Solving Waste Management Data Uncertainties
Case Study of South Moravian Region

Michal Hejč, Jiří Hruběčk

Abstract

Imprecision of input data is an important characteristic feature of environmental monitoring. When making evaluations, conclusions and the decisions from collected data, one has to be very careful not to make fatal mistake. The main task of information and communication technology (ICT) is to deal with the primary data uncertainty and thus to reduce the risk of mistake. At present new approaches and methodologies to handle data uncertainty are explored as opposite to standard. This paper presents comparison of such new approach against standard one in the case study of the assessment of waste management indicators in the South Moravian Region of the Czech Republic. The new approach brings more modular way of dealing with uncertainty and is supposed to bring better results and less workload than standard approach.

1. Introduction

Waste management consumes enormous amount of financial resources and requires sufficient knowledge of decision makers at all levels of management – that means the need for quality data. This need is even increased by the possibility of various environmental risks in case of wrong decision. The data quality is, according to the standard ISO 8402, a characteristic of an entity as a whole that give the capability to satisfy explicit and implicit needs. Many articles have been published (Lacuška/Janář/Hruběčk 2004), (Hruběčk et al 2003), (Kral/Zemlicka 2005) about the environmental data and their weaknesses – uncertainties. So the lack of quality of environmental data is well known and we have to face classic problem – to get reasonable conclusions from imprecise input. It’s possible, but it depends on invested effort and the methods used.

We will discuss this problem with respect to the regional waste management indicators assessment (Hruběčk/Hejč 2005) in the South Moravian Region of the Czech Republic (CZ). These indicators are based on core set indicators of European Environment Agency. Indicators have to be evaluated annually in 12 regions of CZ and also in the Ministry of Environment of CZ (MoE), where standard data validation methodology is used (Vejnar/Mlnáříková 2002). Their evaluation and comparison with planned ones make the basis of accepted decisions for decision makers of the regional governments and the MoE.

2. Case study

Primary data come from annual reports of all waste originators and treatment facilities. These annual reports contain the identification of waste and its amount, origin and destination. The data have to satisfy needs of indicator assessment. There are two main general sets of indicators: waste production set and waste treatment set. These two general sets consist of particular indicators (there are about 50 evaluated indicators differing mostly by type of waste or type of subject – waste originator or treatment facility).

---

1 Faculty of Informatics, Masaryk University Brno, Czech Republic, E-mail: 3148@mail.muni.cz, Internet: www.fi.muni.cz
2 Centre of Biostatistics and Analysis, Faculty of Medicine and Faculty of Science, Masaryk University Brno, Czech Republic, E-mail: hrebicek@cba.muni.cz, Internet: www.cba.muni.cz
Primary data uncertainty is caused by various factors on local and regional level, e.g. measurement errors, estimation of waste weight imprecision, errors during manual input of the data into electronic form (human factor) and even unlawful behaviour of subjects (they intentionally refer incomplete or misguided reports, which is a big problem and may be one of the key factors of uncertainty).

The paper describes the standard process of identification of the uncertainty in the primary waste management data and evaluation of their quality with respect to the Regulation 2150/2002/EC, on waste statistics. The new developed modular approach is also presented and compared with the standard one. Results of this comparison are interpreted with respect to possibilities of future development of standards for data quality verification in the CZ with respect to the Regulation 2150/2002/EC.

2.1 Standard Approach
This is the approach complying the Regulation 2150/2002/EC and used by the MoE (Vejnar/Mlinárová 2002). It uses primitive knowledge database (formed only by reports from previous years) and the interval arithmetic to identify the data uncertainties. Identified uncertainty is not directly cured, but it is consulted with the place of data origin and the data are repaired only in the case of confirmation of the error (the confirmation itself is considered as uncertain, too). This clumsy procedure is mandatory as the report and the data are parts of duties given by law.

Interval arithmetic is here used to address the amount of reported waste uncertainty. However, this method does not provide adequate information about the nature of output uncertainty as all the uncertainties are forced into one arithmetic interval (Kutscher/Schulze, 1993). The probability distributions of the inputs are not generally known (with some significant exceptions). That means probabilistic analysis cannot be used.

Validated data are used for assessment of all waste indicators. That means all of the results are less or more harmed by uncertainty. But we have consistent time series and the results of evaluation of these time series have good quality because this approach is used for several years unchanged.

2.2 Modular Approach
This approach is currently explored in Masaryk University of Brno and has been already used in the case study of waste management indicators assessment in (Hrbiček/Hejč 2005).

Main methods dealing with uncertainties are interval arithmetic, fuzzy and probabilistic approach. Each of them has its strengths and weaknesses and needs different algorithm and data representation (Hrbiček/Hejč, Holubek 2006). The assessment of indicators from the same set of data generates the need to have distinct types of evaluations (and methods) for distinct indicators and applied on different parts of data. This is satisfied by the new modular approach. Every module has its own data model with some constants and parameters (parameters will be the subject of automated optimization).

Some modules required the extension of knowledge database used in our case. Public information from eGovernment sources has been used, later we can employ some databases not available to public and other eGovernment sources. In other words we will need every more or less possible source to allow the automated process of model shaping to have sufficient choices. Knowledge database is treated as certain at this stage, later it will be treated as uncertain in the same manner as the examined primary data.

As there is no possibility to know the truth about the data, we have to use new probabilistic module - "Trustworthiness" (which estimates the data trustworthiness rating and is subject of automated optimization, too). Also estimated number of errors is counted and is the subject of optimization.

Last module ("Optimization") contains various less or more subjective optimization criteria. The strongest criterion is the equation between production and treatment indicators. Next iteration of the
automated process means the need for substitution of uncertain data or request for correction of reports (Trustworthiness module), this need is satisfied by particular waste module and its model and the comparison with optimisation criteria is conducted (Optimization module).

2.3 Results Comparison

Figure 1 shows whole workflow information and waste management processes of data collection and validation, assessment of indicators, conclusions. The main difference (and the contribution of the new method) can be seen in the estimation of uncertainty and the feedback.

![Standard approach compared to new approach to handle the uncertainty in the data.](image)

Good example of the difference between the approaches is the way how they treat municipal waste data. Standard approach uses intervals of possible production of given types of waste. Values are taken from its knowledge database. The data out of the interval are treated as uncertain. The second approach has its module – “Municipal Waste” in this case. This module contains a model of municipal waste production. Constants in the model are taken from enhanced knowledge database and parameters are the subject to optimization. This particular module requires numerous enhancements of knowledge database (number of inhabitants and more then 15 other particular enhancements). Trustworthiness module is used to identify the uncertain data. Optimization module processes the parameters of both previous modules and gives the result by given optimization criteria.

As the first result of evaluation (for the year 2004) we can say that the confirmation of possible errors in the place of data origin (which itself is treated as uncertain, too) shouldn’t be used in some cases – direct substitution is used instead (for example Municipal Waste module revealed that 12.000 tons of municipal waste in Brno was counted twice and many other cases of mistakes, which do not need confirmation – they can be substituted directly). Saved workload should be invested in more precisely conducted confirmation of other “suspects” to lower their uncertainty (for example in the cases when model lacks required constants and we need to rely only on confirmation) or it should be invested in some other
confirms 15 confirmations and no substitutions in our case. Standard approach replaces most of these confirmations by substitutions and adds the need for 70 more new confirmations.

New approach also reveals the gaps of the data, not revealed before (standard approach reveals no gaps and modular one revealed more than 100 gaps in the total of 150,000 items of database). These should be cured either by substitution or by law means – requests for completion (as the report is a part of duties given by the law). As the data are evaluated annually, the next steps in exploration are supposed to be done during the evaluation for the year 2005.

3. Conclusions

Standard approach of dealing with the data uncertainties is burdened by inherent limitations – difference in quality of various data sources results in inconsistency of uncertainty treatment and thus in the low quality of information and waste indicators generated by the data. It is not possible to easily overcome the problem without application of some new approach.

The new developed approach is currently explored and used in parts. Expected benefits comply with the Regulation 2150/2002/EC and are worth the exploration, because they offer significant improvement of the information, carried by the data. This means finance savings and less environmental risk in our case.

As the process of evaluation becomes more automated, the new approach can generate even less workload or the same workload with better results. But the exploration is far to be complete at this moment and there is still much to do – especially in the field of particular waste modules.

Bibliography


408