 to 0.98. The values of these indicators show that although the concentration values obtained by the different devices are not the same, there is a strong correlation between them that causes the concentration values to change in a similar way.

These observations also confirm the results of the series analysis in terms of frequency of changes over time. The results of the series decomposition into the periodograms corresponding to the individual oscillation frequencies by Fourier analysis show very high convergence in this respect. Correlation coefficients calculated for periodograms (values from 0.88 to 0.98) show that the structure of fluctuations in PM₁₀ concentrations obtained by different methods is very close to one another.

The main components analysis proves that the values obtained by the majority of devices are affected by the same or similar factors. The charge values of the components are very similar to each other, and the resulting model translates a decisive part of the variability of the observed concentrations.

The analysis of autocorrelation the series of concentrations shows that there is a large autocorrelation with series delayed by many periods (from 0.75-0.83 for 1-day delay, to 0.5 for 15-day delay). It can be concluded that, thanks to the charge of information on previous events, stored in series, it is possible to use these data to build an effective model describing the behavior of PM₁₀ concentrations.

References

1. J. Gębicki, K. Szymańska, *Atmos. Environ.* **54**, (2012)
2. G. Dongarrà, E. Manno, D. Varrica a, M. Lombardo, M. Vultaggio, *Atmos. Environ.* **44**, (2010)
3. A. Zwozdziak, L. Samek, I. Sowka, L. Furman, M. Skretowicz, *Sci. World. J.* (2012)
4. G. Majewski, W. Rogula-Kozłowska, P.O. Czechowski, A. Badyda, A. Brandyk, *Atmosphere*, **6**, (2015)
5. W. Rogula-Kozłowska, G. Majewski, P.O. Czechowski, P. Rogula-Kopiec, *EPE* **43**, 1 (2017)
6. *Ustawa prawo ochrony środowiska*, 27 kwietnia 2001r. Dz.U. z 2017 poz. 519
7. *Health risk assessment of air pollution. General principles* (2016), <http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2016/health-risk-assessment-of-air-pollution.-general-principles-2016>
8. A. Badyda, J. Grellier, P. Dąbrowiecki, *Adv. Med. Exp. Bio.* **944** (2016)
9. A. Badyda, P. Dąbrowiecki, P.O. Czechowski, G. Majewski, *Respir. Physiol. Neurobiol.* **209**, (2014)
10. *Powietrze atmosferyczne -- Standardowa grawimetryczna metoda pomiarowa do określania stężeń masowych frakcji PM₁₀ lub PM_{2,5} pyłu zawieszonego*, PN-EN 12341:2014, 24 lipca 2014r.
11. *Dyrektywa Parlamentu Europejskiego i Rady w sprawie jakości powietrza i czystsze go powietrza dla Europy*, 2008/50/WE, 21 maja 2008 r.
12. J. Gębicki, K. Szymańska, *Pol. J. Environ. Stud.* **20**, 6 (2011)
13. *Ambient Air – Automated measuring systems for the measurement of the concentration of particulate matter (PM₁₀; PM_{2,5})*, CEN/TS 16450, European Committee for Standardization (2013)
14. G.M. Fichtenholz, *Rachunek różniczkowy i całkowity*, Tom 3 (Państwowe Wydawnictwo Naukowe, Warszawa, 1985)
15. A. D. Aczel, *Complete Business Statistics* (Richard D. Irwin, Inc., 1989)