

Distr.  
GENERAL

CES/SEM.47/15  
13 March 2002

ENGLISH ONLY

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE  
EUROPEAN COMMUNITIES**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROSTAT**

**Joint UNECE/Eurostat Seminar on Integrated Statistical  
Information Systems and Related Matters (ISIS 2002)**  
(17-19 April 2002, Geneva, Switzerland)

Topic II: Secure communications and confidentiality

**CONFIDENTIALITY ISSUES  
FOR STATISTICAL DATABASE QUERY SYSTEMS**

**Invited Paper**

Submitted by National Center for Health Statistics, CDC/DHHS, United States of America<sup>1</sup>

**I. INTRODUCTION**

1. National statistical offices (NSOs) are responsible to collect, verify and refine statistical data in order to make reliable information available to policy makers and the public. By law or regulation and ethical practice, the NSO must do so in a manner that preserves the *confidentiality* of data pertaining to individual entities such as persons, businesses, and health care providers. Notably exempt from confidentiality concerns are statistics pertaining to governmental units.
2. Prior to 1960, NSOs made statistical information available primarily in the form of computed or estimated *tabulations*, defined by cross-classification of only one, two or a small number of variables. The NSO determined which tabulations to release, first in printed form and later also in electronic form. Confidentiality protection, more recently called *statistical disclosure limitation*, was accomplished by suppressing or combining selected tabulations or entire sets of tabulations or, less frequently, by altering tabulations slightly through rounding or incorporation of random noise. In effect, the NSO first determined which tabulations were worth releasing and then released correspondingly less information in consideration of confidentiality and data quality concerns.
3. During the 1960s, first with the Continuous Work History Sample of the U.S. Social Security Administration, followed by Public Use Microdata Samples (PUMS) from the 1960 and subsequent U.S. Decennial Censuses, NSOs began releasing *statistical microdata files* comprising records pertaining to individual entities (mostly, persons). The data user was now free to create all conceivable summaries from the

---

<sup>1</sup> Prepared by Lawrence H. Cox (LCOX@CDC.GOV)

unit record data and, equally important, to fit statistical, demographic or econometric models to the microdata. Statistical disclosure limitation became focused on altering or removing selected microdata records.

Longitudinal data presented confidentiality problems that remain largely unsolved. Emerging research is directed towards fitting the data to complex statistical models and releasing instead model-derived *synthetic microdata* and/or the models themselves. Disclosure limitation for tabulations and microdata are provably complex, both theoretically and computationally.

4. NSOs are considering allowing data users direct access to statistical data bases, either on a public or restricted access basis, via a *statistical data base query system*. This heightens confidentiality concerns and issues, and will motivate disclosure limitation research in coming decades. In this paper, we investigate through examples some of the confidentiality and data usefulness problems raised by the advent of statistical data base query systems. We focus on two query paradigms: tabulations from a data base organized as a large multi-dimensional contingency table and simple statistical models derived from the data base, namely, ordinary least squares regressions and best linear unbiased prediction (*kriging*) models for spatial data.

## II. THE STATISTICAL DATABASE

5. For purposes here, a *statistical database* is equivalent to an *n-dimensional contingency table*: an enumeration of the units from a sample or population with respect to  $n$  cross-classified categorical variables. Each categorical variable  $i$  comprises  $d_i$  mutually exclusive and exhaustive *characteristics*  $c_{ik}$ . The size of the  $n$ -dimensional contingency table is  $d_1 d_2 \dots d_n$ . Each *internal entry*  $t_{i_1 i_2 \dots i_n}$  of the table equals the number of units with characteristics  $(i_1, i_2, \dots, i_n)$ . Internal entries therefore assume nonnegative integer values. This characterization is quite general and flexible. In particular, if every record in the underlying microdata file is uniquely identified by some combination of characteristics, then this characterization encompasses the underlying microdata file. If not, at least in principle the same characterization is achieved by including an additional dimension defined by a unique identifier, such as social security number.

6. The table has many *marginal totals* corresponding to sums along one or more dimensions. *k-dimensional marginal totals* are totals along  $(n - k)$  dimensions. General mathematical notation for marginal totals is available, but somewhat cumbersome. Section IV deals with complexities in  $n$ -dimensional tables, namely, properties that hold, e.g., in two dimensions, but fail entirely or in certain instances in higher dimensions. Examples will be drawn from three and four dimensional tables. Notation specific to these examples will be provided as needed.

## III. CONFIDENTIALITY ISSUES IN STATISTICAL DATABASES

7. If a sample or population unit (*entity*) has one or more characteristics unique from those of the other units, then a third party potentially can identify the entity based on these *identifying characteristics*. In some instances, the simple act of identification is a breach of confidentiality. More typically, identification is based on fewer than the full set of  $n$  characteristics, resulting in disclosure of the remaining nonidentifying characteristics. If precisely two entities possess certain characteristics, then each potentially can identify the other and disclose confidential information. In general, *statistical disclosure* in contingency tables occurs when small counts are released or can be inferred. What constitutes “small” varies from one NSO to another, with, for example, the value being any positive value less than five at the U.S. Census Bureau while at the U.S. Internal Revenue Service and at Statistics New Zealand the *threshold* is not five but three.

8. The number of entries in a  $n$ -dimensional contingency table typically is large and grows quickly with increasing dimension  $n$ . For example, even with all categorical variables dichotomous, the number of internal entries in a 30-dimensional table exceeds one billion. Most internal entries and higher dimensional marginal totals are likely to be small, in fact, zero or one. In this context, our notion of a *statistical database query system* is as follows. The database user can query the system as often as it likes, but each request must be for a marginal total. Of course, correct answers cannot be provided to queries corresponding to marginal totals not exceeding the threshold, but typically doing so in and of itself does not prevent a third party from deducing small entries, due to the additive structure of the table. Further *disclosure limitation* is required.

9. In two dimensional tables, it is possible to *round* all entries and totals in a manner that preserves additivity of internal entries to marginal totals. If all entries are rounded to multiples of the threshold, then disclosure limitation is complete. Similarly, it is possible to *perturb* entries slightly using additive random noise while preserving additivity. Small values remain, but the imprecision introduced through the perturbation is regarded as sufficient for disclosure limitation. Unfortunately, as demonstrated in the next section, it is not always possible to round or perturb entries in this manner in dimension  $n > 2$ . A third disclosure limitation method, *complementary suppression*, viz., the process of selectively suppressing entries to mask small entries, is complicated (indeed, *NP-hard*) even in two dimensions.

10. One approach to disclosure limitation in an  $n$ -dimensional statistical database is to answer only queries corresponding to lower dimensional marginal totals. The confidentiality issue is then whether the released totals can be used to infer small values. There are three aspects to this problem.

11. The first is: Can small values be inferred deterministically? This would be accomplished through manipulation of linear (additive) relationships between entries and the released marginal totals. This is essentially a problem in mathematical programming: Is the *feasible region* delimited (*constrained*) by the released marginals and nonnegativity of entries sufficient to ensure that each entry takes on at least one value at or above the threshold? Normally, this would correspond to a sequence of linear programming problems—one to minimize and one to maximize each internal entry or marginal of interest over the feasible region, resulting in *exact bounds*  $[\min \{t_{i_1, \dots, i_n}\}, \max \{t_{i_1, \dots, i_n}\}]$  for internal entries. This is a challenging but for the most part computationally tractable undertaking. Unfortunately, because entries must be integer, to yield *exact integer bounds*  $[\min \{t_{i_1, \dots, i_n} : t \text{ integer}\}, \max \{t_{i_1, \dots, i_n} : t \text{ integer}\}]$ , the NSO apparently is confronted with a massive integer programming problem, impossible to solve. This is explored in Section IV.

12. The second aspect of the problem is: Can small values be inferred probabilistically? This would be accomplished using distributional models from the theory of log linear models and simulation. Some of the underlying mathematical issues here overlap with those raised in exact integer bounding. This problem is not addressed further here.

13. The third aspect of the problem is: How to manage the query response strategy? The confidentiality problem is dynamic, namely, the response to successive new query potentially increases information about unreleased internal entries and marginals. One solution is to respond to queries on a flow basis, refusing any query that breaches confidentiality, and ending when no further queries can be answered safely. Another approach is to predetermine a (maximal) set of queries that can be mutually answered safely and only to release information in response to these queries. Both approaches are computationally intensive and complex. These problems are not addressed further here.

**IV. PROPERTIES OF HIGHER DIMENSIONAL TABLES**

14. This section comprises a series of examples demonstrating the failure in higher dimensions of properties enjoyed by two-dimensional tables. Attempt is made to keep examples as uncomplicated as possible in order to emphasize essential features. All examples are of modest size and, with the exception of one four-dimensional table, are three-dimensional.

15. Cox and Ernst (1982) demonstrate that in two-dimensional contingency tables *controlled rounding*, viz., rounding entries to a fixed integer rounding base while assuring that rounded and original entries differ by less than the base and that additivity to marginals is preserved, always can be accomplished. In addition, it is possible to ensure that any original entry equal to a multiple of the base remains fixed (*zero-restrictedness property*). Example 1, from Causey, Cox and Ernst (1985), depicts the internal entries of a three-dimensional table of size 2x2x2. Examination reveals that zero-restricted controlled rounding is not possible for Example 1, and consequently is not assured in three and higher dimensions. Ernst (1989) exploits this fact to construct a three-dimensional table for which a controlled rounding does not exist.

0.5	0
0	0.5

0	0.5
0.5	0

Example 1: Zero-Restricted Controlled Rounding Fails in Three Dimensions  
(source: Causey, Cox and Ernst 1985, p. 907)

16. *Controlled random perturbation* is based on selecting a small positive perturbation value and alternately adding and subtracting it to/from original values while preserving additivity to marginals. Zero counts cannot be reduced, and therefore random perturbation must be zero-restricted. Arguments entirely analogous to those for controlled rounding show that controlled random perturbation is always possible in two-dimensional tables. Example 2, a three-dimensional table of size 3x3x3, from Cox (2000), illustrates that controlled perturbation fails in three and higher dimensions. The “\*” symbol denotes any positive value. It is not possible to alternate +/- movement of a positive quantity between nonzero values (\*) while preserving additivity to fixed marginals.

0	*	*
*	0	*
*	*	0

*	*	0
*	0	*
0	*	*

*	0	*
0	0	0
*	0	*

Example 2: Controlled Random Perturbation Fails in Three-Dimensions (\* = positive entry)  
(source: Cox 2000, Example 6)

17. Two vectors of nonnegative integers whose entries add to a common value are *consistent*. In two dimensions, a consistent pair of integer vectors assures the existence of one (typically, more) two-dimensional contingency table whose one-dimensional marginal (row and column) totals are given by the respective vectors. However, in n-dimensions, n consistent vectors of nonnegative integers do not necessarily comprise the (n-1)-dimensional marginal totals for any n-dimensional contingency table. This is demonstrated in the three-dimensional Example 3 of size 3x4x6, from Vlach (1986), in which consistent integer two-dimensional marginals define a unique nonnegative table which, however, contains some noninteger entries. Indeed,

consistent integer marginals can lead to an entirely *infeasible* situation, viz., no integer or even continuous table exists, as demonstrated in Example 4, from Cox (2000). The “+” sign indicates the dimension over which the marginal is computed, e.g., in Example 3,

$$t_{+61} = \sum_{i=1}^4 t_{i61} = 0 \text{ and, in Example 4, } t_{13+} = \sum_{k=1}^3 t_{13k} = 3 \text{ .}$$

$$(t_{ij+}) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (t_{i+k}) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \quad (t_{+jk}) = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Example 3: Consistent Integer Marginals Fail to Assure an Integer Three-Dimensional Table  
(source: Vlach 1986, p. 77)

$$(t_{ij+}) = \begin{pmatrix} 1 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 1 \end{pmatrix}, \quad (t_{+jk}) = (t_{i+k}) = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

Example 4: Consistency Fails to Assure a Feasible Three-Dimensional Table  
(source: Cox 2000, Example 4a)

18. Assessment of disclosure risk in statistical tables and tabulations, referred to as *disclosure audit*, is the process by which to address the question of paragraph 11: Is the table safe from deterministic attempts to infer small values? This requires a mechanism for determining exact lower and upper bounds for each internal entry. In two dimensions, this can be accomplished using simple formulae (Cox 2002). In higher dimensions,, such formulae are not available except in specialized cases. It might appear that exact bounds could be computed using linear programming: For each internal entry  $x$ , solve one linear program to compute  $\min \{x\}$  and a second to compute  $\max \{x\}$ . This is tractable computationally and can be accomplished with far fewer optimizations if interrelationships between bounds are exploited. This process would be sufficient for disclosure audit under any of the following conditions. One, if all *extremal points* of the linear programming polytope were integer-valued. Two, if every exact lower and upper bound occurred at one or more integer-valued points of the polytope, and an algorithm were available to direct the linear program to one such point for each bound. Three, the *integer rounding property* (Nemhauser and Wolsey 1988, 594-598) holds for each bound, viz., the exact integer bound corresponds to rounding the exact continuous bound down or up, respectively, to the nearest integer. The first condition holds in two dimensions, and therefore so do the second and third. Unfortunately, all three conditions fail in higher dimensions, meaning that linear programming is not a viable method on which to base procedures for disclosure audit in general higher dimensional tables.

19. Failure of the first condition is illustrated in Example 5, which displays all prescribed two-dimensional marginals for a set of 4x4x4 three-dimensional tables, from Cox (2000). Failure of the second condition is illustrated by Example 6, which displays a noninteger extremal solution at which  $\max \{t_{312}: t \text{ integer}\} = 1$  is achieved on the polytope of 3x3x3 three-dimensional tables with all one-dimensional marginals prescribed,

from Cox (2002). Failure of the integer rounding property is illustrated by Example 7, which displays one totally integer table corresponding to an extremal point among  $2 \times 2 \times 2 \times 2$  four-dimensional tables with all two-dimensional marginals prescribed, due to Sturmfels (2002). The two-dimensional marginal totals corresponding to this table define a feasible region in 16-dimensional space formed by intersection of a five-plane with the first orthant. Nevertheless, this feasible region contains precisely one totally integer point, shown in Example 7. Thus, the integer rounding property fails because, e.g.,  $\max \{t_{1121}\} = 5/3$  but  $\max \{t_{1121}: t \text{ integer}\} = 0$ .

$$t_{ij+} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad t_{i+k} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad t_{+jk} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Example 5: Table with Fractional Continuous Exact Bound ( $\max \{t_{331}\} = 1/2$ )  
(source: Cox 2000, Example 7c)

$$t_{ij1} = \begin{pmatrix} 0 & 0 & 1.5 \\ 0 & 0 & 0 \\ 0.5 & 0 & 0 \end{pmatrix}, \quad t_{ij2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad t_{ij3} = \begin{pmatrix} 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 \end{pmatrix}$$

Example 6: Noninteger Extremal Point for  $\max \{t_{312}: t \text{ integer}\} = 1$ , Subject to Prescribed One-Dimensional Marginals  $t_{i++} = t_{j++} = t_{+++} = (2 \ 1 \ 2)$   
(source: Cox 2002, Example 18)

$$t_{ij11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad t_{ij12} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad t_{ij21} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad t_{ij22} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Example 7: Unique Integer Point for Four-Dimensional Table with Prescribed Two-Dimensional Marginals  
(source: Sturmfels 2002)

## V. LINEAR AND SPATIAL PREDICTION USING STATISTICAL DATABASES

20. An alternative output model for a statistical database is to release only regression coefficients as requested by users. Refusing, perhaps, to release regressions representing nearly perfect fit, this appears to be a safe release strategy. While for the most part this may be so, it is possible to construct scenarios under which disclosure occurs. Such scenarios, while unlikely to occur in practice, are instructive towards developing strategies for safe release. One such scenario is presented in the next paragraph.

21. Under simple linear ordinary least squares regression, assume that the user has requested regression of Y (say, income) on X (say, age) for all p database units with specific characteristics (say, statisticians in a particular city under the age of 80). The database returns the regression coefficient  $\beta$ . Next, the user requests the same regression, but this time for all (p + m) database units satisfying more general characteristics (say, statisticians in the city under the age of 90). The database returns the regression

coefficient  $\beta^*$ .  $\bar{x}$ ,  $\bar{y}$  denote the X- and Y-means of the m additional database records. As regression reproduces means, then

$$\begin{aligned}\bar{Y}_{90} &= \bar{X}_{90}\beta^* \\ (p\bar{Y}_{80} + m\bar{y})/(p + m) &= \bar{X}_{90}\beta^*\end{aligned}$$

Therefore,

$$\begin{aligned}\bar{y} &= ((p + m)/m)\bar{X}_{90}\beta^* - (p/m)\bar{Y}_{80} \\ &= \bar{x}\beta^* + (p/m)(\bar{X}_{80}\beta^* - \bar{Y}_{80})\end{aligned}$$

viz.,  $\bar{y}$  can be precisely determined. If  $m = 1$  and the one statistician in the city of age 80-90 can be identified, then that statistician's income is precisely determined. If  $m = 2$ , then either of the two elderly statisticians could subtract his or her income from  $\bar{y}$  and again precisely determine the income of the other statistician. In general, if  $m$  is small, some disclosure is possible.

22. The question arises: Does adding noise to the x-variables limit disclosure in regression outputs? The simple linear regression is:  $Y = X\beta$ . Add zero-mean IID noise to the X-data  $X_p = \{x_i\}_{i=1}^p$ . In lieu of releasing the true regression, the NSO generates zero-mean IID noise  $\epsilon_p$  and creates p noisy data points  $(x_k + \epsilon_k, y_k)$ . Simple linear regression on the noisy data results in the regression model:

$$\begin{aligned}Y &= (X + \epsilon)\gamma = X[Cov(X_p + \epsilon_p, Y_p)/Var(X_p + \epsilon_p)] \\ &= X(1/(1 + Var(\epsilon_p)/Var(X_p)))\beta\end{aligned}$$

The user now requests an updated regression that in addition includes m additional data points  $\{x_{p+k}\}_{k=1}^m$ . m additional noisy data points  $\{(x_{p+k} + \epsilon_{p+k}, y_{p+k})\}_{k=1}^m$  are created and an updated regression performed:

$$Y = (X + \epsilon_{p+m})\gamma^* = X(1/(1 + Var(\epsilon_{p+m})/Var(X_{p+m})))\beta^*$$

If, as often is the case,  $Var(\epsilon_{p+m})$  is known, then disclosure can be achieved as in the first section. Otherwise, as  $Var(\epsilon_{p+m})/Var(X_{p+m})$  is small, approximate disclosure is possible.

23. (Ordinary) kriging is a method for best linear unbiased prediction of spatially referenced data. Observations  $\{Z(x_1), \dots, Z(x_s)\}$  are made at known locations  $X = \{x_1, \dots, x_s\}$  identified, e.g., by latitude and longitude, and are fit to a covariance model  $Cov(Z(x), Z(x'))$ , from which a spatial (kriging) model is developed and used to predict the value of  $Z(x)$  at unobserved locations  $x$ . See Cressie (1993, Chapter 3) for details. If, e.g.,  $Z$  is Gaussian, then the best linear unbiased predictor is given by:

$$\hat{Z}(x) = \sum_{i=1}^s \lambda_i(x)Z(x_i)$$

with  $\sum_{k=1}^s Cov(Z(x_i), Z(x_k)) \lambda_k(x) = Cov(Z(x), Z(x_i))$  and  $\sum_{i=1}^s \lambda_i(x) = 1$ .

24. The confidentiality issue is whether it is safe for the NSO to release the kriging model. The answer is no: Because  $\hat{Z}(x_i) = Z(x_i)$  and because locations are typically public knowledge, release of the kriging model results in exact disclosure of  $Z$ -data at the observed locations  $\mathbf{X}$ .

25. What disclosure limitation options are available to the NSO? It is not possible to add noise  $\epsilon$  to the locations, as the  $Z(x_i + \epsilon_i)$  are unknown. One possibility is as follows: (1) Kriging based on  $\{Z(x_i)\}$ , resulting in  $\hat{Z}(x)$ . (2) Generate zero-mean IID noise  $\epsilon$ . (3) Kriging based on  $\{\hat{Z}(x_i + \epsilon_i)\}$ , resulting in  $\hat{Z}_\epsilon(x)$ . (4) Release  $\hat{Z}_\epsilon(x)$ .

26. A second, more speculative, possibility is: (1) *Jiggle* the covariance matrix, viz., given  $\mathbf{K} = (\text{Cov}(\mathbf{Z}(x_i), \mathbf{Z}(x_j)))$ , create  $\mathbf{K}_\epsilon = \mathbf{K} + \epsilon$ . (2) Kriging based on  $\mathbf{K}_\epsilon$ . This however is tricky as the effects of small perturbations to entries of  $\mathbf{K}$  on covariance and the resulting spatial model are unclear. In particular, it is not clear if or how to ensure that  $|Z(x_i) - \hat{Z}_\epsilon(x_i)|$  is sufficiently large, but not too large.

## VI. CONCLUSIONS

27. It can be argued that the next evolution in the release of statistical data by NSOs is statistical data base query systems. This moves the NSO into the arena of releasing tabulations from high dimensional and linked tabular structures. This on the one hand magnifies disclosure risk and on the other presents seemingly monumental theoretical and computational challenges to the NSO as it attempts to assess and limit statistical disclosure.

28. Strategies for releasing statistical models in lieu of original data or tabulations have been proposed to address confidentiality concerns. Based on simple evidence gained by examining linear regression and spatial prediction models, it is clear that the advantages and limitations of doing so need to be carefully assessed.

## References

- Causey, B.D., L.H. Cox and L.R. Ernst (1985). Applications of transportation theory to statistical problems. *Journal of the American Statistical Association* **80**, 903-909.
- Cox, L.H. (2000). Properties of multi-dimensional statistical tables. Submitted.
- \_\_\_\_\_ (2002). Bounds on entries in 3-dimensional contingency tables subject to given marginal totals. **Inference Control in Statistical Databases** (J. Domingo-Ferrer, ed.), Lecture Notes in Computer Science 2316. Heidelberg: Springer-Verlag, 21-33.
- Cressie, N.A.C. (1993). **Statistics for Spatial Data**. New York: Wiley-Interscience.
- Ernst, L.R. (1989). Further applications of linear programming to sampling problems. Technical Report– Census/SRD/RR-89-05. Washington, DC: U.S. Census Bureau, Department of Commerce, 33 pp. Available: <http://www.census.gov/srd/www/byname.html>
- Nemhauser, G.L. and L.A. Wolsey (1988). **Integer and Combinatorial Optimization**. New York: John Wiley & Sons, Inc.
- Sturmfels, Bernd (2002). Personal communication. March 5, 2002.
- Vlach, M. (1986). Conditions for the existence of solutions to the three-dimensional planar transportation problem. *Discrete Applied Mathematics* **13**, 61-78.