CHEMICAL ENGINEERING TRANSACTIONS
VOL. 61, 2017
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors:Petar SVarbanov, Rongxin Su, Hon Loong Lam, Xia Liu, Jiří J Klemeš

ISBN978-88-95608-51-8; **ISSN** 2283-9216
DOI: 10.3303/CET1761162

Verification of Information in Large Databases by Mathematical Programming in Waste Management

Radovan Šomplák[a*], Vlastimír Nevrlý[a], Veronika Smejkalová[b], Martin Pavlas[a], Jakub Kůdela[b]

[a]Institute of Process Engineering, Faculty of Mechanical Engineering, Brno University of Technology – VUT Brno, Technická 2896/2, 616 69 Brno, Czech Republic
[b]Institute of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology – VUT Brno, Technická 2896/2 616 69 Brno, Czech Republic
Radovan.Somplak@vutbr.cz

The obligation to register production and waste management leads to a formation of a large-scale database. The reporting obligation concerns immediately a large number of subject that may cause discrepancies in reported data. The paper presents an approach for error detection in large data files. Errors are reflected as inconsistency in total production and processing or in transportation between two nodes. In this case, the area (node), that has sent the waste, registers a different quantity than the node that has received the waste. The database of waste management is an essential source of information for many calculations and analysis, which further open up the scope for the realisation of projects, so it is important to have accurate data.
This paper presents an approach for identifying errors in the database using mathematical programming techniques. This issue was solved as a task of network flow with an emphasis on the force of mass balance in nodes. The objective is to make the amount of produced and delivered waste to each node equal to the amount that was there processed or removed. This is required with the minimum modification of the input data. Weights are introduced to distinguish high and low-quality data by assigning bigger values to arcs where sent amount correspond with quantity received. In this case, there is no reason to consider the data as erroneous. This tool has been tested through a case study on the database of waste management in the Czech Republic. The considered network consists of 206 nodes representing municipalities, which corresponds to 42,230 edges (possible flows). The output from the calculation is a large amount of data, which are in terms of approximation to initial values interpreted as maps.
However, the tool could be used for other areas of records and databases, where there is a transfer of any material flow. In the further research, the model can be supplemented by specific constraints arising from additional information for the specific application. In this case, decision-making about network flow would be done with taking into account the shortest distance between producers and treatment facilities.

1. Introduction

In many fields, the central databases, devoted to acquiring, producing, handing over and processing of certain material flow, contain some inconsistencies in reported data. This problem arises mainly because of the involvement of a large number of entities that have the obligation to report. As an example of fields, where inconsistencies can emerge in the registration database, can be mentioned: records of income and expenses - tax documents, receipt and delivery of goods, reporting in waste management.
The dependence of conclusive results from many optimisation-based models on accurate data inputs is indisputable. An inaccuracy in input data can lead to erroneous or suboptimal decisions, especially in network flow models such as Šomplák et al. (2013) or in supply chain models like Čuček et al. (2011) and Stille et al. (2011) for P-graphs, or other problems considered in process engineering, where uncertainties are included. Several different approaches have been proposed to tackle this issue. Roupec et al. (2013) proposed a hybrid algorithm for network design problem with uncertain demands. A solid assessment of the Origin-Destination

Please cite this article as: Šomplák R., Nevrlý V., Smejkalová V., Pavlas M., Kůdela J., 2017, Verification of information in large databases by mathematical programming in waste management, Chemical Engineering Transactions, 61, 985-990 DOI:10.3303/CET1761162

matrix estimation techniques was presented by Bera and Rao (2011). The study by Karlaftis and Vlahogianni (2011), and Sun et al. (2006) investigate the trade-off between using more traditional statistical methods and neural networks in transportation problems. A similar investigation was conducted in the problems concerning failure and reliability predictions by Zio et al. (2012). In the case of the supply chain problems, the input data which creates the network (edges and nodes) have to be often verified using different techniques. The proposed articles utilise only one value for each parameter.
In some databases, the particular information might be reported by more subjects differently. This fact makes it easier to identify potential errors. First entity gives the amount of handover material, and the second one notes down the takeover. These amounts are supposed to be equal, but often this is not true. The flow principle is illustrated in Figure 1. It schematically describes the material flow through the network (from a source node over transfer nodes to target facility). The inconsistencies arise always between two neighbouring nodes, where for example source node handover certain amount (H1), but transfer node does not takeover the same amount (T1). The equivalent situation can be detected in H2 and T2, see Figure 1.

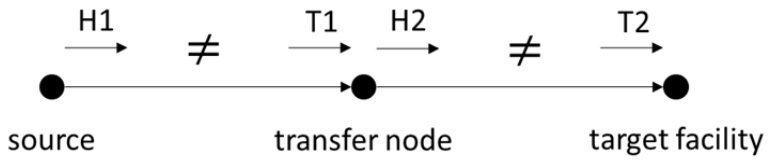

Figure 1: Schematic representation of reporting flows

Due to error illustrated in Figure 1, accurate values of material flow are unknown. These errors significantly restrict supply chain planning, where it is necessary to have a quality data from the previous period. Usually the errors are supposed to be random, but in this case, the systematic error is assumed. It is because of reporting one value by more subjects, where the error is expected from either of them.
This paper presents an approach based on optimisation methods which identify errors leading to inconsistencies in databases (Section 2). Section 3 focuses on a case study and the mentioned process is applied to the issue of waste management registration in the Czech Republic. The result is the model of production and assessment of transportation effectivity, which identify deficiencies in processing capacities within regions.

2. Mathematical model

The mathematical model is described within the waste management application. As noted previously, there are two sources of information for each shipment of waste. The first one is the node sending the waste, and these values are marked by index (-). The second information comes from a node receiving the waste designated by index (+). These values are considered as two equivalent scenarios. The verification of data should guarantee the validity of the mass balance in the nodes with minimal change in input data.

Sets:

$i \in I$ index of the node
$j \in J$ index of the arc

Parameters:

A_{ij}^{-} incidence matrix for sending
A_{ij}^{+} incidence matrix for receiving
x_j^{-} amount of waste shipped on the arc j according to the scenario -
x_j^{+} amount of waste shipped on the arc j according to the scenario +
w_j weight of the arc j, $w_j \in (0;1)$
o_i waste production in the node i
z_i waste processing in the node i
a threshold of zero penalization

Variables:

ε_j^{-} error on the arc j, scenario -
ε_j^{+} error on the arc j, scenario +

τ_i error in the node i
y_i penalization

Positive variables:
ε_j^{-+} positive part of the error ε_j^-
ε_j^{--} negative part of the error ε_j^-
ε_j^{++} positive part of the error ε_j^+
ε_j^{+-} negative part of the error ε_j^+
y_i^+ positive part of the penalization y_i
y_i^- negative part of the penalization y_i

$$\min \sum_{j \in J} \left(\varepsilon_j^{-+} + \varepsilon_j^{--} + \varepsilon_j^{++} + \varepsilon_j^{+-}\right) w_j + \sum_{i \in I} y_i^+ \tag{1}$$

s.t.

$$\sum_{j \in J} A_{ij}^- \left(x_j^- + \varepsilon_j^-\right) + \sum_{j \in J} A_{ij}^+ \left(x_j^+ + \varepsilon_j^+\right) + o_i + \tau_i - z_i = 0, \qquad \forall i \in I \tag{2}$$

$$x_j^- + \varepsilon_j^- = x_j^+ + \varepsilon_j^+, \qquad \forall j \in J \tag{3}$$
$$x_j^- + \varepsilon_j^- \geq 0, \qquad \forall j \in J \tag{4}$$
$$o_i + \tau_i \geq 0, \qquad \forall i \in I \tag{5}$$
$$\varepsilon_j^- = \varepsilon_j^{-+} - \varepsilon_j^{--}, \qquad \forall j \in J \tag{6}$$
$$\varepsilon_j^+ = \varepsilon_j^{++} - \varepsilon_j^{+-}, \qquad \forall j \in J \tag{7}$$
$$y_i = \operatorname{sgn}(a)(\tau_i - a o_i), \qquad \forall i \in I \tag{8}$$
$$y_i = y_i^+ - y_i^-, \qquad \forall i \in I \tag{9}$$

The objective function in Eq(1) minimises weighted total error on arcs and sum of positive part of the penalization for exceeding waste production. The construction of weights w_j is described below. Constraint Eq(2) is the equation for conservation of the mass balance in each node. The balance on each arc ensures equality of scenarios (+,-) for the sum of the data and found error. Constraints Eq(4) specifies non-negativity of flow on each arc for scenario (-). The same property for scenario (+) is ensured due to Eq(3). Similarly, Eq(5) is condition for non-negativity production. The errors for both scenarios (+,-) are decomposed into positive and negative part in Eq(6) and Eq(7). The assessment of penalty function for each node provides Eq(8). The calculation of a parameter (threshold for zero penalty) is described below. The last Eq(9) splits penalization y_i into positive and negative part. In the case of the waste registration, inconsistencies can be found also in total waste production and processing. This is probably caused by incorrect records and the errors are expected especially on the production side. This error rate on the production side is caused by recalculating the production according to the rules of the annual report, so a high precision of processing data may be assumed. This inconsistency problem can be solved by integrating penalty into optimisation tasks, specifically in the objective function Eq(1). The illustration of penalty function is shown in Figure 2 and arising equation that this principle transforms into a model is Eq(8). The computation of threshold for zero penalty a was set according to Eq(10), which correspond to the ratio of an average change of production and average production to maintain the balance with the processing. Parameter a_i determines the ratio of error for each producer, which can be reached without being penalized, assuming that each producer participates in the error equally. The sign of a is determined by differences between production and treatment from Eq(10) and affects the form of equation Eq(8).

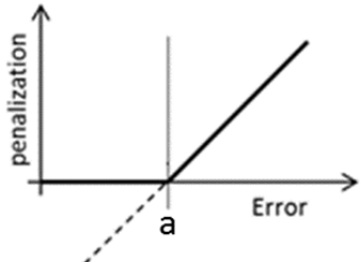

Figure 2: The choice of penalty function

$$a = \frac{\sum_{i \in I}(z_i - o_i)}{\sum_{i \in I} o_i} \tag{10}$$

More significant errors occur in the data about the waste streams. In addition, the recalculation of production does not take into account the flow due to lack of awareness of processing site for a specific producer. To sum up, an inconsistency in the database may arise in the following manner: incorrect evidence, incorrect data processing, reporting to company headquarters, missing re-count of production and others. To distinguish between these inconsistencies, the weights w_j were introduced. Each weight is based on the similarity of the transmitted and received quantities. The more these values are equal, the greater weight is chosen, to ensure that the respective error is lower. To ensure this feature, weights were selected as in the equation Eq(11).

$$w_j = \begin{cases} 1, & x_j^-, x_j^+ = 0 \\ 1 - \dfrac{|x_j^- - x_j^+|}{\max(x_j^-, x_j^+)}, & otherwise \end{cases} \qquad \forall j \in J \tag{11}$$

As a result of the presented mathematical model, errors in production and transported amount of waste for all arcs are estimated. The results indicate the efficiency of the transport of waste and self-sufficiency of each region with regard to the processing capacities and the appropriateness of the current deployment. On this basis, the region with great potential for economic and environmental improvements can be identified.

3. Case study

The case study is an example of model utilisation for the balance of mixed municipal waste in the Czech Republic in 2015. The input data were provided by the Ministry of the Environment of the Czech Republic. One of the contributions of this computation is information about the waste production and therefore also about the necessity of waste export. The modelled production for each node consist of sum of input data (o_i) and error (τ_i). In the most of territorial units, the resulting τ_i has ended with threshold value (a), only few of them has been penalized. It has been given by waste flows with a big weight (w_i).

Figure 3 illustrates a comparison between the modelled waste production and processing for each region. At first sight, significant differences can be observed. Some regions have enough capacity for the processing the waste they produce (their own waste) and can even import a foreign waste, e.g. Central Bohemian Region, Pardubice Region, Liberec Region, or South Moravian Region. Transport over long distances is not desirable, but it should be noted that some exports between the regions are advantageous for geographical and environmental reasons. Regions Plzeň and Moravian-Silesian are characterised by a lack of capacity. These areas are dependent on the processing in other territories. It points out on the potential for more efficient transport of waste relative to the economic and environmental aspects, in the case, when processing capacities would be increased.

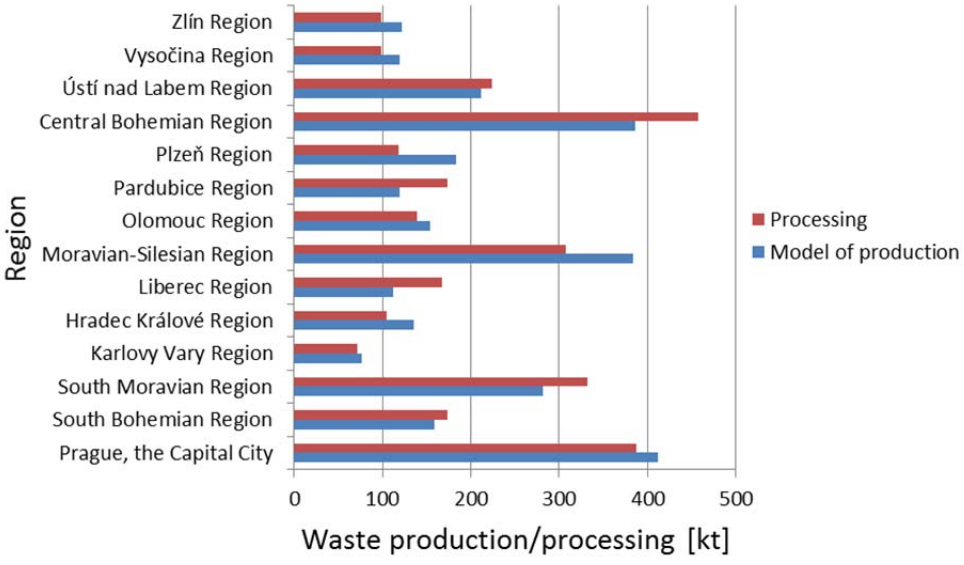

Figure 3: Comparing the production and processing of mixed municipal waste in the various regions.

The optimum operation has Karlovy Vary Region, where the quantities of treated waste are nearly the waste production. If the production after the balance is in accordance with processing, it does not mean that there is not import and export. The Ústí nad Labem Region is worth mentioning. It has an acceptable ratio of production and processing, however, considerable interregional transport takes place. This is due to the size of the territory and a small number of processing sites. The similar situation is in the Prague, the Capital City. This is a specific case in the Czech Republic, the region comprises of the only city. The attention should be also concentrated on the regions Zlín and Vysočina, which show an almost ideal data in terms of processing on their own territory.
The average distance in km for transportation of produced waste is depicted in Figure 4, which shows two types of transport (within and outside the region). In this respect, the Hradec Králové Region greatly exceeds all others and this is the only region that is so different. The problem is not the fact that region exports waste beyond its own area, but overall travelled kilometres, which cause a high impact on the environment. This happens due to the inappropriate deployment of processing capacities, but reveal the space for efficiency improvement.

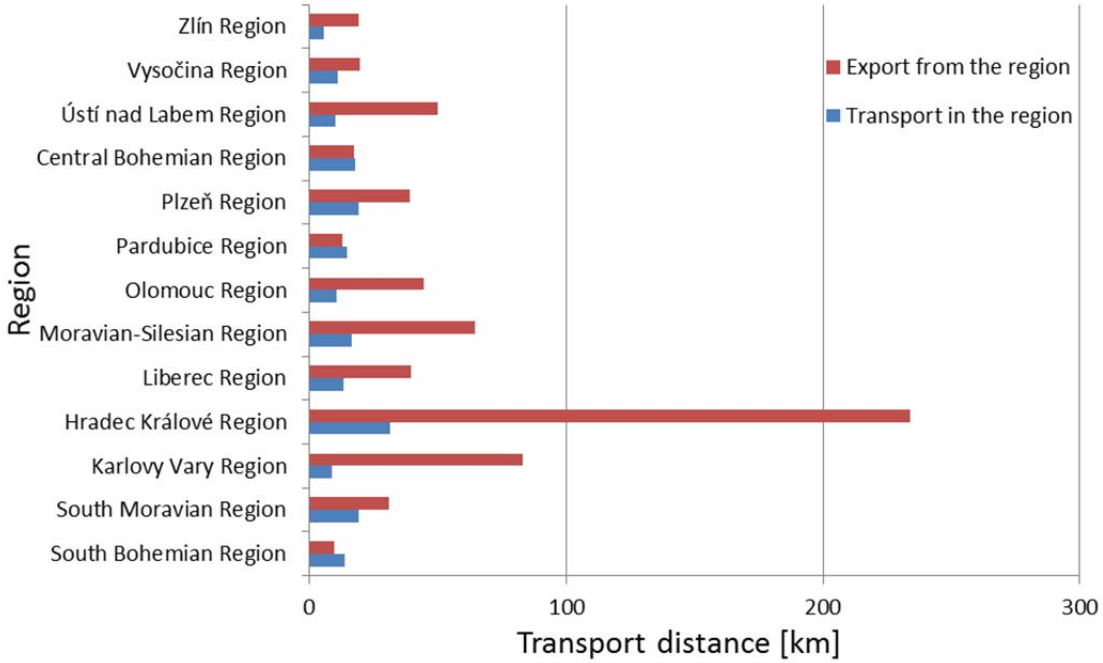

Figure 4: The average transported distance in km related to one produced tonne of waste

Generally, exporting usually dominates, with some exceptions. The reason is the inefficient distribution of processing facilities within the particular region, even though some regions are almost self-sufficient. According to the stated aspects, it can be concluded, that waste is treated effectively in regions Vysočina and Zlín despite the little lack of their own processing capacity. Although they are forced to process part of the production elsewhere, it is multiple times less than in other cases in terms of km.

Table 1: The aggregated results of waste flows beyond the Karlovy Vary Region

Micro-region	—>O			O—>		
	Handover [t]	Reporting [t]	Result [t]	Reporting [t]	Takeover [t]	Result [t]
Aš	56	9	9	930	492	1,320
Cheb	398	11	16	11,760	12,427	11,756
Karlovy Vary	3,664	658	664	3,955	10,966	4,613
Kraslice	0	0	0	53	46	57
Mariánské Lázně	83	0	0	4,012	6,511	4,360
Ostrov	328	0	25	4,771	10,750	5,055
Sokolov	31,656	8,780	27,105	3,346	3,138	4,905

The detailed illustration of results for the Karlovy Vary Region is in Table 1. This region includes a total of seven lower territorial units – micro-regions. The symbol **—>O** means the total import to the micro-region and handover (tagged as H in Figure 1) denotes the amount of waste which flows from micro-regions beyond the Karlovy Vary Region. The reporting (tagged as T in Figure 1) in this context means the amount of waste which was reported in the particular micro-region. Conversely, **O—>** is a total export from the particular micro-region and the reporting here swaps the role with the takeover.

4. Conclusions

This paper has introduced the mathematical model, which allows estimation of errors in current flows which were reported in the waste management database. The computation was presented at the case study of mixed municipal waste in the Czech Republic in 2015. The results demonstrate the potential for more efficient transport of waste within individual regions. It shows the dependence on other regions and the degree of export. A big contribution is in the identification of regions with insufficient capacities (high export) or bad deployment of processing facilities (high export and import). This analysis could serve as the initialization process before computing another optimisation tasks about potential construction of new treatment facilities. For routing in cities, Viktorín et al. (2016) proposed differential evolution algorithm while the whole topic is covered by Šomplák et al. (2013), where authors allocate facilities to optimally transport the waste. In the future research, these approaches could be integrated into one multiphase tool.

Acknowledgments

The authors gratefully acknowledge financial support provided by Technology Agency of the Czech Republic within the research project No. TE02000236 "Waste-to-Energy (WtE) Competence Centre" and by the project Sustainable Process Integration Laboratory – SPIL, funded as project No. CZ.02.1.01/0.0/0.0/15_003/0000456, by Czech Republic Operational Programme Research and Development, Education, Priority 1: Strengthening capacity for quality research.

References

Bera S., Rao K.V.K., 2011. Estimation of origin-destination matrix from traffic counts: the state of the art. European Transport, 49, 3-23.

Čuček L., Klemeš J. J., Varbanov P.S., Kravanja Z., 2011. Life cycle assessment and multi-criteria optimization of regional biomass and bioenergy supply chains. Chemical Engineering Transactions, 25(1), 575 - 580.

Karlaftis M.G., Vlahogianni E.I., 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation Research Part C: Emerging Technologies, 19(3), 387-399.

Roupec J., Popela P., Hrabec D., Novotný J., Olstad A., Haugen K.K., 2013. Hybrid algorithm for network design problem with uncertain demands. World Congress on Engineering and Computer Science, 1, 554-559.

Stille Z., Bertók B., Friedler F., Fan L. T., 2011. Optimal design of supply chains by P-graph framework under uncertainties. Chemical Engineering Transactions, 25(1), 453 - 458.

Sun S., Zhang C., Yu G., 2006. A Bayesian Network Approach to Traffic Flow Forecasting. IEEE Transactions on Intelligent Transportation Systems, 7(1), 124-132.

Šomplák R., Procházka V., Pavlas M., Popela P., 2013. The Logistic Model for Decision Making in Waste Management. Chemical Engineering Transactions, 35(1), 817 - 822.

Viktorín A., Hrabec D., Pluháček M., 2016. Multi-chaotic differential evolution for vehicle routing problem with profits. 30th European Conference on Modelling and Simulation, 30, 245-251.

Zio E., Broggi M., Golea L.R., Pedroni N., 2012. Failure and reliability predictions by infinite impulse response locally recurrent neural networks. Chemical Engineering Transactions, 26(1), 117 - 122.