AUSTRIAN JOURNAL OF STATISTICS
Volume 27 (1998), Numbers 1&2, 99–107

Performance Evaluation of Ground Water Reference Material by Statistical Methods

Karina Schramm-Nielsen[1], Jill Merry[2], Bente A. Nyeland[3], Niels H. Spliid[4], Henrik Spliid[1]
[1]IMM, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, [2]VKI, Water Quality Institute, Hoersholm, [3]NERI, National Environmental Research Institute, Roskilde, [4]Danish Institute of Agricultural Sciences Research Centre, Flakkebjerg, Denmark

Abstract: As part of the Danish ground water monitoring program a multi-component reference material for laboratory use has been developed. After the certification procedure was completed, further analysis was undertaken to investigate any possible effect of the analytical methods used by the laboratories. In the present case, a group of similar compounds were selected for the statistical analyses. Data was imbalanced, therefore a mixed model approach has been applied using the MIXED procedure of SAS®. Method effect was significant in two cases out of six. Some of the data sets caused problems with non positive definite G-matrices, probably because of too sparse data.

Keywords: Non Positive Variances, Mixed Procedure, Mixed Model, Reference Material, Unbalanced Data.

1 Introduction

The formal definition of a reference material is given in ISO (1992).

> A material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the calibration of a apparatus, the assessment of a measurement method or for assigning values to materials.

Environmental reference materials are used in chemical laboratories to assess analytical quality on environmental samples. A chemical reference material contains the same components and has a matrix comparable with environmental samples to be analysed. The need for chemical reference materials arises from the fact that environmental samples are often complex and difficult to analyse. Samples of soil, lake water, sludge, wastewater etc. contain a lot of compounds other than the one(s) relevant in a particular analysis. These compounds are likely to cause difficulties in the analytical procedure and may obscure the measurement of the compounds in focus. Most chemical analytical procedures include a calibration step where the measuring equipment is calibrated with solutions of the pure compounds to be measured. It is the large difference in chemical composition between the complex unknown sample and the solution of the pure measurand which calls for the use of chemical reference materials. The environmental reference material has the advantage over the pure solution of having a more complex composition and it is, therefore, more representative in relation to real samples. Furthermore, it has a known content of the measurand because certified chemical reference materials are labelled with certified

consensus values. Consensus values are often determined by means of interlaboratory comparisons according to ISO (1989).

The object of this paper is to evaluate the different analytical methods applied by the participants in a certification procedure.

The reference material in question contains 22 pesticides in a single ampoule. In sample preparation and analysis there are variations that depend on the pesticide to be analysed. As a result of this, correlation between groups of compounds treated by the same preparative or analytical procedures is expected. Because of the multidimensional nature of the material and the restrictions imposed by analytical techniques it is desirable to investigate laboratory and method performance more closely. This study focuses on a selection of pesticides and laboratories from the original data set.

2 Certification Data

The original data set contained 22 records with certification data. Measurements of 22 pesticides were recorded from about 20 laboratories. The material was certified in accordance with the directions given ISO (1993), ISO (1989), ISO (1992). The design of the certification study was a fully nested design with three factors: laboratory, day and sample. The participants were asked to perform analyses on 2 different days analysing two ampoules with replicates on each day. The purpose of the certification study is to establish the reference value of the material. The mathematical model applied for certification is

$$Y_{ijkl} = \mu + L_i + D(L)_{j(i)} + A(LD)_{k(ij)} + E_{l(ijk)} \tag{1}$$

μ represents the overall mean and the sources of variation are

Table 1: Sources of variation in data collected for certification of pesticide reference material

Source of Variation	Symbol	Indices	Assumptions
Laboratories	L_i	$i = 1$ to a	$L_i \in N(0, \sigma_0^2)$
Days within laboratories	$D(L)_{i(j)}$	$j = 1, 2$	$DL_{i(j)} \in N(0, \sigma_1^2)$
Replicates within days (ampoules)	$A(LD)_{k(ij)}$	$k = 1, 2$	$A(LD)_{k(ij)} \in N(0, \sigma_2^2)$
Error	$E_{l(ijk)}$	$l = 1, 2$	$E_{l(ijk)} \in N(0, \sigma_r^2)$

The pesticide analyses are complicated, therefore it cannot be expected that all data sets will be complete. Thus, the data sets were unbalanced in the sense that the number of laboratories who analysed (i.e. the number of observations) differed from one compound to another and in many cases the total number of results for one compound from one laboratory was less than 8 (two ampoules analysed with 2 replicates on two days).

3 Subsets of Data

6 compounds out of the 22 available were chosen for this study. The 6 compounds were selected for further investigations with regard to method of analysis applied by the laboratory.

Full information about methods of analysis was not available for all laboratories. Among the laboratories who did provide the full information about methods of analysis, a number were selected based on their considerable experience in pesticide analyses. The 6 selected pesticides were chosen on the basis of similarity with respect to chemical properties. The reason for choosing a group of compounds with chemical resemblance was the expectation that differences caused by method of analyses would then affect several compounds.

Table 2: Selected pesticides and laboratories. The method code covers a combination of sample preparation and detection method. Methods AA-AC include the same type of sample preparation but with different detection methods. Methods DD-DF form a second group with common sample preparation and varying detection methods.

	Method					
Compound	AA	AB	AC	DD	DE	DF
Cyanazine	22		3,5,6,9,11,14		10	8,12
Dichlorprop	16	4,7,13	3,5,6,9,11,14	2	10	8,12
Dimethoate	16,22		3,5,6,9,11,13,17		10	8,12
2,4-D	16,22		3,5,6,9,11,14,17	2	4,10	8,12
Isoproturon	16,22	4	3,5,6,9,14,17		10	8,12
Linuron	16,22		3,5,12,14		4,9,10	8

The above Table 2 illustrates an important aspect of the data imbalance. Information about methods of analysis was registered in the certification procedure, but the original design was not intended for investigation of methods. As mentioned earlier, the certification data were unbalanced as the result of the complicated analyses of 22 pesticides. In this part of the analysis it was decided to leave out methods used by only one laboratory since keeping such data in the model would mean confounding the effect of laboratory with method.

4 Hypothesis

The purpose of the present study was to investigate whether differences between results from different methods of analysis could be detected.

5 Model and Method

A mixed model approach is applied. The structure from the certification study, described in equation (1), is expanded by a systematic effect of method. The method effect is placed as the top level in the nested structure so that laboratory, days and ampoules are nested within method. The response y for a selected pesticide is modelled by an expanded version of the linear model (1) where the systematic effect for method is added, assuming that $\sum_i m_i = 0$.

$$Y_{ijklm} = \mu + m_i + L(M)_{j(i)} + D(ML)_{k(ij)} + A(MLD)_{l(ijk)} + E_{ijklm}$$

The random effects are as described in Table 1. This model can be rephrased in a matrix structure where the response is modelled by

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are design matrices and $\mathbf{e}$ is the residual. β is a vector of method parameters and $\mathbf{u}$ is a matrix of variance components:

$$\boldsymbol{\beta} = \begin{bmatrix} m_A \\ m_B \\ m_C \\ m_D \\ m_E \\ m_F \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} \sigma^2_{L(M)} & & \\ & \sigma^2_{D(LM)} & \\ & & \sigma^2_{A(MLD)} \end{bmatrix}$$

Normality is assumed for the random parameters so that

$$\mathrm{E}\begin{bmatrix} \mathsf{u} \\ \mathsf{e} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \mathrm{Var}\begin{bmatrix} \mathsf{u} \\ \mathsf{e} \end{bmatrix} = \begin{bmatrix} \mathsf{G} & \mathsf{0} \\ \mathsf{0} & \mathsf{R} \end{bmatrix}$$

The variance of the response y is described as

$$\mathbf{V} \;=\; \mathbf{Z}\mathbf{G}\mathbf{Z}' \;+\; \mathbf{R}$$

The rank of $\mathbf{X}$ is p and the residuals are calculated as

$$\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

The expression for the REsidual Maximum Likelihood (REML) is

$$l_R(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{n-p}{2}\log|\mathbf{r}'\mathbf{V}^{-1}\mathbf{r}| - \frac{n-p}{2}\{1+\log[2\pi/(n-p)]\} \quad (2)$$

The objective function is minimized to obtain estimates of the $\mathbf{G}$ and $\mathbf{R}$ matrices using PROC MIXED in SAS®. Solving the mixed model equation*s

$$\begin{bmatrix} \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{Z}+\widehat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

yields the estimates of the fixed and random effects in the model.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}$$

$$\hat{\mathbf{u}} = \widehat{\mathbf{G}}\mathbf{Z}'\widehat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

The full model is reduced stepwise, removing explanatory factors one at a time starting with the lowest level in the nested structure (ampoules). After each reduction the difference in χ^2 statistics is examined to determine the significance of the omitted factor. All tests were performed at a 5 % significance level.

Table 3: Stepwise reduction using the procedure MIXED in SAS®

PROC MIXED - Likelihood Ratio Test
Full model (0) $Y_{ijklv} = \mu + m(i) + L(M)_{j(i)} + D(ML)_{k(ij)} + S(MLD)_{l(ijk)} + E_{v(ijkl)}$
Test procedure, χ^2-test:$\chi^2_{Full} - \chi^2_{Reduced1} \in \chi(1)_{0.95}$
Reduced model (1) $Y_{ijklv} = \mu + m(i) + L(M)_{j(i)} + D(ML)_{k(ij)} + E_{v(ijkl)}$
Test procedure, χ^2-test:$\chi^2_{reduced1} - \chi^2_{reduced2} \in \chi(1)_{0.95}$
Reduced model (2) $Y_{ijklv} = \mu + m(i) + L(M)_{j(i)} + E_{v(ijkl)}$

6 Results

The result of the analyses can be outlined as follows:

For the compounds included in this study the statistical analyses resulted in models as follows. For linuron the full model was reduced to the first stage (reduced model (1)) and the effect from method was found to be significant at $\alpha = 0.05$. For cyanazine, isoproturone, and dichlorprop the effect from method was not significant at $\alpha = 0.05$, and the model could not be reduced further.

It is well known among analytical chemists in this field that laboratory variation is present, however it was not expected that the day to day variance would be significant and furthermore larger than the interlaboratory variance. A closer look at some of the laboratories revealed that in some cases, the day to day variance $(\sum_k \bar{X}_{ijk..} - \bar{X}_{ij...})^2/(n_{ijk..})$ within the laboratory was larger than the interlaboratory variance $((\bar{X}_{ij...} - \bar{X}_{.....})^2/(n_{ij...}))$. Some examples are given here.

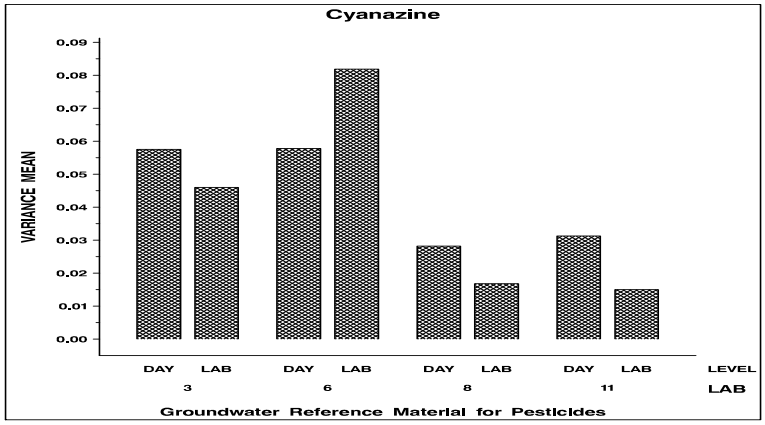

Figure 1: Cyanazine - selected laboratories: Variance between days within each laboratory and variance between laboratory and overall mean.

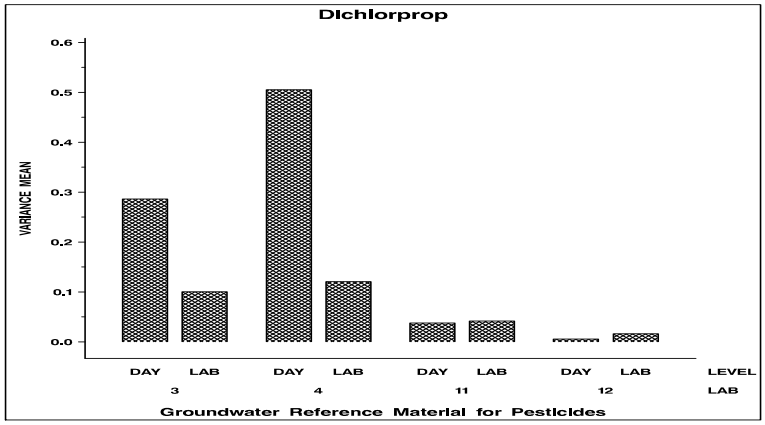

Figure 2: Dichlorprop: - selected laboratories: Variance between days within each laboratory and variance between laboratory and overall mean.

This clearly indicates that in many of the laboratories the analytical methods are not in statistical control. This finding was surprising because the laboratories were selected for their long experience in pesticide analyses and that the selected compounds are not new, but have been analysed for a number of years. The large day to day variance observed is likely to disguise possible method effects in the model. The estimation of covariance parameters for the effects in the reduced model (1) shows that variance estimates for $\sigma^2_{L(M)}$ are zero in three out of four cases when applying the estimation

$$\sigma^2_{L(M)} = \max\{0,\ \ k_L^{-1}(MS_{L(M)} - MS_{D(LM)})\}$$

k_L is the coefficient for $\sigma_{L(M)}$ in the expected mean square of interlaboratory variance.

For Dimethoate and 2,4-D the estimates of the G-matrices were not positive definite. A possible explanation for this could be, that too many parameters are specified in a system where data are too sparse for an estimation of all the parameters in the model. During analysis it was seen that leaving out the laboratory factor from the model (still keeping the day and ampoule factors) eliminated the problem of a not positive definite G-matrix. If the day-to-day variances is greater than the interlaboratory variance it will give rise to non positive interlaboratory variance estimates.

The possibility of obtaining non positive variance estimates for random effects in nested models has been discussed among others Leone et al. (1968), Thompson (1962) and McHugh and Mielke (1968).

Leone et al. (1968) carried out a Monte Carlo study of the probabilities of obtaining non positive variance estimates in various types of nested designs ranging from the fully balanced nested to the inverted staggered design. The data in this study represent intermediate cases between the fully balanced and the inverted staggered design. In a four stage nested design with $\sigma^2_a = 1$, $\sigma^2_b = 9$, $\sigma^2_c = 9$, the probability of obtaining non positive variance estimates in the balanced, staggered and inverted designs was about 12 % for normal distributed variables.

In other words, non positive variances may occur according to Leone et al. (1968) and also Thompson (1962) and McHugh and Mielke (1968). Thompson states that the reason may be 1) the non positive variance estimates may well be the result of incorrect

Table 4: Covariance parameter estimates calculated by REML

COMPOUND	Cov. Parm.	Estimate
CYANAZIN	LAB(METHOD) DAY(METHOD*LAB) Residual	0.096 0.206 0.063
LINURON	LAB(METHOD) DAY(METHOD*LAB) Residual	0.038 0.660 0.074
ISOPROTURON	LAB(METHOD) DAY(METHOD*LAB) Residual	0.004 0.392 0.052
DICHLORPROP	LAB(METHOD) DAY(METHOD*LAB) Residual	0.126 0.123 0.040

model assumptions and 2) statistical noise obscuring the underlying physical situation. McHugh and Mielke investigated the first point presented by Thompson by removing the assumption of statistical independence for the reason that sampling carried out in the real world takes place on a finite population without replacement. In the resulting model the a_is in the first stage are no longer independent. The same applies to the b_{ij}s in the second stage. If an infinite model is incorrectly applied non positive variance estimates may be the outcome.

In our case, however, statistical independence between laboratories in general cannot be doubted. One might ask whether laboratories using the same method of analysis could be statistically correlated. Data inspection by graphical methods does not support such a hypothesis. The explanation offered by Leone et al. (1968) seems to cover the present case. The most important knowledge gained from the present study, is that even experienced laboratories analysing well known compounds do not have their day to day variance in control.

7 Conclusion

For cyanazine, dichlorprop, dimethoate, 2,4-D, isoproturon and linuron the effect of applied analytical method was significant only for the last compound linuron when applying a mixed model procedure. For cyanazine, isoproturon and dichlorprop the method effect was not significant at α=0.05. In two cases dimethoate and 2,4-D the mixed procedure of SAS® gave rise to not positive definite G-matrices. It was possible to detect method dependence for dimethoate when leaving out an intermediate nested effect from the full model. A common feature of all of the data sets is a general tendency of having larger day to day than interlaboratory variances.

References

ISO Guide 30: Terms and definitions used in connection with reference materials, 1992.

ISO Guide 35: Certification of reference materials - general and statistical principles, 1989.

ISO-REMCO N282 Rev Final Draft (1993): Quality system for the operation of a reference material producer, 1993.

F.C. Leone, L.S. Nelson, N.L. Johnson, and S. Eisenstat. Sampling distributions of variance components ii. empirical studies of unbalanced nested designs. *Technometrics*, 10 (4):719–737, 1968.

R.B. McHugh and P.W. Jr. Mielke. Negative variance estimates and statistical dependence in nested sampling. *J. Amer. Statist. Assoc.*, 63:1000–1003, 1968.

W.A. Jr. Thompson. The problem of negative estimates of variance components. *Annals of Mathematical Statistics*, 33:273–89, 1962.

Authors' addresses:

Karina Schramm-Nielsen
IMM,
Department of Mathematical Modelling,
Technical University of Denmark
DK-2800 Lyngby, Denmark

Tel. (+45) 45 25 33 76
Fax (+45) 45 88 26 73
E-mail: ksn@imm.dtu.dk

Jill Merry
VKI, Water Quality Institute
Agern Allé 11
DK-2970 Hoersholm, Denmark

Tel. (+45) 45 16 93 63

Bente A. Nyeland
NERI, National Environmental Research Institute
P.O. Box 358
Frederiksborgvej 399
DK-4000 Roskilde, Denmark

Tel. (+45) 46 30 13 57

Niels H. Spliid
Danish Institute of Agricultural
Sciences Research Centre
Flakkebjerg, Denmark

Henrik Spliid
IMM,
Department of Mathematical Modelling,
Technical University of Denmark
DK-2800 Lyngby, Denmark

Tel. (+45) 45 25 33 62