27 HOW TO MERGE ENVIRONMENTAL INFORMATION USING META-ANALYSIS

27.1 Introduction

Despite technical progress, e.g. automation of measurements, remote observations etc. environmental studies are still exceptionally expensive, time consuming and labor demanding. These studies are often performed by different research teams, in different place and at different times using different measurements methods and standards. Gradually it turned out that this type of situation, even in case of numerous and extensive research causes many difficulties and uncertainties. Standardization of the environmental investigations became an urgent need. As a result many quality assurance (QA) and quality control (QC) guidelines have been proposed to meet basic scientific standards. A variety of such sophisticated guidelines were developed by United States Environmental Protection Agency to properly deal with environmental data. They form so called life-cycle of data in the EPA Quality System. The main purpose of such approaches is to increase as much as possible probability- based evaluations of environmental experiments instead of judgmental ones. The later approach although cheep and fast depend on an expert knowledge and its experience, which is not reasonable measureable. Besides, it is not possible to estimate the precision of the environmental assessments without statistical methods. However, the situations, when studies meet all QA and QC requirements, are relatively rare. Finally, it is often necessary to use different independent studies whose results should be correctly combined and generalized in order to take well-founded and more precise conclusions than using single study, but it is not easy task. The article explains how to integrate information originated from different studies in such practical situation. A family of statistical methods that are developed especially for combining of various studies leading the same objectives are referred to as meta-analysis. Meta-analysis is a large family of sophisticated methods that go far beyond the well-known classical methods based on classical correlation or regression analysis [1]. It is worth to promote these methods for environmental applications due to their simplicity and high efficiency. Due to small length of the article it was not possible to present in it detailed example of using meta-analysis for environmental applications. This will be done during conference presentation. Instead, the article contains carefully selected literature, related to environmental and related studies [2-5, 8-10, 15, 19].

27.2 Requirements for environmental studies and their results from meta-analysis perspective

In practice, quality of the experimental data arising from different measurements is also very different. Usually much more realistic requirements for results of environmental studies are possible than those described by rigorous QA and QC procedures. The data that meet some minimum requirements for the meta-analysis purposes are referred to as homogenous and exchangeable ones.

This implies that:

- Experimental procedures satisfy the sufficient quality standards on the type of sampling, measurement techniques as well as on laboratory analysis.
- The investigation planed for meta-analysis must have the same goal.
- Research should be performed under approximately the same conditions.
- The underlying effects of measured variables should be non-stochastic and homogenous among surveys. Such effects are referred to as fixed.
- All relevant results are accessible for statistical analysis in appropriate publications or databases. (This does not mean that data must be necessarily free, although this situation is highly desirable). The problem of unavailability of existing data is sometimes called, in statistical jargon, as the file drawer problem.

Sometimes there can be sometimes even worse, if some of basic assumptions 1-5 are still not fulfilled. But even in these cases it is sometime possible to properly perform meta-analysis and integrate the data taking whole benefit from the existence of numerous sets of data. For example, when the condition 3rd is not fulfilled this drawback can be compensated through appropriate selection of weights for the study results, depending on their quality and size. Lack of fulfilling 4th condition can be compensated theoretically using duly selected random-effect model, although it is not trivial task. It is necessary to remember that even the most advanced and sophisticated analysis can not either replace the lack of data or compensate experimental errors.

27.3 Search for data and main causes of publication bias

Searching for relevant data is first and essential step in any meta-analysis. Research may be local, regional, national and even global. This involves the increasing difficulty in finding relevant data. Therefore, below will be described only the problems arising when solving general questions. Nowadays, the main sources of information are electronic databases, with remote access. Even if such data are provided by prestigious, international institutions one should be aware that they are subject to various restrictions and often are not fully representative. Very important are the constraints imposed by language. Currently, most data sources come from English language website. This results in a certain tendency (bias) to narrow the information and change its interpretation, due to both geographical distribution and some organizational factors, as well as cultural ones. This kind of tendency in information has similar reasons to the above-mentioned. Namely, some scientific groups or organizations are cited more frequently than another, e.g. because of their importance, influence on the scientific community. Their publications are more often produced and cited and as a result be taken more often into account during merger of information. There are also another numerous sources of publication biases. For example, more willingly are published these results, that indicate the existence of certain relationships, than the results showing their absence.

27.4 Combining studies by means of P-values

A P-value reflects the strength of evidence supporting a null hypothesis. If the test statistic in a hypothesis test is equal to given value, then the P-value is the probability of observing a test statistic equal to this given value, under assumption that the null hypothesis, H0, is true. If the P-value is less than the significance level, one rejects the null hypothesis, H0.

The importance of methods based on combining studies by means of P-values consists in widespread recognizing them, by scientific community as a solid, working standard. Of course, the initial stages of the meta-analysis, particularly concerning the accuracy and significance of individual studies should be carried out correctly on carefully selected random samples. It concerns to research on the same subject, made by the same or different research teams, using similar methods.

Three main cases where these methods are used are described below.

a) In the ideal case, there are available all the sets of raw data e.g. results of environmental measurements such as water pollution, soil pollution, effects specific to the health of human populations. With all the raw data sets it is possible to perform the statistical analysis of the same null hypothesis H0, preserving as much as possible the same rules associated with different stages of the analysis e.g. for rejecting outliers, clustering of data etc. The analysis should be carried out to verify the same null hypothesis H0 in each case and calculating appropriate P-values. The set of all P-values will be then for in the meta-analysis. Such a way of making meta-analysis is recommended and has the highest quality.

b) Most often intermediate case occurs, in which can be accessed only to selected sets of raw data. Other publishable results that can be used in the meta-analysis should both apply to the same, precise null hypothesis, H0, and be developed by the same statistical methods in which P-values are determined. The quality of the meta-analysis performed in this case, using methods based on combining of P-values, depends mainly on the proper selection of relevant contributions and proper evaluation of the comparability of their results. This subject will be continued below.

c) The third case concerns the relevant research that was conducted by generally recognized scientific teams, leading to verification of the same null hypothesis H0, and to the set of P-values, when these works do not contain direct information about raw data. Now, only possibility to perform the meta-analysis is to use the published P-values.

27.4.1 The Tippet's method

The oldest and the simplest method of combining results from K different experiments using P-values, was proposed by Tippett [18]. This method is based on arranging in ascending order P-values obtained from different experiments, which tested the same null hypothesis H0. Then, the smallest P-value, Pmin , is chosen in order to verify the null hypothesis, H0, for a set of experiments. If this value was smaller than simple function of the proposed level of significance, α, i.e. if Pmin < 1-(1- α)1/K the null hypothesis, H0, is rejected in favour of the alternative hypothesis H1. Despite its simplicity, the Tippet's result of the method is dependent one and only on the study, which shows the strongest dependencies in whole series of the individual studies (i.e. which has the smallest level of significance, α). This lead too often to rejection of the null hypothesis, H0, the more that this fault of the Tippet's method strengthens the common publication bias associated with the greater effort of researchers to detect dependencies, than deny them.

27.4.2 The Wilkinson's method

This method was improved by Wilkinson [20], who generalized it in such a way that it could be used for every Lth P-value, PL. The null hypothesis H0 about the lack of dependence was rejected when the L ≤ Kα,K,L where K is the value read from the appropriate statistical table. Although this method is more resistant to accidental errors, its use is still problematic, because it requires a suitable choice of PL, and consequently did not necessarily lead to clear results, especially when several PL's were selected.

27.4.3 The Fisher's method

Probably, the most common method of results merger in the meta-analysis is one proposed by Fisher [7], known also as inverse chi-square method. This method combines simplicity with relatively high effectiveness. There are also many variations of this method, which due to the small length of the article could not be described here. The interested reader can find their detailed description specified in the bibliography, e.g. [11-13].

The input in the Fisher's method is the set of P-values, whose cardinality K, is equal to the number of independent studies. Denoting the P-values, by Pk, where k = 1, ... K, it follows from P-value definition that 0 <Pk <1 for each k. In the Fisher's method Pk's are treated as independent random variables with uniform distribution in the interval (0,1). It can be also proved that independent random variables of the form -2log(Pk) have the chi-square distribution with degree freedom equal to 2. The sum of such K independent random variables has also the chi-square distribution with degree freedom equal to 2K. This can be expressed mathematically by the following equation:

$$X_{exp}^2 = -2\sum_{k=1}^{K} \log\left(P_k\right) \tag{27.1}$$

This statistic is indicated often in short, by X2(2K), where 2K is the degree freedom, which is equal to doubled number of independent studies, that are the subject of the meta-analysis. When the aggregated value of the experimental statistics is determined, it is then compared to the chosen significance level, α. The later is the probability that the test statistic will reject the null hypothesis H0, when it is true. This can be expressed by the following expression:

$$P\left(X^2(2K) \geq X_{exp}^2\right) = \alpha \tag{27.2}$$

If $P\left(X^2(2K) \geq X_{exp}^2\right) \geq \alpha$ is true, then there is no reason to reject the verified null hypothesis, H0, at the level of significance α (i.e. about lack of any correlation or dependency between the studied variables, evaluated on the basis of the combined K experiments).

If $P\left(X^2(2K) \geq X_{exp}^2\right) < \alpha$ occurs, then the verified null hypothesis H0 is rejected and the alternative hypothesis H1 is accepted at the level of significance α. (For example, it means that combined K studies confirmed occurrence the correlation or dependency between studied variables).

27.5 Weighting individual studies using effect size

27.5.1 Effect size

Methods based on integration of data using P-values assume the equal importance of the individual studies. In other words, it is assumed that equal weights are attributed to these studies. Even having similar quality of the studies, equal size of samples, etc. methods for combining information using P-values are not sufficient to take precisely into account size of a phenomenon associated with a specific P-value, e.g. the strength of correlation between the measured environmental variables, the intensity of studied phenomenon (such as pollution levels) etc.

The basic method of determining the weights, wk, of the individual studies is to determine the reciprocal variance of the so-called a size effect, dk:

$$w_k = \frac{1}{Var[d_k]}\,. \tag{27.3}$$

Expression for dk is given below, using Eq (27.4).

The size effect is a precise, mathematical measure of importance of a k-th significant P-value [16].

In the environmental studies of the effect size is determined usually from a series of k, comparative studies in which measurements (a target group, indicated below with an index T) are compared each time with the control measurements (a control group, indicated below with an index C). A detailed description of determining the effect size by calculating a standardized mean difference between the target group and the control group was firstly developed by Cohen [6]. He suggested that departures from the null hypothesis, H0, about no differences between the average of the target group, $\bar{X}_{Tk}$, and the control group, $\bar{X}_{Ck}$, can be expressed using the same metrics - a standardized mean difference given (in the simplest form) using the following expression:

$$d_k = \frac{\varphi_k(\bar{X}_{Tk}-\bar{X}_{Ck})}{s_k}, \tag{27.4}$$

where: s_k, referred as to for standard deviation of combined k-th study (i.e. calculated for both the target and the control group), and φ_k is correction factor for small samples studies. It can be proved (Cohen, 1969) that they can be determined from the following expressions:

$$s_k^2 = \frac{(N_{Tk}-1)s_{Tk}^2+(N_{Ck}-1)s_{Ck}^2}{N_{Tk}+N_{Ck}-2}, \tag{27.5}$$

And

$$\varphi_k = 1 - \frac{3}{4(N_{Tk}+N_{Ck}-2)-1}. \tag{27.6}$$

In the above formulas N_{Tk}, N_{Ck}, indicate number of measurements in the target and control group, respectively. Similarly, s_{Tk}, s_{Tk} are standard deviations in the target and control group, respectively.

Expressions for estimators of the effect size variance, $Var[d_k]$ are in general very complicated, because they depend on many statistical parameters of the target and control samples, such as their size, means, standard deviations, correction factor, etc. They can be found in the appropriate, specialized literature devoted to the meta-analysis, e.g., Hedges and Olkin (1985).)

A graph showing dependence of $1/Var[d_k]$ as a function of d_k is very useful tool for visualisation a publication bias mentioned in section 3. Sometimes it is referred to as a funnel plot [21]. In general a funnel plot is the graph the sample size against its corresponding effect size. Exemplary funnel plot is shown in fig. 27.1.

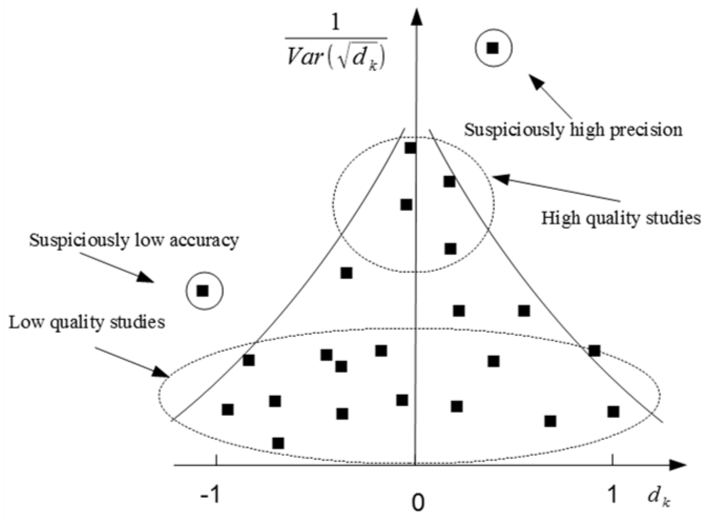

Fig. 27.1 Exemplary funnel plot for multiple independent studies of varying quality

This graph allows for quick comparison of the quality of many independent studies, as well as allows for extracting studies showing poor quality, or having a suspiciously high precision.

27.5.2 The Liptak-Stouffer's method

The simplest and most widely used method of integration of information which takes into account both the P-values and the size effect is the Liptak-Stouffer's one [14]. It is analogous to the Fisher's method, but instead of the logarithms of the P-values, -2log(Pk), it uses the weighted inverse values of the standard normal cumulative distribution function calculated for the P-values, namely $\Phi^{-1}(P_k)$. In the Liptak-Stouffer's method one applies similar assumptions and practices as those described above for the Fisher's method.

The basic statistic used to verify the joint null hypothesis based on multiple independent studies (analogous to the statistics given by Eq. (27.1) is now given by the expression:

$$Z_{exp} = \frac{\sum_{k=1}^{K} w_k \Phi^{-1}(P_k)}{\sqrt{\sum_{k=1}^{K} w_k^2}} \tag{27.7}$$

Additionally, it is necessary to calculate (or read from statistical table) the value of the standard normal cumulative distribution function for the Z_{exp}, i.e. $\Phi(Z_{exp})$.

If $\Phi(Z_{exp}) \geq \alpha$ is true, then there is no reason to reject the common null hypothesis, H0, evaluated in the K different experiments, at the given level of significance α.

Otherwise, if $\Phi(Z_{exp}) < \alpha$ then the common null hypothesis, H0, is rejected, and the common alternative hypothesis, H1, is accepted, at the given level of significance α.

If all weights are equal one (wi=1, for k=1...K) the Liptak-Stouffer's method simplifies to the Stouffer's method [17] based merely on the P-values, which is equivalent to the Fisher's method.

27.6 Examplary application of meta-analysis in soil pollution studies

Example given below, is based on the soil pollution measurements performed in frame of the grant of Ministry of Science and Higher Education of Poland, headed by the author, and entitled "Statistical and geostatistical analysis of the possibility of magnetometric method use for pre-screening of soil pollution by industrial and urban dusts". All measurements were made in Upper Silesian Industrial Area which is one of the most urbanized regions of Poland [23,24].

Soil magnetic susceptibility is important of soil property which enables to assess the level of soil pollution with heavy metals. Dust produced during many industrial processes or fuel combustion contains significant amounts of heavy metals. Simultaneously, dust deposition on the soil surface causes the increase of the soil magnetic susceptibility. For this reason, field magnetometry may be used as a low-cost substitute for expensive and time-consuming chemical analyses [22]. Usually two above-mentioned types of measurement are carried out simultaneously. Measurements of surface magnetic susceptibility can be conducted directly in the field using e.g. the MS2D "Bartington" susceptibility measuring device, integrated with the GPS system. The depth of penetration of this sensor equals about 10 cm, and for this reason these measurements are relatively imprecise. Then the surface magnetic susceptibility measurements are commonly validated by much more accurate chemical measurements, or magnetic susceptibility measurements performed in soil profile. However, the latter measurements are more costly and thus also limited, especially when a large area is studied and many geochemical analyses are needed to perform. To integrate these measurements the cokriging method can be used. To use this method effectively it is necessary to estimate first the Pearson's correlation coefficient between relatively inaccurate surface magnetic susceptibility and precise chemical measurements [23]. To do so, several study sites were selected, each of which covered an area of about a few square kilometers. Then the correlation coefficients between surface magnetic susceptibility and chemical measurements were calculated and analysed in terms of the feasibility of the cokriging method. However, because of complex environmental factors very different values of the Pearson's correlation coefficient with different significance levels may be obtained in each study site. To combine properly the whole information about these different correlation coefficients the meta-analysis methods can be used. In the discussed example, five different study sites (denoted, hereafter, as to for A, B, C, D, and E) were selected in which above-described magnetometric and geochemical measurements (namely, Fe, Mn, Zn, Pb, Cd, Cu, Cr, Ni, Co concentrations in the soil) were performed. In order to reduce the significant anthropogenic influx, these study areas were located preferentially in forests. Below, it will be demonstrated how it is possible to aggregated the correlation coefficient's obtained from these different study sites between surface magnetic susceptibility and Pb measurements concentrations, using as an example, the Fisher's and Stouffer's methods. The number of the pairs of chemical and magnetometric measurements in the same places, in sites A, B, C, D, and E, were 17, 15, 15, 17, and 20, respectively.

The following correlation coefficients were obtained for measurements performed in these regions: 0.7986, 0.5593, 0.4565, 0.4231, and 0.8758 with the significance levels equal to 0.001, 0.03, 0.087, 0.091, and 0.001, respectively. It can be seen from these data, that both correlation coefficients and significance levels differ very much and that twice significance level exceed 5% which was a pre-specified value. This should raise considerable concern whether the observed correlations truly exist, and whether the cokriging method can be used. Using the Fisher's method $\chi 2calc=19.2486$ is obtained, and the combined P-value, i.e. $P[\chi 2(10) > 19.2486] \leq 0.0373$ is thus smaller than 5%. Therefore one can conclude that a positive correlation, aggregated over all 5 sites, exists between surface magnetic susceptibility measurements and Pb concentrations. When using the inverse normal method (the Stouffer's one) one get similar results. Thus, $Zcalc= -5.48356$, and $P+=\Phi(Zcalc) < 0.001$ which is significantly smaller than 5%, which leads to the same conclusion, as in the case of the Fisher's method. Similar calculations should be performed for the rest of investigated soil pollutants.

Conclusions

Meta-analysis is a family of sophisticated methods that go far beyond the well-known classical methods of data merger based on classical correlation or regression analysis. It is worth to use these methods in environmental applications due to their simplicity, high efficiency, and importance. However, these methods are not trivial, and they can be they can be part of more complex statistical calculations. Therefore is necessary to have considerable experience in applying these methods, together with knowledge of the details of studied problem. This can be easily achieved through the cooperation of a specialist in a given field with an environmental statistician.

REFERENCES

[1] Blair A., Burg J., Foran J., Gibb H., Greenland S. Morris R., Raabe G., Savitz D., Teta J., Wartenberg D.: Guidelines for application of meta-analysis in environmental epidemiology. ISLI Risk Science Institute. Regul. Toxicol. Pharmacol 1995, 22(2):189-97.

[2] Boffetta P., Tredaniel J., Greco A.: Risk of Childhood Cancer and Adult Lung Cancer after Childhood Exposure to Passive Smoke: A Meta-Analysis. Env. Health. Persp. 2000, 108(1): 73-82.

[3] Boon D.Y., Soltanpour P.N.: Lead, Cadmium, and Zinc Contamination of Aspen Garden Soils and Vegetation. J.Environmental Quality 1992, 21: 82-86.

[4] Boyle K.J., Poe G.L., Bergstrom, J.C.: What do we know about groundwater values? Preliminary implications from a meta-analysis of contingent valuation studies. American Journal of Agricultural Economics 1994, 76(5): 1055-1061.

[5] Cobb G.P., Sands K., Waters M., Wixson B.G., Dorward-King E.: Accumulation of Heavy Metals by Vegetables Grown in Mine Wastes. Environmental Toxicology and Chemistry 2000, 19: 600-607.

[6] Cohen, J.: Statistical Power Analysis for the Behavioural Sciences. New York. Springer-Verlag 1969.

[7] Fisher, R.A. 1948. Combining independent tests of significance. American Statistician 2, 30.

[8] Gurevitch J. and Hedges L.V.: Statistical issues in ecological meta-analyses. Ecology 1999, 80: 1142-1149.

[9] Hasselblad V.: Meta-analysis of environmental health data. Sci. Total Environ 1995. 160-161:545-58.

[10] Hedges, L.V.: Modeling publication selection effects in meta-analysis. Statistical Science 1992, 7: 246-255.

[11] Hedges L.V., Olkin I.: Statistical Methods for Meta-analysis. Orlando, F.L. Academic Press 1985.

[12] Hunt, M.: How Science Takes Stock. The Story of Meta-Analysis. New York. Russell Sage Foundation 1997.

[13] Hunter J.E. and Schmidt F.L.: Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. Newbury Park. Sage Publications 1990.

[14] Liptak T.: On the combination of independent tests. Magyar Tudomanyos Akademia Matematikai Kutato Intezentek Kozlemenyi 1958, 3: 171-197.

[15] Rosenberger R.S. and Loomis, J.B.: Using meta-analysis for benefit transfer: insample convergent validity tests of an outdoor recreation database. Water Resources Research 2000, 36(4): 1097-1107.

[16] Rosenthal R.: Parametric measures of effect size. In H. Cooper & L.V. Hedges (Eds.), The Handbook of Research Synthesis (pp. 231-244). New York: Russel Sage Foundation 1994.

[17] Stouffer S.A. Suchman E.A., De Vinney L.C., Star S.A., and Willams, R.M, Jr.: The American Soldier, Volume I. Adjustment during Army Life. Pricenton, NJ: Pricenton University Press 1949.

[18] Tippett, L.H.C. 1931. The Methods of Statistics. London: Williams&Norgate.

[19] van den Bergh, J.C.J.M., Button, K.J., Nijkamp, P., & Pepping, G.C.: Meta-analysis in Environmental Economics. Boston: Kluwer Academic Publishers 1997.

[20] Wilkinson B. 1951. A statistical consideration in psychological research. Psychological Bulletin 48: 156-158.

[21] Light R.J., Pillemer D.B.: Summing Up: The science of Reviewing Research. Cambridge, MA. Harvard University Press 1984.

[22] Strzyszcz Z. and Magiera T.: Heavy metal contamination and magnetic susceptibility in soils of southern Poland. Phys. Chem. Earth 1998, 23, 1127–1131.

[23] Zawadzki J., Fabijańczyk .P, Magiera T., Strzyszcz Z.: Study of litter influence on magnetic susceptibility measurements of urban forest topsoils using the MS2D sensor. Environmental Earth Sciences 2010, 61(2): 223-230.

[24] Zawadzki J., Magiera T., Fabijańczyk P., Kusza G.: Geostatistical 3-dimensional integration of measurements of soil magnetic susceptibility. Environmental Monitoring and Assessment 2012, 184(5):3267-78.