I. MAN AND BIOSPHERE

ON SOME RELATIONSHIPS BETWEEN ECOLOGY AND HEALTH

Alexander Sadovski

Abstract: Three mathematical and statistical methods - namely correlation, Redundancy and Procrustean analysis were applied to analyze relationships between ecological factors of environment and health indicators. They used data from the International Organization for Environment Protection (UNEP) and the World Bank (World Bank). Environmental factors include emissions of various pollutants - N_2O , NO_x , PM_{10} , CO_2 , and economic factors such as renewable energy resources. Of the indicators of health are considered life expectancy, mortality, morbidity from diabetes and tuberculosis. The results provide a basis for a series of discussions and conclusions.

Key words: ecological factors, health indicators, correlation, Redundancy analysis, *Procrustean analysis*

INTRODUCTION

The most famous modern definition of health was created during a Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (*Official Records of the World Health Organization, no. 2, p. 100*) and entered into force on 7 April 1948: **"Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity"** The Definition has not been amended since 1948.

During the Ottawa Charter for Health Promotion in 1986 (WHO, 1986) WHO said that health is: **"a resource for everyday life, not the objective of living. Health is a positive concept emphasizing social and personal resources, as well as physical capacities."**

The health of individual people and their communities are affected by a wide range of contributory factors. People's good or bad health is determined by their environment and situations - what is happening and what has happened to them, says WHO. WHO says that the following factors probably have a bigger impact on our health than access and use of health care services:

- Where we live
- The state of our environment
- Genetics
- Our income
- Our education level
- Our relationship with friends and family.

Ecology is the scientific analysis and study of interactions among organisms and their environment. It is an interdisciplinary field that includes biology, geography, and Earth science. Ecology includes the

study of interactions organisms have with each other organisms, and with abiotic components of their environment.

Environment literally means surrounding and everything that affect an organism during its lifetime is collectively known as its environment. In another words **"Environment is sum total of water, air and land interrelationships among themselves and also with the human being, other living organisms and property"**. It includes all the physical and biological surrounding and their interactions.

Components of the environment are:

PHYSICAL: air, water, soil, housing, climate, geography, heat, light, noise, debris, radiation, etc.

BIOLOGICAL: man, viruses, microbial agents, insects, rodents, animals and plants, etc.

PSYCHOSOCIAL: cultural values, customs, beliefs, habits, attitudes, morals, religion, education, lifestyles, community life, health services, social and political organization.

Environmental health is the study and management of environmental conditions that affect the health and well-being of humans.

The people are changing their environment. The release of new forces, the development of new substances, the induction of new behaviour patterns, the creation of new desires has changed man's environment.

Day by day he faces conditions that are new to his collective experience; circumstances that the cleverest minds before have foreseen but dimly; effects that could not have been predicted. The operations he performed upon his environment were intended to develop a situation advantageous to him; but through inadequate pre-knowledge, limited thinking, or plain mischance the unexpected happened (Lee DHG., 1964).

The purpose of the present study, which is an example of environmental health study, to apply mathematical and statistical methods in order to reveal relationships between ecological factors of environment and health indicators.

MATERIAL AND METHODS

Materials for conducting our study are reliable public available date from two sources.

1. Data from World Bank for 248 countries (see World development indicators):

Life expectancy at birth, total (years), Diabetes prevalence (% of population ages 20 to 79), Incidence of tuberculosis (per 100,000 people), Prevalence of HIV, total (% of population ages 15-49), CO₂ emissions (metric tons per capita), Methane emissions (kt of CO₂ equivalent), PM₁₀ air pollution, mean annual exposure (micrograms per cubic meter), Renewable energy consumption (% of total final energy consumption), Urban population (% of total), Health expenditure, total (% of GDP), Immunization, DPT (% of children ages 12-23 months).

2. Data from UNEP Environmental Data Explorer for 240 countries (UNEP, 2016):

Life expectancy (years), Crude Birth Rate (per Thousand people), Crude Death Rate (per Thousand people), Maternal Mortality Ratio (Deaths per 100 000 Births), Under-Five Mortality Rate (Deaths per Thousand Births), Infant mortality (probability of dying between birth and exact age 1, expressed as

deaths per 1,000 births), Diseases of the Respiratory System (Number of Deaths), Reported Malaria Cases (Total Number), Emissions of CH₄ - (National Reports, UNFCCC), Emissions of CO - (National Reports, UNFCCC), Emissions of CO₂ - (National Reports, UNFCCC), Emissions of SO₂ (National Reports, UNFCCC), Emissions of N₂O (National Reports, UNFCCC), Emissions of NO_x (National Reports, UNFCCC), Concentrations of Particulate Matter less than 10 microns (PM₁₀), Consumption of Ozone-Depleting Substances, Emissions of SO₂ - Total (Model Estimations, EDGAR v.4.2), Emissions of NO_x - Total (Model Estimations, EDGAR v.4.2), Emissions of HFCs (all Gases).

First step of our study is a Descriptive statistical analysis presented in the following Tables 1 and 2.

Looking to the Tables, it is evident that the UNEP set contains many missing values for some of emissions in comparison with World Bank data. It will make analysis more complicated. Also it should be taken in consideration that distributions of almost all variables deviate significant from the Normal (Gaussian) distribution, which is evident from the values of skewness and kurtosis. For Normal distributions they are 0 and 3. This also complicates the analysis. Charts show the differences between actual and Gaussian distributions.

Next step is study the correlations between indicators in World Bank set. They are given in Table 3.

Table 1. Descriptive Statistics for World Bank data.

	N	Minimum	Maximum	Mean	Std. Dev.	Skewness	Kurtosis
LifeExpect	231	48.938	83.832	71.106	8.305	-0.696	-0.272
Diabetes	240	1.540	37.060	8.645	5.154	2.263	7.597
Tuberculosis	229	0.000	852.000	112.825	149.529	2.258	6.081
HIV	123	0.100	27.700	2.559	4.993	3.263	11.136
CO ₂ emiss	232	0.021	44.019	4.883	6.052	2.803	11.513
Methane	168	0.000	7515150	304949	913495	5.194	31.704
PM ₁₀ pollut	218	4.363	70.133	19.042	11.267	1.426	2.179
RenewEnerg	217	0.009	96.646	31.609	29.104	0.817	-0.666
UrbanPop	245	8.550	100.000	58.439	23.405	-0.062	-1.002
Hexpend	221	1.273	19.720	6.850	2.869	1.064	2.229
Immuniz	224	24.000	99.000	88.342	12.612	-2.179	5.731

Table 2. Descriptive Statistics for UNEP data.

	N	Minimum	Maximum	Mean	Std. Dev.	Skewness	Kurtosis
LifeExpect	195	61.581	88.290	77.881	6.728	-0.367	-0.739
BirthRate	195	6.662	36.255	14.725	6.276	1.076	0.214
DeathRate	195	3.105	17.727	9.208	3.231	0.548	-0.460
MaterMort	179	0.000	2000.000	310.994	426.509	1.814	3.031
MortUFive	195	0.867	78.130	15.509	16.259	1.571	2.107
InfMort	195	0.685	47.412	11.688	11.271	1.259	0.692
MortResp	21	338.000	101124	9615	22009	3.983	16.744
MalariaCase	77	3.000	12343400	672184	2042719	4.394	20.938
CH4	42	0.77	552010	44249	112476	4.066	16.294
CO	40	0.01	61343	3808	11258	4.189	18.688
CO2	42	83.22	5375000	329184	867595	5.140	29.281
SO2	212	0.00	17241.00	236.80	1263.00	11.998	158.267
N2O	42	2.93	395794.00	24278.16	62648.35	5.405	31.831
NOx	39	0.33	11882.00	853.96	2071.39	4.502	22.390
PM10	178	6.48	174.67	44.84	32.08	1.615	2.728
Ozone	185	0.00	4770.00	167.40	456.00	6.801	59.581
EmisSO2	211	0.00	41565.80	554.38	3019.90	12.259	164.185
EmisNOx	213	0.00	21684.00	499.61	1904.40	8.628	85.256
EmisHFC	133	0.00	302.82	5.67	30.06	8.468	77.869

Table 3. Correlations table for World Bank indicators.

	LifeExpect	Diabetes	Tuberculosis	HIV	CO2emiss	Methane	PMpollut	RenewEnerg	UrbanPop	Hexpend	Immuniz
LifeExpect	1.0000	0.1908	-0.6959	-0.4829	0.4632	-0.0089	-0.1439	-0.6693	0.6350	0.3356	0.5627
Diabetes	0.1908	1.0000	-0.1394	-0.1489	0.2440	0.0234	0.0303	-0.3303	0.1014	0.0295	0.0398
Tuberculosis	-0.6959	-0.1394	1.0000	0.7088	-0.3364	0.0383	-0.0126	0.4151	-0.4025	-0.0909	-0.3592
HIV	-0.4829	-0.1489	0.7088	1.0000	-0.0937	-0.0395	-0.2334	0.1064	-0.2258	0.1513	-0.0247
CO2emiss	0.4632	0.2440	-0.3364	-0.0937	1.0000	-0.0117	0.0710	-0.5440	0.4455	0.0571	0.3069
Methane	-0.0089	0.0234	0.0383	-0.0395	-0.0117	1.0000	0.3136	-0.0692	-0.0747	0.0849	-0.0452
PMpollut	-0.1439	0.0303	-0.0126	-0.2334	0.0710	0.3136	1.0000	0.0710	-0.0579	-0.3564	-0.0303
RenewEnerg	-0.6693	-0.3303	0.4151	0.1064	-0.5440	-0.0692	0.0710	1.0000	-0.5349	-0.1918	-0.4357
UrbanPop	0.6350	0.1014	-0.4025	-0.2258	0.4455	-0.0747	-0.0579	-0.5349	1.0000	0.2607	0.3242
Hexpend	0.3356	0.0295	-0.0909	0.1513	0.0571	0.0849	-0.3564	-0.1918	0.2607	1.0000	0.2132
Immuniz	0.5627	0.0398	-0.3592	-0.0247	0.3069	-0.0452	-0.0303	-0.4357	0.3242	0.2132	1.0000

It is evident that there are not significant correlations ($r > 0.7$) between health and environmental indicators. So we cannot fit linear regression models of health as dependent variables against independent environmental variables. Only correlation between Tuberculosis and HIV is slightly positive ($r = 0.7088$), which is difficult to explain.

Factor analysis gives results that three health indicators – LifeExpect, Diabetes and Tuberculosis

from all four explain 93.82% of the total variance. From environmental indicators we select CO2emiss, Methane and PMpollut. So we shall search dependencies between those variables.

Quite different is the story with UNEP indicators. In Tables 5 and 6 are presented separately correlations of health and environmental indices

Table 4. Correlations for World bank three health and three environmental indicators.

	LifeExpect	Diabetes	Tuberculosis	CO2emiss	Methane	PMpollut
LifeExpect	1.0000	0.1406	-0.7272	0.4179	-0.0172	-0.1715
Diabetes	0.1406	1.0000	-0.1997	0.4108	0.0257	0.3569
Tuberculosis	-0.7272	-0.1997	1.0000	-0.3394	0.0379	0.0196
CO2emiss	0.4179	0.4108	-0.3394	1.0000	-0.0158	0.0987
Methane	-0.0172	0.0257	0.0379	-0.0158	1.0000	0.3186
PMpollut	-0.1715	0.3569	0.0196	0.0987	0.3186	1.0000

Table 5. Correlation table for UNEP health indicators.

	LifeExpect	BirthRate	DeathRate	MaterMort	MortUFive	InfMort
LifeExpect	1.0000	-0.7611	0.3207	-0.6767	-0.8733	-0.8830
BirthRate	-0.7611	1.0000	-0.6637	0.7403	0.7954	0.7912
DeathRate	0.3207	-0.6637	1.0000	-0.4311	-0.3968	-0.4150
MaterMort	-0.6767	0.7403	-0.4311	1.0000	0.7508	0.7509
MortUFive	-0.8733	0.7954	-0.3968	0.7508	1.0000	0.9890
InfMort	-0.8830	0.7912	-0.4150	0.7509	0.9890	1.0000

Table 6. Correlation table for UNEP environmental indicators.

	CH4	CO	CO2	SO2	N2O	NOx	PM10	Ozone	EmisSO2	EmisNOx	EmisHFC
CH4	1.0000	0.9217	0.8683	0.7401	0.8779	0.9445	-0.0939	0.7887	0.9553	0.8772	0.7699
CO	0.9217	1.0000	0.9053	0.8200	0.9126	0.9663	-0.1013	0.8537	0.9584	0.9240	0.8578
CO2	0.8683	0.9053	1.0000	0.8995	0.9784	0.9732	-0.0303	0.9782	0.9290	0.9954	0.9768
SO2	0.7401	0.8200	0.8995	1.0000	0.8971	0.8800	0.1153	0.8992	0.8331	0.9052	0.8960
N2O	0.8779	0.9126	0.9784	0.8971	1.0000	0.9697	-0.0790	0.9778	0.9017	0.9874	0.9630
NOx	0.9445	0.9663	0.9732	0.8800	0.9697	1.0000	-0.0567	0.9303	0.9778	0.9809	0.9227
PM10	-0.0939	-0.1013	-0.0303	0.1153	-0.0790	-0.0567	1.0000	-0.0227	-0.0144	-0.0357	-0.0289
Ozone	0.7887	0.8537	0.9782	0.8992	0.9778	0.9303	-0.0227	1.0000	0.8491	0.9825	0.9906
EmisSO2	0.9553	0.9584	0.9290	0.8331	0.9017	0.9778	-0.0144	0.8491	1.0000	0.9320	0.8491
EmisNOx	0.8772	0.9240	0.9954	0.9052	0.9874	0.9809	-0.0357	0.9825	0.9320	1.0000	0.9761
EmisHFC	0.7699	0.8578	0.9768	0.8960	0.9630	0.9227	-0.0289	0.9906	0.8491	0.9761	1.0000

In the lower triangle part are given in bold significant correlations, which indicates that we should extract a smaller set of almost independent health and environmental indicators to be related in the subsequent analyses. It is done by factor analysis of principal components.

Factor analysis shows that only three health indicators, namely: Under-Five Mortality Rate (MortUFive), Crude Death Rate (DeathRate) and

Maternal Mortality Ratio (MaterMort) are able to explain 94.56% of the total variance and are considered in the next analyses.

Analogous factor analysis of environmental indices gives that three indicators, namely: EmisSO2, PM10 and EmisHFC explain almost 97.7% of the total variance. They will be analyzed together with shown health indicators.

Table 7. Correlations for UNEP three health and three environmental indicators.

	DeathRate	MortUFive	MaterMort	PM10	EmisSO2	EmisHFC
DeathRate	1.0000	-0.3283	-0.3912	-0.3187	0.1053	0.1826
MortUFive	-0.3283	1.0000	0.7688	0.1508	-0.0612	-0.1104
MaterMort	-0.3912	0.7688	1.0000	0.0594	-0.0107	-0.0413
PM10	-0.3187	0.1508	0.0594	1.0000	0.0713	-0.0159
EmisSO2	0.1053	-0.0612	-0.0107	0.0713	1.0000	0.9191
EmisHFC	0.1826	-0.1104	-0.0413	-0.0159	0.9191	1.0000

Next section presents the results of Redundancy and Procrustean analysis.

RESULTS AND DISCUSSION

In a general case the researcher collects simultaneously data from two groups of variables, presented by the matrix of independent (predictor) variables

$$\mathbf{X} = [\mathbf{X}_{ij}]_{n,p}$$

and matrix of dependent (output) variables

$$\mathbf{Y} = [\mathbf{Y}_{ik}]_{n,q}$$

where n is number of observations or variants of experiment, p - number of independent variables or their functions, q - number of dependent variables. The problem here is to look for the relationship between dependent health and independent environmental ones. This can be done by means of Redundancy analysis. Application of Redundancy analysis is based on the results of Ivakhnenko, Sadovski and Wollenberg (Ivakhnenko A., 1982; Sadovski A., 1985; Sadovski A., 1998; Wollenberg A., 1977). We use it on both sets of data. Redundancy index is a measure of the extent to which the variables X can describe or predict variables Y . It is given by the expression

$$R^2(Y : X) = \frac{\text{tr}(R_{YX} R_{XX}^{-1} R_{XY})}{\text{tr}(R_{YY})},$$

were R being the correlation matrix of Y and X variables.

Results shows that redundancy criterion between three health and three environmental indicators from Word Bank data is $R^2 = 0.1188$. Analogous redundancy index for UNEP data is $R^2 = 0.0086$. These low values indicate that dependencies between these health and environmental indicators are not linear.

In several environmental, health and demographic studies questions arise to compare results of multivariate analyses derived from different data by the same analysis or derived from different analyses of the same data. An appropriate method for such comparative studies is the so-called "Procrustes analysis". The theory and analytical procedures of Procrustes analysis are presented in works of P. Schonemann and R. Carroll (1970) and J. Gower

(1971). With this name are indicated least-squares problems where independent and dependent variables are matrices. Sadovski apply this method in comparing demographic and environmental studies (Sadovski A, 1995). We shall use Procrustes analysis to compare results from World Bank and UNEP data.

The problem is defined mathematically as follow: Two orthogonal matrices T and S are searched for, which transform given square matrix A in a given square matrix B , that the matrix $E = T'AS - B$ to be minimal in least-squares sense. In our study matrix A is correlation matrix from World Bank data (Table 4) and matrix B is correlation matrix from UNEP data (Table 7). Two distance matrices are derived from these data and calculated by the formula

$$d_{ij} = \sqrt{1 - r_{ij}}.$$

where r_{ij} are correlation coefficients. The distance matrix of World Bank data and analogous of UNEP data are given in Tables 8 and 9.

Table 8. Distance matrix **Ad** from World Bank data.

	LifeExpect	Diabetes	Tuberculosis	CO2emiss	Methane	Pmpollut
LifeExpect	0.0000	0.9270	1.3142	0.7630	1.0086	1.0824
Diabetes	0.9270	0.0000	1.0953	0.7676	0.9871	0.8019
Tuberculosis	1.3142	1.0953	0.0000	1.1573	0.9809	0.9902
CO2emiss	0.7630	0.7676	1.1573	0.0000	1.0079	0.9494
Methane	1.0086	0.9871	0.9809	1.0079	0.0000	0.8255
Pmpollut	1.0824	0.8019	0.9902	0.9494	0.8255	0.0000

Table 9. Distance matrix **Bd** from UNEP data.

	DeathRate	MortUFive	MaterMort	PM10	EmisSO2	EmisHFC
DeathRate	0.0000	1.1525	1.1795	1.1483	0.9459	0.9041
MortUFive	1.1525	0.0000	0.4808	0.9215	1.0301	1.0538
MaterMort	1.1795	0.4808	0.0000	0.9698	1.0053	1.0204
PM10	1.1483	0.9215	0.9698	0.0000	0.9637	1.0079
EmisSO2	0.9459	1.0301	1.0053	0.9637	0.0000	0.2844
EmisHFC	0.9041	1.0538	1.0204	1.0079	0.2844	0.0000

Here are results from Procrustes analysis. Numerical values of the criterion for accuracy of solution's approximation **Bd** is

$$\varepsilon = 0.0099,$$

which testify for a very good concordance of the results. The interpretation follows from the values of diagonal elements (trace) of the residual matrix E .

It is evident that the largest are deviations for DeathRate and EmissSO2 (-0.1599 and 0.1090 respectively). Also shows weak relationship between MaterMort and MortUFive. From here a conclusion could be drawn that these indicators cannot explained sufficiently well connection between health and environmental characteristics.

Table 10. Residual Matrix **E** (deviations from original data)

	DeathRate	MortUFive	MaterMort	PM10	EmisSO2	EmisHFC
DeathRate	-0.1599	0.0059	0.0197	-0.0198	-0.0054	-0.0355
MortUFive	0.0059	0.0228	-0.2191	-0.0473	0.0235	0.0313
MaterMort	0.0197	-0.2191	0.0133	-0.0250	0.0100	0.0106
PM10	-0.0198	-0.0473	-0.0250	-0.0917	-0.0163	0.0053
EmisSO2	-0.0054	0.0235	0.0100	-0.0163	0.1090	-0.3073
EmisHFC	-0.0355	0.0313	0.0106	0.0053	-0.3073	0.1065

CONCLUSIONS

We suggest a methodological approach to study the relationships between indicators of health and environment by using correlation, Redundancy and Procrustean analysis, and present preliminary results from these analysis on two sets of data published by the World Bank and UNEP. These methods may be used successfully for evaluation of the links between environmental quality and health indices. Correlations between health and environment are not sufficient to explain these relationships, as they represent only linear dependencies. Apparently we cannot fit linear regression models of health as dependent variables against independent environmental variables. Relationships are more complicated and probably non-linear. This was confirmed by the conducted Redundant analysis. Also there are internal links and dependencies within the groups of health and environmental indicators. The Procrustean analysis is able to reveal strong and weak links between such indicators.

The main conclusion of this study is the need for numerous reliable publicly available data to obtain credible conclusions about the impact of the environment on health. Otherwise, some environmental groups can manifest speculative theories and allegations.

REFERENCES

- Canadian Public Health Association, Global Change and Public Health: Addressing the Ecological Determinants of Health. Discussion Document, 2015.
- Lee D.H.G., Environmental Health and Human Ecology. Am. J. Pub. H., Vol. 54, 1964, No. 1, 7-10.
- Gower J., Statistical methods of comparing different multivariate analyses of the same data, Mathematics in the Archeological and Historical Sciences, (eds. F. Hodson et al.), Edinburgh, Univ. Press, 1971, 138-149.
- Ivakhnenko, A., Inductive method of self-organizing models of complex systems, Naukova Dumka, 1982, Kiev.
- Kovats, S (eds.), Health Climate Change Impacts Summary Report Card, Living With Environmental Change, 2015.
- Sadovski A., Self-organizing Models of Biological and Environmental Systems. J. of Balkan Ecology, vol. 1, 1998, No. 2, 15-20.
- Sadovski A., Method of comparing Environmental and Demographic studies. Automation and Informatics, Mini symposium EEEP '95, 7 - 9 November, 1995, Sofia.
- Sadovski A., Redundancy Analysis of Yield and Biometric Indices in fertilizing Wheat with Macro elements. Plant Physiology, Vol. 7, 1985, 215-217, Sofia (in Bulg.)
- Schonemann P., R. Carroll., Fitting one matrix to another under choice of a central dilation and a rigid motion. Psychometrika, Vol. 35, 1970, 245-255.
- UNEP, The UNEP Environmental Data Explorer. United Nations Environment Programme. 2016, <http://ede.grid.unep.ch>.
- World Bank. <http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators#>
- WHO, The Ottawa Charter for Health Promotion. 1986, <http://www.who.int/healthpromotion/conferences/previous/ottawa/en/#>
- WHO, WHO Public Health & Environment. Global Strategy Overview, 2011, 1-11.
- Wollenberg A. van den, Redundancy Analysis an alternative for Canonical Correlation Analysis. Psychometrika, Vol. 42, 1977, 207-219.

ВЪРХУ НЯКОИ ЗАВИСИМОСТИ МЕЖДУ ЕКОЛОГИЯТА И ЗДРАВЕТО

Александър Садовски

Резюме: Три математико-статистически методи, а именно – корелационен, редувантен и прокрустов анализ са приложени за анализиране зависимости между екологични фактори на околната среда и индикатори за здравето. Използвани са данни от Международната организация за опазване на околната среда (UNEP) и Световната банка (World Bank). Екологичните фактори включват емисии на различни вредни вещества – N_2O , NO_x , PM_{10} , CO_2 , както и икономически фактори като възобновяеми енергийни ресурси. От индикаторите за здраве се разглеждат продължителност на живота, смъртност, заболяемост от диабет и туберкулоза. Резултатите дават основа за редица дискусии и заключения.

Ключови думи: екологични фактори, здравни индикатори, корелация, редувантен анализ, Прокрустов анализ

Проф. Д-р Александър Н. Садовски

Академик на МАНЕ

Тел. 02-981-5586

e-mail: alexsad@hotmail.com

Prof. Dr. Alexander N. Sadovski

Academician of the IEAS

Phone: +3592-981-5586

e-mail: alexsad@hotmail.com