Critical values for 33 discordancy test variants for outliers in normal samples up to sizes 1000, and applications in quality control in Earth Sciences

Surendra P. Verma[1,*], Alfredo Quiroz-Ruiz[1], and Lorena Díaz-González[2]

[1] Centro de Investigación en Energía , Universidad Nacional Autónoma de México, Priv. Xochicalco s/n, Col. Centro, Apartado Postal 34, Temixco 62580, Mexico.
[2] Posgrado en Ingeniería (Energía), sede Centro de Investigación en Energía, Universidad Nacional Autónoma de México, Priv. Xochicalco s/n, Col. Centro, Apartado Postal 34, Temixco 62580, Mexico.
** spv@cie.unam.mx*

ABSTRACT

In two earlier papers (Verma and Quiroz-Ruiz, 2006, Rev. Mex. Cienc. Geol., 23, 133-161, 302-319) precise critical values for normal univariate samples of sizes n up to 100 have been reported. However, for greater n, critical values are available only for a few tests: N1 for n up to 147, N4k2 for n up to 149, N6, N14 and N15 (for the latter three tests, critical values were reported for only n=200, 500, and 1000). This clearly demonstrates the need for proposing new critical values for n>100 through an adequate statistical methodology. Therefore, modifications of our earlier simulation procedure as well as new, precise, and accurate critical values or percentage points (with four to eight decimal places; average standard error of the mean ~0.00000003–0.0039) of 15 discordancy tests with 33 test variants, and each with seven significance levels α = 0.30, 0.20, 0.10, 0.05, 0.02, 0.01, and 0.005, for normal samples of sizes n up to 1000, viz., n_{min} (1)100(5)200(10)500(20)1000, are reported. For the first time in the literature, the standard error of the mean is also reported explicitly and individually for each critical value. Similarly, a new methodology involving artificial neural network (ANN) was used, for the first time in published literature, to obtain interpolation equations for all 33 discordancy test variants and for each of the seven significance levels. Each equation was fitted using 76 simulated data for n from 100 to 1000 for a given test and significance level. Extremely small sums of squared residuals (~5.5×10^{-8} – 8.4×10^{-5}; generally <10^{-5}) in the ANN equations fitted for n=100 to 1,000 were obtained. As a result, the applicability of these discordancy tests is now extended up to 1000 observations of a particular parameter in a statistical sample. The new most precise and accurate critical values will result in more reliable applications of these discordancy tests than have been possible so far in various scientific and engineering fields, particularly for quality control in Earth Sciences. The multiple-test method with new critical values was shown to perform better than both the box-and-whisker plot and the "two standard deviation" methods used by some researchers, and is therefore the recommended procedure for handling experimental data.

Key words: outlier methods, normal sample, two standard deviation method, 2s method, reference materials, Monte Carlo simulation, critical values, Dixon tests, skewness, kurtosis, artificial neural network, ANN, statistics, petroleum hydrocarbon, Nd isotopes, BCR-1.

RESUMEN

En dos trabajos anteriores (Verma and Quiroz-Ruiz, 2006, Rev. Mex. Cienc. Geol., 23, 133-161, 302-319) se han reportado valores críticos precisos para pruebas de discordancia en muestras normales univariadas n hasta 100. Sin embargo, para n >100, se dispone solamente de valores críticos para las pruebas: N1 para n hasta 147, N4k2 para n hasta 149, N6, N14 y N15 (para las últimas tres pruebas, valores críticos han sido reportados solamente para n=200, 500 y 1000). Esto demuestra claramente la necesidad de proponer nuevos valores críticos para n>100 mediante una metodología estadística apropiada. Por lo tanto, se reportan las modificaciones del procedimiento de la simulación así como valores críticos o puntos porcentuales nuevos y más precisos y exactos (con cuatro hasta ocho puntos decimales; el error estándar de la media ~ 0.00000003 – 0.0039) para 15 pruebas de discordancia con 33 variantes, y cada una con siete niveles de significancia α = 0.30, 0.20, 0.10, 0.05, 0.02, 0.01 y 0.005, para muestras normales con tamaño n hasta 1000, viz., n_{min} (1)100(5)200(10)500(20)1000. Por primera vez en la literatura, se reporta el error estándar de la media explícitamente y en forma individual para cada valor crítico. De igual manera, una nueva metodología que consiste en la aplicación de redes neuronales artificiales (ANN, por sus siglas en inglés) fue usada, por primera vez en la literatura publicada, para obtener ecuaciones de interpolación para las 33 variantes de las pruebas de discordancia y para cada uno de los siete niveles de significancia. Cada ecuación fue ajustada con los 76 datos de las simulaciones para n desde 100 hasta 1,000 correspondientes a cada prueba y cada nivel de significancia. Sumas de cuadrados de los residuales extremadamente pequeñas (~5.5×10^{-8} – 8.4×10^{-5}; generalmente <10^{-5}) fueron obtenidas en el ajuste de las ecuaciones por ANN para n =100 a 1,000. Como consecuencia, la aplicabilidad de las pruebas de discordancia ha sido extendida hasta 1,000 observaciones de un determinado parámetro en una muestra estadística. Los valores críticos nuevos y mucho más precisos y exactos resultarán en aplicaciones más confiables de las pruebas de discordancia que han sido posibles hasta ahora en una variedad de campos de las ciencias e ingenierías, particularmente para el control de calidad en Ciencias de la Tierra. El método de pruebas múltiples con nuevos valores críticos proporcionó mejores resultados que los métodos de la gráfica de "box y whisker" y de "dos desviaciones estándar" usados por algunos investigadores y, por lo tanto, el presente método estadístico es el más recomendado para el manejo de datos experimentales.

Palabras clave: métodos de valores desviados, muestra normal, prueba de dos desviaciones estándar, 2s, materiales de referencia, simulación Monte Carlo, valores críticos, pruebas de Dixon, sesgo, curtosis, redes neuronales artificiales, RNA, estadística, hidrocarburos de petróleo, isótopos de Nd, BCR-1.

INTRODUCTION

Two recent papers (Verma and Quiroz-Ruiz, 2006a, 2006b) have reported a highly precise and accurate Monte Carlo type simulation procedure for N(0,1) random normal variates and presented new, precise, and accurate critical values for seven significance levels α = 0.30, 0.20, 0.10, 0.05, 0.02, 0.01, and 0.005, and for sample sizes n up to 100 for 15 discordancy tests with 33 variants. Table 1 summarizes these tests. However, for greater n, only a few critical values are available in the literature (Barnett and Lewis, 1994; Verma, 2005). These values are for tests: N1 (n up to 147); N4k2 (n up to 149); N6, N14 and N15 (for the latter three tests, critical values with only two decimal places were reported for only n =200, 500, and 1000).

Reference materials (RMs) are routinely used for quality control in Earth Sciences (*e.g*., Verma, 1997, 1998, 2005; Velasco-Tapia *et al.*, 2001; M.P. Verma, 2004; Lozano and Bernal, 2005; Guevara *et al.*, 2005; Sang *et al.*, 2006; Santoyo *et al.*, 2006; Papadakis *et al.*, 2007). In other fields of science and engineering also, quality control through RMs has become mandatory, for example, in biology and medicine (Okamoto *et al.*, 1996; Dybczyński *et al.*, 1998; Patriarca *et al.*, 2005); environmental sciences (Gill *et al.*, 2004; Graybeal *et al.*, 2004; Farre *et al.*, 2006); and food research (In´t Veld, 1998; Langton *et al.*, 2002; Gabrovská *et al.*, 2006).

When a large number of laboratories around the world participate in a cooperative study of a RM, the number of individual data (n) for a given chemical element in that RM can exceed 100. In these cases, at present the multiple-test method initially proposed by Verma (1997) and practiced by Verma (1998, 2005) and Verma and Quiroz-Ruiz (2006a, 2006b), among others, is not likely to be appropriately applicable due to the unavailability of precise critical values for n >100 for most discordancy tests (Table 1). This clearly demonstrates the need for proposing new critical values for n >100 through an adequate statistical methodology. Requirements of critical values for large n (>100) also exist in an altogether different field of molecular and cellular proteomics (Xia *et al.*, 2006; Murray Hackett, written communication, June 2007).

For the present work, we have included most discordancy tests for normal univariate samples (15 tests with 33 test variants; see Table 1) for simulating new, precise, and accurate critical values for the same seven

Table 1. Fifteen discordancy tests with 33 test variants for univariate normal samples (modified after Barnett and Lewis, 1994; Verma, 1997, 2005; Verma and Quiroz-Ruiz, 2006a, 2006b).

Test code *		**Value(s) tested**	**Test statistic**	**Test significance**	**Applicability of test $n_{min} - n_{max}$**		
					Literature pre-2006 (less precise values) **	Literature 2006 (more precise values) ***	This work, 2008 (most precise values) ****
N1	Upper	$x_{(n)}$	$TN1_{(u)} = (x_{(n)}-\bar{x})/s$	Greater	3 – 100	3 – 100	3 – 1000
	Lower	$x_{(1)}$	$TN1_{(l)} = (\bar{x}-x_{(1)})/s$	Greater	3 – 100	3 – 100	3 – 1000
N2 (two-sided)	Extreme	$x_{(n)}$ or $x_{(1)}$	$TN2 = Max$: $\{(x_{(n)}-\bar{x})/s, (\bar{x}-x_{(1)})/s\}$	Greater	3 – 100	3 – 100	3 – 1000
N3	k=2 Upper	$x_{(n)}$, $x_{(n-1)}$	$TN3_{(2u)} = (x_{(n)}+x_{(n-1)}-2\bar{x})/s$	Greater	5 – 100	5 – 100	5 – 1000
	k=3 Upper	$x_{(n)}$, $x_{(n-1)}$, $x_{(n-2)}$	$TN3_{(3u)}=(x_{(n)}+x_{(n-1)}+x_{(n-2)}-3\bar{x})/s$	Greater	7 – 100	7 – 100	7 – 1000
	k=4 Upper	$x_{(n)}$, $x_{(n-1)}$, $x_{(n-2)}$, $x_{(n-3)}$	$TN3_{(4u)}=(x_{(n)}+x_{(n-1)}+x_{(n-2)}+x_{(n-3)}-4\bar{x})/s$	Greater	9 – 100	9 – 100	9 – 1000
	k=2 Lower	$x_{(1)}$, $x_{(2)}$	$TN3_{(2l)}=(2\bar{x}-x_{(1)}-x_{(2)})/s$	Greater	5 – 100	5 – 100	5 – 1000
	k=3 Lower	$x_{(1)}$, $x_{(2)}$, $x_{(3)}$	$TN3_{(3l)}=(3\bar{x}-x_{(1)}-x_{(2)}-x_{(3)})/s$	Greater	7 – 100	7 – 100	7 – 1000
	k=4 Lower	$x_{(1)}$, $x_{(2)}$, $x_{(3)}$, $x_{(4)}$	$TN3_{(4l)}=(4\bar{x}-x_{(1)}-x_{(2)}-x_{(3)}-x_{(4)})/s$	Greater	9 – 100	9 – 100	9 – 1000
N4	k=1 Upper	$x_{(n)}$	$TN4_{(1u)}=S^2_{(n)}/S^2$	Smaller	3 – 100	3 – 100	3 – 1000
	k=2 Upper	$x_{(n)}$, $x_{(n-1)}$	$TN4_{(2u)}=S^2_{(n),(n-1)}/S^2$	Smaller	4 – 100	4 – 100	4 – 1000
	k=3 Upper	$x_{(n)}$, $x_{(n-1)}$, $x_{(n-2)}$	$TN4_{(3u)}=S^2_{(n),(n-1),(n-2)}/S^2$	Smaller	6 – 100	6 – 100	6 – 1000
	k=4 Upper	$x_{(n)}$, $x_{(n-1)}$, $x_{(n-2)}$, $x_{(n-3)}$	$TN4_{(4u)}=S^2_{(n),(n-1),(n-2),(n-3)}/S^2$	Smaller	8 – 100	8 – 100	8 – 1000
	k=1 Lower	$x_{(1)}$	$TN4_{(1l)}=S^2_{(1)}/S^2$	Smaller	3 – 100	3 – 100	3 – 1000
	k=2 Lower	$x_{(1)}$, $x_{(2)}$	$TN4_{(2l)}=S^2_{(1),(2)}/S^2$	Smaller	4 – 100	4 – 100	4 – 1000
	k=3 Lower	$x_{(1)}$, $x_{(2)}$, $x_{(3)}$	$TN4_{(3l)}=S^2_{(1),(2),(3)}/S^2$	Smaller	6 – 100	6 – 100	6 – 1000
	k=4 Lower	$x_{(1)}$, $x_{(2)}$, $x_{(3)}$, $x_{(4)}$	$TN4_{(4l)}=S^2_{(1),(2),(3),(4)}/S^2$	Smaller	8 – 100	8 – 100	8 – 1000
N5	k=2 Upper–Lower	$x_{(n)}$, $x_{(1)}$	$TN5_{(ul)}=S^2_{(n),(1)}/S^2$	Smaller	4 – 100	4 – 100	4 – 1000
N6	k=2 Upper–Lower	$x_{(n)}$, $x_{(1)}$	$TN6_{(ul)}=(x_{(n)}-(x_{(1)})/s$	Greater	3 – 100	3 – 100	3 – 1000
N7 (r_{10})	Upper	$x_{(n)}$	$TN7=(x_{(n)}-x_{(n-1)})/ (x_{(n)}-x_{(1)})$	Greater	3 – 30	3 – 100	3 – 1000
N8 (two-sided)	Extreme	$x_{(n)}$, or $x_{(1)}$	TN8=Max: $\{(x_{(n)}-x_{(n-1)})/ (x_{(n)}-x_{(1)})\}$ $\{(x_{(2)}-x_{(1)})/ (x_{(n)}-x_{(1)})\}$	Greater	4 – 100	4 – 100	4 – 1000

Table 1 (continued). Fifteen discordancy tests with 33 test variants for univariate normal samples (modified after Barnett and Lewis, 1994; Verma, 1997, 2005; Verma and Quiroz-Ruiz, 2006a, 2006b).

Test code *		Value(s) tested	Test statistic	Test significance	Applicability of test $n_{min} - n_{max}$		
					Literature pre-2006 (less precise values) **	Literature 2006 (more precise values) ***	This work, 2008 (most precise values) ****
N9 (r_{11})	Upper	$x_{(n)}$	$TN9_{(u)} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(2)})$	Greater	4 – 30	4 – 100	4 – 1000
	Lower	$x_{(1)}$	$TN9_{(l)} = (x_{(2)} - x_{(1)}) / (x_{(n-1)} - x_{(1)})$	Greater	4 – 30	4 – 100	4 – 1000
N10 (r_{12})	Upper	$x_{(n)}$	$TN10_{(u)} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(3)})$	Greater	5 – 30	5 – 100	5 – 1000
	Lower	$x_{(1)}$	$TN10_{(l)} = (x_{(2)} - x_{(1)}) / (x_{(n-2)} - x_{(1)})$	Greater	5 – 30	5 – 100	5 – 1000
N11 (r_{20})	Upper pair	$x_{(n)}, x_{(n-1)}$	$TN11_{up} = (x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(1)})$	Greater	4 – 30	4 – 100	4 – 1000
	Lower pair	$x_{(1)}, x_{(2)}$	$TN11_{lp} = (x_{(3)} - x_{(1)}) / (x_{(n)} - x_{(1)})$	Greater	4 – 30	4 – 100	4 – 1000
N12 (r_{21})	Upper pair	$x_{(n)}, x_{(n-1)}$	$TN12_{up} = (x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(2)})$	Greater	5 – 30	5 – 100	5 – 1000
	Lower pair	$x_{(1)}, x_{(2)}$	$TN12_{lp} = (x_{(3)} - x_{(1)}) / (x_{(n-1)} - x_{(1)})$	Greater	5 – 30	5 – 100	5 – 1000
N13 (r_{22})	Upper pair	$x_{(n)}, x_{(n-1)}$	$TN13_{up} = (x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(3)})$	Greater	6 – 30	6 – 100	6 – 1000
	Lower pair	$x_{(1)}, x_{(2)}$	$TN13_{lp} = (x_{(3)} - x_{(1)}) / (x_{(n-2)} - x_{(1)})$	Greater	6 – 30	6 – 100	6 – 1000
N14	Extreme	$x_{(n)}$, or $x_{(1)}$	$TN14 = \left[\frac{n^{1/2} \{\sum_{i=1}^{n} (x_i - \bar{x})^3\}}{\{\sum_{i=1}^{n} (x_i - \bar{x})^2\}^{3/2}} \right]$	Greater	5 – 100	5 – 100	5 – 1000
N15	Extreme	$x_{(n)}$, or $x_{(1)}$	$TN15 = \left[\frac{n\{\sum_{i=1}^{n} (x_i - \bar{x})^4\}}{\{\sum_{i=1}^{n} (x_i - \bar{x})^2\}^2} \right]$	Greater	5 – 100	5 – 100	5 – 1000

* Test code (N series) is from Barnett and Lewis (1994), whereas test code (*r* series) is for Dixon tests (see Dixon, 1951); tests N14 and N15 are respectively the skewness and kurtosis tests. The symbols for test statistics $TN1_{(u)}$, $TN1_{(l)}$,$TN2$, etc. have been proposed by Verma (2005) and used by Verma and Quiroz-Ruiz (2006a, 2006b). The subscripts (*u*), (*l*), (2*u*), and (2*l*) are, respectively, upper (the highest), lower (the lowest), upper pair, and lower pair observations. The test statistics are self explanatory except the statistics of the type "*reduced* sum of squares" / "*total* sum of squares" for example, $S^2_{(n)}/S^2$ for test N4–k=1, proposed by Grubbs (1950, 1969), which need some explanation. For an ordered array $x_{(1)}, x_{(2)}, x_{(3)}, \ldots x_{(n-2)}, x_{(n-1)}, x_{(n)}$, the S^2 term is calculated using all data $S^2 = \Sigma^{n}_{i=1}(x_{(i)} - \bar{x})^2$, where $\bar{x}$ is the arithmetic mean ($\bar{x} = \Sigma_{i=1}^{n} x_{(i)}/n$), whereas $S^2_{(n)}$ is computed from the (n-1) remaining data $x_{(1)}, x_{(2)}, x_{(3)}, \ldots$, $x_{(n-2)}, x_{(n-1)}$, after eliminating the highest datum to be tested $x^2_{(n)}$ (see the subscript (n) in the term $S_{(n)}$ as follows: $S^2_{(n)} = \Sigma^{n-1}_{i=1}(x_{(i)} - \bar{x}_n)^2$ where $\bar{x}_n = \Sigma_{i=1}^{n-1} x_{(i)}/(n\text{-}1)$. The other statistics of the type $S^2_{(n)}/S^2$, such as $S^2_{(1)}/S^2$ or $S^2_{(n),\,(n-1)}/S^2$ are calculated in a similar manner. For more details, see Verma (2005).

** For literature values see books by Barnett and Lewis (1994) and Verma (2005).

*** Verma and Quiroz-Ruiz (2006a, 2006b) increased n_{max} to 100 by simulating more precise and accurate critical values for all discordancy tests.

**** Finally, note that, in the present work, n_{max} has been increased to 1000 for all discordancy tests (see Tables A1-A40 of the electronic supplement), and when critical values were already available for this $n_{min} - n_{max}$ range, the new values are shown to be more precise and accurate than even Verma and Quiroz-Ruiz (2006a, b) (see Fig. 1 for comparison of standard errors of the simulated critical values). Because critical values were simulated for n_{min}(1)100(5)200(10)500(20)1000 (see Tables A1-A40 and A41 of the electronic supplement), interpolation equations using 76 newly generated critical values for n from 100 to 1,000 (a total of 140 equations for all discordancy tests and $\alpha = 0.30, 0.20, 0.10, 0.05, 0.02, 0.01$, and 0.005) were proposed for correctly obtaining the "missing" values for n between 100 and 1000 (see Table 2 and Tables A42-A60 of the electronic supplement). For more information on these tests and their applications, see references cited in Verma and Quiroz-Ruiz (2006b).

significance levels (α = 0.30 to 0.005) and for n up to 1000, viz., n_{min} (1)100(5)200(10)500(20)1000 (where n_{min} is the minimum number of data that could be tested by a given statistical test; see Table 1), using a simulation procedure slightly modified after Verma and Quiroz-Ruiz (2006a, 2006b). Further, a novel approach is followed, for the first time in the literature, for presenting these new critical values along with the respective standard errors and for interpolating the simulated critical values using artificial neural network (ANN). These results are useful in all fields of science and engineering, especially in quality control in Earth Sciences. We present a few examples of the application of all normal univariate tests (Table 1) for which we have reported new, most precise critical values in this paper.

DISCORDANCY TESTS

We will not repeat the explanation of discordancy tests; the reader is referred to Barnett and Lewis (1994), Verma (2005), or the recent papers by Verma and Quiroz-Ruiz (2006a, 2006b). The 15 tests with their 33 variants for which critical values were simulated are listed in Table 1.

SIMULATION PROCEDURE FOR MOST PRECISE AND ACCURATE CRITICAL VALUES

Our highly precise and accurate Monte Carlo type simulation procedure has already been described in detail (Verma and Quiroz-Ruiz, 2006a, 2006b) and, therefore, will not be repeated here. However, some required changes will be mentioned.

In our present work, the simulations were of sizes 500,000 for tests N3-N5 and N7-N13; 1,000,000 for N14; and 2,000,000 for N1, N2, N6, and N15. They were repeated ten times (each using a different set of 500,000,000 to 2,000,000,000 random normal variates). Different simulation sizes (500,000 to 2,000,000) were appropriate to optimize the simulation time required for the use of personal computers and to obtain, at the same time, "acceptable" simulation errors for all tests. For tests N2, N5-N8, N14 and N15, the final mean critical value or percentage point ($\bar{x}$) and its standard error ($se_{\bar{x}}$) for each n and α were estimated from ten repetitions. However, for tests such as N1 (Table 1) two independent test statistics (one for an upper and the other for a lower outlier) were simulated and thus 20 independent results could be obtained from the same simulation scheme as reported earlier (Verma and Quiroz-Ruiz, 2006a, 2006b). Besides test N1, because of the existence of the upper and lower versions of the statistic (Table 1), 20 results of critical values and their error estimates were also obtained for tests N3–k=2,3,4, N4–k=1,2,3,4, and N9-N13. For all these tests, therefore, $\bar{x}$ and $se_{\bar{x}}$ calculations were based on 20 independent results.

RESULTS OF NEW CRITICAL VALUES

Both $se_{\bar{x}}$ and $\bar{x}$ data for 33 discordancy test variants (Table 1), for n from n_{min} (3, 4, 5, 6, 7, 8, or 9, depending on the type of statistic to be calculated) up to 1000, viz., n_{min} (1)100(5)200(10)500(20)1000, and α = 0.30, 0.20, 0.10, 0.05, 0.02, 0.01, and 0.005 (corresponding to confidence level of 70% to 99.5%, or equivalently significance level of 30% to 0.5%), are summarized in Tables A1-A40 (40 tables in the electronic supplement; 20 odd-numbered tables for $se_{\bar{x}}$ and 20 even-numbered tables for $\bar{x}$). Thus, our data presentation approach is novel because, for the first time in the literature, the precision estimates are explicitly tabulated for each critical value. For example, in Table A1 the rounded $se_{\bar{x}}$ values are presented individually for each n and α, whereas in Table A2 the rounded $\bar{x}$ values are similarly listed for test N1; the rounding procedure follows the guidelines suggested by Verma (2005). Similarly, $se_{\bar{x}}$ and $\bar{x}$ values are presented consecutively for the remaining tests N2 to N15 in Tables A3-A40.

For all cases, our present values are more reliable (error is given by a small number on the third up to the eighth decimal place) than the earlier literature values (compiled by Barnett and Lewis, 1994; Verma, 2005), including those reported by Verma and Quiroz-Ruiz (2006a, 2006b). In fact, the errors of these literature critical values, except those by Verma and Quiroz-Ruiz (2006a, 2006b), are not precisely known. A synthesis of standard errors of the mean for all tests is presented in Table A41 of the electronic supplement. The errors of the present critical values for n up to 1,000 (Table A41) range as follows: ~0.00000009–0.0007 for test N1 (see also Table A1); ~0.00000003–0.0009 for test N2 (Table A3), ~0.00005–0.0019 for N3–k=2 (Table A5); ~0.00009–0.0020 for N3–k=3 (Table A7); ~0.00010–0.0021 for N3–k=4 (Table A9); ~0.00000023–0.00040 for N4–k=1 (Table A11); ~0.00000007–0.00025 for N4–k=2 (Table A13); ~0.0000017–0.00021 for N4–k=3 (Table A15); ~0.0000021–0.00018 for N4–k=4 (Table A17); ~0.00000005–0.00035 for N5–k=2 (Table A19); ~0.00000005–0.0012 for N6–k=2 (Table A21); ~0.000016–0.0005 for N7 (Table A23); ~0.000015–0.0006 for N8 (Table A25); ~0.000015–0.00028 for N9 (Table A27); ~0.000016–0.00032 for N10 (Table A29); ~0.000015–0.00028 for N11–k=2 (Table A31); ~0.000008–0.00025 for N12–k=2 (Table A33); ~0.000011–0.00024 for N13–k=2 (Table A35); ~0.000023–0.0012 for N14 (Table A37); and ~0.000015–0.0039 for N15 (Table A39).

The much greater precision (and accuracy) of critical values simulated by Verma and Quiroz-Ruiz (2006a, 2006b) as compared to the literature values was already documented. Here, we compare the mean values of the standard errors for n up to 100 for all tests obtained in the present work with those obtained by Verma and Quiroz-Ruiz (2006a, 2006b) and show that the most precise critical values than ever attempted in the literature are now being reported (Table A41; see also Figure 1). And this improvement is due

to the fact that much larger simulation sizes of 500,000 to 2,000,000 are used in the present work, which have resulted in smaller standard errors than was earlier possible from sizes of 100,000 to 500,000. Further, when in the present work the sample sizes were exactly the same as those in Verma and Quiroz-Ruiz (2006b), for example, 500,000 for tests N3–k=2, 3, and 4 (see footnote of Table A41 for a correction and explanation), the errors were exactly the same (see diamond symbols that plot right at the diagonal line in Figure 1). For these cases, not only the errors were exactly the same, but also the critical values were identical in both simulations (this work and Verma and Quiroz-Ruiz, 2006b). This is an interesting observation and testifies the high reproducibility of our simulation procedure because the earlier simulations (Verma and Quiroz-Ruiz, 2006b) were programmed in C whereas the present simulations were programmed in a different language (Java) and were run on a different and faster personal computer equipped with a different processor than our earlier work.

As our earlier tables for sample sizes up to 100 (Verma and Quiroz-Ruiz, 2006a, 2006b), these new critical value data, along with their individual uncertainty estimates, are available in other formats such as *txt*, *Excel*, or *Statistica*, on request from the authors (S.P. Verma spv@cie.unam.mx, A. Quiroz-Ruiz aqr@cie.unam.mx, or L. Díaz-González ldg@cie.unam.mx). Similarly, the interpolation equations (see below) can also be obtained in a *doc* file with plain text format.

RESULTS OF INTERPOLATIONS OF CRITICAL VALUES USING ARTIFICIAL NEURAL NETWORK (ANN)

A new methodology was developed that involved the use of ANN for obtaining the best-fitted interpolation equations. This was actually required because, for $100 \leq n \leq 1000$, critical values were not simulated for all n (see Table 1 for tests and Tables A1-A40 for information on simulated n). No attempt was made in the present work to fit equations to critical values for $n < 100$ mainly because precise and accurate critical values for all $n < 100$ have already been simulated (Verma and Quiroz-Ruiz, 2006a, 2006b; this work). Therefore, interpolation equations were actually not required for small n. Prior to our work, different kinds of interpolation or fitting (Bugner and Rutledge, 1990; Rorabacher, 1991; Verma *et al.*, 1998) to low precision critical values available in the literature (see Barnett and Lewis, 1994; Verma, 2005; Verma and Quiroz-Ruiz, 2006a, 2006b) were used for this purpose.

This is the first time in published literature that ANN was used for fitting highly sophisticated equations to the most precise and accurate simulated critical value data for n between 100 and 1000, with extremely small sums of squares of residuals and thence for predicting interpolated critical values with the smallest error. Details on the ANN can be found in Hassoun (1995) or Haykin (1999).

The fitted equations for test N1 using all critical values

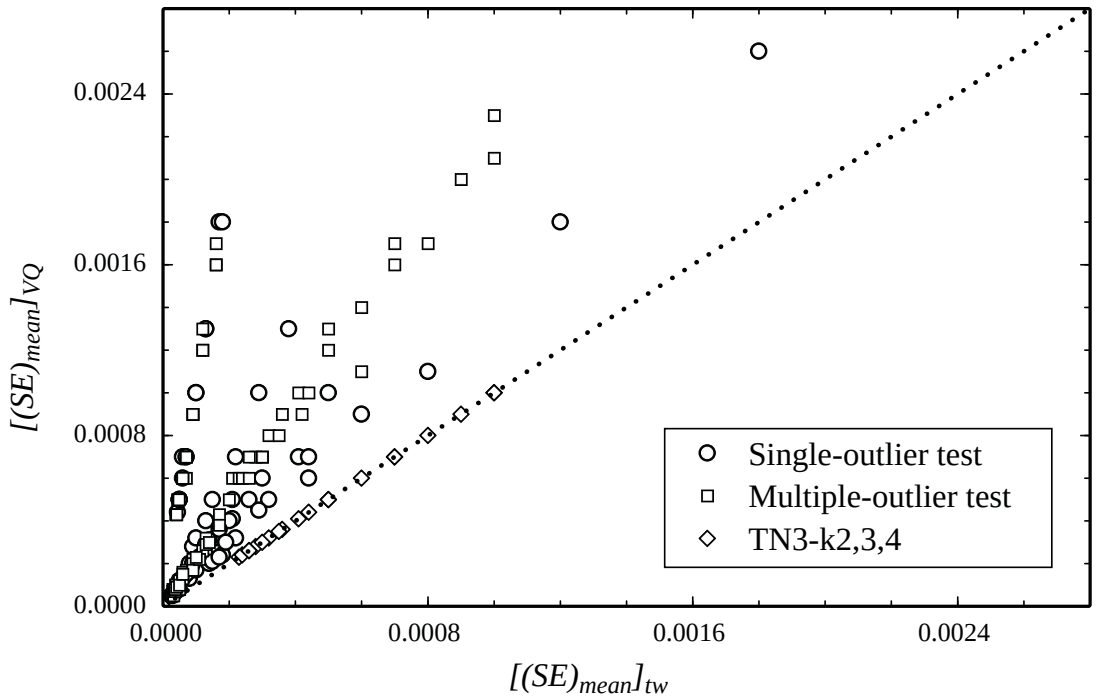

Figure 1. Comparison of new average values of standard errors of mean critical values obtained in the present work for all tests (15 tests with 33 variants) with those recently reported by Verma and Quiroz-Ruiz (2006a, 2006b) for $\alpha = 0.30$ to 0.005. The diagonal line represents those cases for which the present errors are the same as those obtained by Verma and Quiroz-Ruiz (2006a, 2006b); see text for more details.

listed for n from 100 to 1000 (Table A2) for each α (from 0.30 to 0.005) are presented in Table 2. The values of the sum of squared residuals of each equation ($\Sigma(SIM–ANN)^2$) for the 76 simulated critical values, corresponding to n = 100(5)200(10)500(20)1000 and α = 0.30, 0.20, 0.10, 0.05, 0.02, 0.01, and 0.005, used for equation fitting are also included in Table 2. For the remaining tests (N2 to N15), the equations are summarized in Tables A42-A60 of the electronic supplement.

The fitting quality parameter $\Sigma(SIM–ANN)^2$ for the interpolation equations for n=100(5)200(10)500(20)1000 (Tables 2 and A42-A60) range as follows: ~2.4×10^{-7} – 4.1×10^{-6} for test N1 (see Table 2); ~4.1×10^{-7} – 7.8×10^{-6} for test N2 (Table A42), ~1.8×10^{-6} – 3.0×10^{-5} for N3–k=2 (Table

Table 2. Interpolation equations fitted from ANN to 76 simulated critical values of Test N1 (for n between 100 and 1000; Table A2), used for computing interpolated precise critical values for 100 > n > 1000 (see Table A2 for those n for which critical values were simulated and for which interpolated critical values were required).

CL / SL / α	ANN Model	$\Sigma(SIM - ANN)^2$	Interpolation equation
70% / 30% / 0.30	4-1	2.9393×10^{-7}	$[CV_{TN1}^{0.30}]_{ANN} = (-0.001243\tanh(-0.012306n + 11.6994)) + (-0.7967\tanh(-0.00088756n - 0.20499)) + (1.8356\tanh(0.0035369n + 0.95374)) + (1.8669\tanh(0.01155n + 0.99749)) - 0.9585$
80% / 20% / 0.20	4-1	2.3957×10^{-7}	$[CV_{TN1}^{0.20}]_{ANN} = (-0.0011524\tanh(-0.011938n + 11.3938)) + (-0.6805\tanh(-0.0009135n - 0.11186)) + (2.0193\tanh(0.0034746n + 0.9971)) + (-2.0527\tanh(-0.011673n - 1.0372)) - 1.0934$
90% / 10% / 0.10	5-1	5.1654×10^{-7}	$[CV_{TN1}^{0.10}]_{ANN} = (0.0021069\tanh(0.01564n - 15.0842)) + (-0.001136\tanh(-0.012227n + 10.2114)) + (-0.34746\tanh(-0.0012735n + 0.38373)) + (-1.7327\tanh(-0.0033703n - 0.86195)) + (1.6954\tanh(0.011746n + 0.92192)) + 0.018995$
95% /5% / 0.05	5-1	9.8369×10^{-7}	$[CV_{TN1}^{0.05}]_{ANN} = (0.0028268\tanh(0.015256n - 14.9292)) + (0.0019383\tanh(0.011301n - 9.4696)) + (-0.3394\tanh(-0.0013487n + 0.38418)) + (1.9633\tanh(0.0035924n + 0.92815)) + (-1.9421\tanh(-0.012611n - 0.98245)) - 0.28933$
98% / 2% / 0.02	5-1	1.76×10^{-7}	$[CV_{TN1}^{0.02}]_{ANN} = (1.3706\tanh(0.012409n - 15.7713)) + (0.0049522\tanh(0.0096589n - 8.7164)) + (0.11713\tanh(0.0022176n - 1.4029)) + (1.4478\tanh(0.0026146n + 0.65921)) + (-1.4643\tanh(-0.010622n - 0.78004)) + 2.4662$
99% / 1% / 0.01	5-1	3.038×10^{-7}	$[CV_{TN1}^{0.01}]_{ANN} = (0.00017091\tanh(-0.033778n + 13.7692)) + (-0.63729\tanh(-0.00069273n - 0.17292)) + (3.5681\tanh(0.0097377n + 1.2801)) + (-3.3628\tanh(-0.0027313n - 1.3041)) + (3.5326\tanh(-4.6921n - 1.5225)) + 0.4037$
	5-1	4.059×10^{-6}	$[CV_{TN1}^{0.005}]_{ANN} = (0.0022291\tanh(0.015463n - 15.3358)) + (0.0012184\tanh(0.012468n - 10.7273)) + (-0.25014\tanh(-0.0014606n + 0.61062)) + (1.696\tanh(0.0034361n + 0.81002)) + (-1.6503\tanh(-0.013057n - 0.7881)) + 0.87585$

CL: Confidence level (%); SL: Significance level (%); α: Significance level; *ANN*: Artificial Neural Network; $\Sigma(SIM - ANN)^2$ = sum of squares of residuals for n = 100 to n = 1000. The first number in the ANN Model column refers to the number of neurons used at the input side of the ANN; only one output neuron was always used. Note the total number of terms in a given equation is the sum of the numbers of input and output neurons. The fitting quality parameter $\Sigma(SIM - ANN)^2$ is the total sum of squares of the difference between the simulated critical value (*SIM*) and that predicted by the (*ANN*) equation for the 76 simulated values corresponding to n = [100(5)200(10)500(20)1000] for a given CL (see Table A2 for the *SIM* values for n =100 to 1000 used for this fitting). Note that independent equations were fitted for each confidence level (70% to 99.5%) or significance level α (0.30 to 0.005). $[CV_{TN1}^{0.30}]_{ANN}$ in interpolation equations is the critical value (*CV*) for test *TN*1 and significance level α = 0.30 obtained by *ANN* methodology. The parameter n is the sample size of the critical value to be computed from the equation for a given significance level (α). $[CV_{TN1}^{0.05}]_{ANN}$ and $[CV_{TN1}^{0.01}]$ are the most commonly used critical values and the corresponding CL/SL/α are shown in ***italic bold*** face. Note also that Verma (1997) recommended the strict level of α = 0.01 be used in application of the multiple-test method. The other *CV* values in interpolation equations are similarly explained.

A43); ~4.1×10^{-6} – 7.6×10^{-5} for N3–k=3 (Table A44); ~7.7×10^{-6}–8.4×10^{-5} for N3–k=4 (Table A45); ~1.9×10^{-8} – 4.8×10^{-7} for N4–k=1 (Table A46); ~1.0×10^{-8} – 8.2×10^{-7} for N4–k=2 (Table A47); ~9.3×10^{-8} – 8.7×10^{-7} for N4–k=3 (Table A48); ~1.9×10^{-7} – 8.8×10^{-7} for N4–k=4 (Table A49); ~7.5×10^{-8} – 5.3×10^{-7} for N5–k=2 (Table A50); ~6.7×10^{-7} – 2.2×10^{-5} for N6–k=2 (Table A51); ~4.2×10^{-7} – 1.3×10^{-6} for N7 (Table A52); ~2.7×10^{-7} – 9.7×10^{-7} for N8 (Table A53); ~5.1×10^{-7} – 2.0×10^{-5} for N9 (Table A54); ~5.7×10^{-7} – 9.6×10^{-6} for N10 (Table A55); ~1.4×10^{-7} – 8.0×10^{-7} for N11–k=2 (Table A56); ~1.9×10^{-7} – 7.2×10^{-6} for N12–k=2 (Table A57); ~2.8×10^{-6} – 2.9×10^{-5} for N13–k=2 (Table A58); ~5.7×10^{-7} – 3.0×10^{-5} for N14 (Table A59); and ~1.8×10^{-7} – 2.8×10^{-5} for N15 (Table A60). Thus, the fitting quality parameter $\Sigma (SIM–ANN)^2$ for the interpolation equations was generally $<10^{-5}$ (and always $<10^{-4}$).

These equations can be used to compute precisely the interpolated critical values for all n between 100 and 1000, for which such values are not listed in Tables A2-A40 (see even-numbered tables). Thus, precise critical values can be made available for all n between n_{min} and 1000, viz., n_{min} (1)1000 (see Table 1 for more information on all tests and their n_{min} values). Figure 2 shows, as an example, the simulated critical values for test N1 (CV_{TN1}) for all values of α (0.30 to 0.005) as a function of n (from 100 to 1000). The respective interpolation equations are also plotted using dotted or dashed curves. Note these equations very closely match the simulated critical values to such an extent that the curves cannot be properly observed in Figure 2.

APPLICATIONS IN SCIENCE AND ENGINEERING

The tests (Table 1) after extending their applicability to samples of sizes up to 1,000, can be applied to all examples earlier summarized by Verma and Quiroz-Ruiz (2006a, 2006b). These include all the following fields (but are not limited to them): Agricultural and Soil Sciences; Aquatic Environmental Research; Astronomy; Biology; Biomedicine and Biotechnology; Chemistry; Electronics; Ecology; Geochronology; Geodesy; Geochemistry; Isotope Geology; Medical Science and Technology; Meteorology; Paleontology; Petroleum Hydrocarbons and Organic Compounds in Sediment Samples; Quality Assurance and Assessment Programs in Biology and Biomedicine, in Cement Industry, in Food Science and Technology, in Environmental and Pollution Research, in Nuclear Science, in Rock Chemistry, in Soil Science, and in Water Research; Structural Geology; Water Resources; and Zoology. Further, our new critical values for n up to 1,000 will be equally useful for applying these discordancy tests to identify outliers

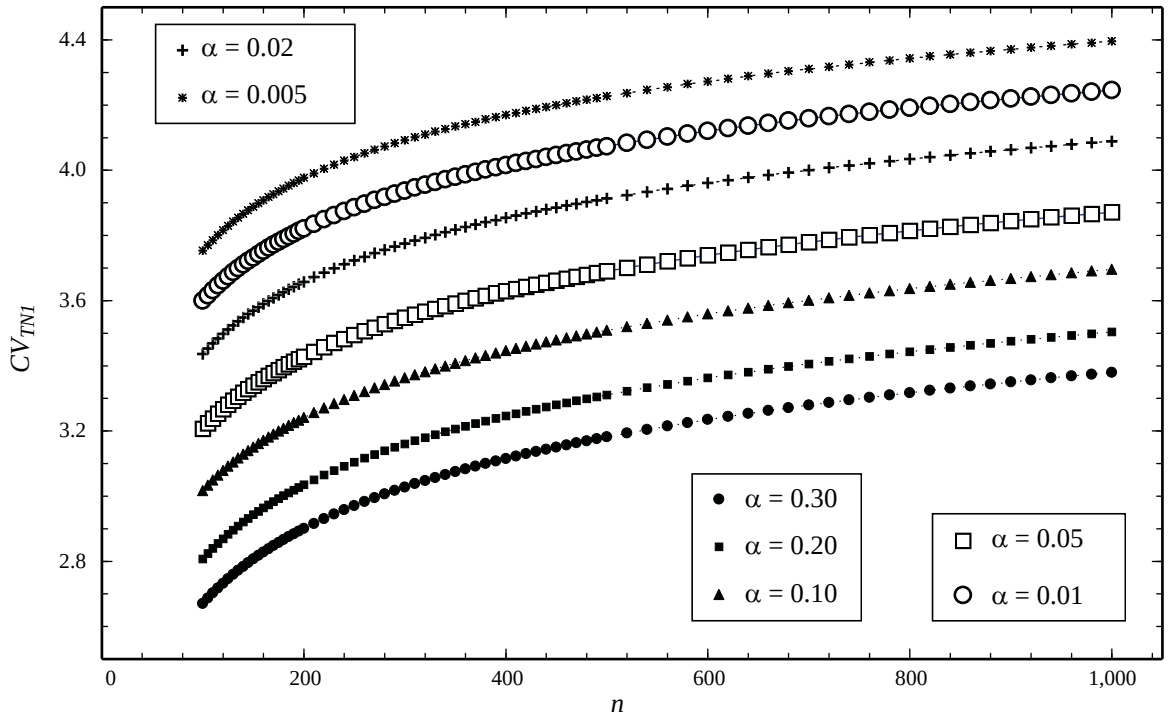

Figure 2. Interpolation curves (drawn from the corresponding ANN equations presented in Table 2) for test N1 with all significance levels α = 0.30, 0.20, 0.10, 0.05, 0.02, 0.01, and 0.005. Note that the fitted curves are "hidden" below the actually simulated critical values (see symbols used for each α shown as inset) for n = 100(5)200(10)500(20)1000; this is because both sets (fitted curves and the corresponding simulated values) are in very close agreement (see Table A2 for the actual critical values). Larger symbols were used for more frequently used significance levels (α = 0.05 and 0.01; the latter is the recommended significance level to be used to test experimental data for possible outliers).

in linear regressions, such as those recently employed by Verma *et al.* (2006).

APPLICATIONS IN QUALITY CONTROL IN EARTH SCIENCES

As stated earlier in the Introduction section, the most important requirement for simulating critical values for n >100 was for processing interlaboratory data for RMs. This is confirmed from the information synthesized in Table 3. Thus, the data for all chemical elements, without exception, in all RMs (Table 3) can now be processed for possible outliers and thence for correctly computing central tendency (or location) and dispersion (or scale) parameters (see Verma, 2005, for more details) using outlier-based statistical methods.

Specific examples

We now present examples or case histories to illustrate the use of all discordancy tests for which new critical values have been obtained in this work. For these applications, we chose the strict confidence level of 99% (*i.e.*, we used the 99% CL, or 1% SL, or 0.01 α column; see the respective critical values in Tables A2 to A40 – even-numbered tables in the electronic, supplementary data file).

Example 1. Comparison of multiple-test method with box-and-whisker plot method: Chlorinated pesticides and petroleum hydrocarbons in a sediment reference sample

Verma and Quiroz-Ruiz (2006b) used interlaboratory data for one sediment RM (IAEA-417; IAEA–International Atomic Energy Agency) to show that, for detecting outliers, the multiple-test method of Verma (1997) performed better than the box-and-whisker plot method used by Villeneuve *et al.* (2002). Here, we use, as the first example, a different sediment RM (IAEA-408) from an interlaboratory study by Villeneuve *et al.* (1999) to highlight the use of the multiple-test method and compare its performance with the box-and-whisker plot method used by the original authors. The individual data for nine selected compounds (six chlorinated pesticides and three petroleum hydrocarbons) were compiled in Table A61 (see electronic supplement) from the original report by Villeneuve *et al.* (1999). The multiple-test method (Table 1) consisted in applying all nine single-outlier (with 13 test variants) and seven multiple-outlier tests (with 20 test variants) at the strict confidence level of 99% to a given set of data although, if desired, a lesser number of tests or a less strict confidence level, such as 95%, could instead be used. The final concentration data along with the basic statistical information are summarized in Table 4. A greater number of discordant outliers were detected in the data for most compounds listed in Table A61, by the multiple-test method than the box-and-whisker plot method used by the original authors (Villeneuve *et al.*, 1999). Consequently,

Table 3. Improved applicability of the method of multiple tests (15 tests with 33 variants) for some RM databases (modified after Verma and Quiroz-Ruiz, 2006a).

Reference Material (RM)		Pre-2006 application (of Dixon tests) possible for # of		During 2006 application (of multiple test method) possible for # of		Present application (of multiple test method) possible for # of		Literature reference(s)
Code	Description	Major elements	Trace elements	Major elements	Trace elements	Major elements	Trace elements	
AGV-1	Andesite from U.S.G.S.	0	15	4	39	14 (all)	54 (all)	Gladney *et al.* (1992); Velasco-Tapia *et al.* (2001)
BCR-1	Basalt from U.S.G.S.	1	12	2	35	15 (all)	62 (all)	Gladney *et al.* (1990)
G-1	Granite from U.S.G.S.	0	36	5	55	15 (all)	56 (all)	Gladney *et al.* (1991)
G-2	Granite from U.S.G.S.	2	28	7	46	16 (all)	58 (all)	Gladney *et al.* (1992)
GSP-1	Granodiorite from U.S.G.S.	3	25	6	51	16 (all)	58 (all)	Gladney *et al.* (1992)
PM-S	Microgabbro (Scotland) from G.I.T.-I.W.G.	5	19	8	47	16 (all)	48 (all)	Govindaraju *et al.* (1994, 1995); Verma (1997)
W-1	Diabase from U.S.G.S.	0	21	3	47	14 (all)	57(all)	Gladney *et al.* (1991); Velasco-Tapia *et al.* (2001)
WS-E	Whin Sill dolerite (England) from G.I.T.-I.W.G.	5	15	8	44	16 (all)	46 (all)	Govindaraju *et al.* (1994, 1995); Verma *et al.* (1998)

U.S.G.S.: United States Geological Survey; G.I.T.-I.W.G.: Groupe International de Travail "Etalons analytiques des mineraux, minerais et roches or International Working Group "Analytical Standards of Minerals, Ores, and Rocks".

Table 4. Comparison of our multiple-test method (15 tests with 33 test variants) with the box and whisker plot and "two standard deviation" methods using concentrations of organochlorine pesticides and petroleum hydrocarbons in a sediment RM (IAEA-408; Villeneuve *et al.*, 1999; see Table A61 for individual data) and Sm, Nd and $^{143}Nd/^{144}Nd$ in a rock RM (BCR-1 from U.S.G.S.; Gladney *et al.*, 1990; see Tables A62-A64 for individual data).

Chemical variable	Initial statistics				Final statistics (this work)					Final statistics (literature)				
	n_{in}	$\bar{x}_{in}$	s_{in}	R_{in}	o_f	n_f	$\bar{x}_f$	s_f	R_f	o_l	n_l	$\bar{x}_l$	s_l	R_l
Sediment RM: IAEA-408														
HCB (ng/g)	32	250	1400	0.16 – 7910	10	22	0.43	0.15	0.18 – 0.71	7	25	0.45	0.19	0.16 – 0.85
pp'DDE (ng/g)	36	3	11	0.3 – 67.5	5	31	1.3	0.6	0.38 – 2.24	2	34	1.4	0.63	0.3 – 2.7
pp'DDD (ng/g)	31	3	10	0.101 – 56.7	5	26	1.0	0.6	0.20 – 2.34	1	30	1.1	0.74	0.56 – 1.7
PCB28 (ng/g)	16	1.1	0.9	0.11 – 3.49	4	12	0.64	0.32	0.11 – 1.19	2	14	0.74	0.4	0.11 – 1.51
PCB101 (ng/g)	23	1.3	0.5	0.52 – 2.46	0	23	1.3	0.5	0.52 – 2.46	0	23	1.3	0.54	0.52 – 2.46
PCB138 (ng/g)	23	1.7	0.9	0.2 – 4.05	3	20	1.5	0.6	0.20 – 2.4	1	22	1.6	0.78	0.2 – 3.32
Nephthalene (ng/g)	14	83	190	2 – 717	5	8	26	14	7 – 46.9	1	13	34	25	2 – 93
Chrysene (ng/g)	19	120	235	3 – 1030	5	13	36	17	8 – 60	3	16	38	20	3 – 65.1
Fluo-ranthene (ng/g)	21	110	80	10 – 356	4	16	73	35	10 – 128	3	18	85	47	10 – 192
Rock RM: BCR-1														
Sm (µg/g)	274	6.7	0.6	5.1 – 10.8	10	264	6.60	0.40	5.1 – 7.72	10	264	6.6	0.4	---
Sm (µg/g) *					24	250	6.61	0.35	5.5 – 7.55	---	---	---	---	---
Nd (µg/g)	242	29.6	4.2	11 – 50.9	22	220	29.1	1.9	22 – 35	11	231	29	2	---
Nd (µg/g) *					34	208	29.1	1.6	25 – 34	---	---	---	---	---
$^{143}Nd/^{144}Nd$	102	0.512646	0.000031	0.512566 – 0.512732	0	102	0.512646	0.000031	0.512566 – 0.512732	9	93	512.64 **?**	0.03 **?**	---

n: number of data; $\bar{x}$: mean; s: standard deviation; R: range; o: number of discordant outliers detected by a given method; the subscripts $_{in, f}$ and $_l$ refer to, respectively, the initial, final (after applying all discordance tests listed in Table 1; this work), and the box-and-whisker plot method for IAEA-408 (literature, Villeneuve *et al.*, 2002) or the two standard deviation method for BCR-1 (literature, Gladney *et al.*, 1990); the difference between n_{in} and n_f gives the number of discordant outliers detected by the discordancy tests (see Table 1 for these tests); similarly, the difference between n_{in} and n_l gives the number of discordant outliers detected by the box-and-whisker plot method (Villeneuve *et al.*, 1999) for IAEA-408 or by the two standard deviation method (Gladney *et al.*, 1990). Sm and Nd data in BCR-1 were processed without consideration of the analytical methods. *: The rows marked by an asterisk report the statistical results obtained by processing for outliers the data from individual method groups, applying ANOVA test to isolate the data from a method group that are significantly different from the remaining groups, combining the remaining data, and then performing once again the outlier tests to this combined dataset before computing the final statistical parameters. **?**: For the values 512.64 and 0.03 (from Gladney *et al.*, 1990), identified by the **?** mark in the $^{143}Nd/^{144}Nd$ row, it is not clear if these values should be 0.51264 and 0.00003, respectively. The nine outliers detected by these authors are also erroneously done (see the text of the present paper for an explanation for this error in the original reference Gladney *et al.*, 1990).

smaller standard deviations (and somewhat different mean values, although probably not statistically different) were obtained from the multiple-test method as compared to the box-and-whisker plot method for all cases except one compound (PCB101) for which none of the two methods detected any outliers. Note also that, because of the presence of outliers, the initial mean and standard deviation data strongly differ from the final statistical parameters. Finally, had we applied the multiple-test method at a less strict confidence level of 95% (which will probably be consistent with the box-and-whisker plot method although details on the respective confidence level were not provided by Villeneuve *et al.*, 1999), a greater number of outliers would have been identified in most cases than those obtained at 99% confidence level, with the consequent reduction of the standard deviation values obtained by our method and possible changes in the mean values (Table 4).

Application of significance tests, such as F-test and Student-t test (see Verma, 2005 for more details on significance tests), to evaluate the performance of the multiple-test method versus the box-and-whisker plot method is not advisable using the present data, because rather small number of degrees of freedom (7–33; the final number of data remaining after outlier elimination vary from 8 to 31 for n_f and from 13 to 34 for n_l in Table 4) are involved. We consider these numbers too small for quality control purposes. Therefore, an objective statistical comparison of the multiple-test method and the box-and-whisker plot method should be done using larger datasets.

Nevertheless, as in Verma and Quiroz-Ruiz (2006b), but using different datasets, we now conclude that the multiple-test method exemplified in this work can be advantageously used in future to arrive at the final statistical parameters in such interlaboratory studies.

Example 2. Comparison of multiple-test method with "two standard deviation" (2s) method: Two chemical elements and one isotopic ratio in a geochemical reference material (BCR-1) from U.S.G.S.

We present the example of just two elements (petrogenetically important trace elements Sm and Nd) and one widely used radiogenic isotopic ratio ($^{143}Nd/^{144}Nd$) in a rock RM (Columbia River basalt) BCR-1 from the United States Geological Survey (U.S.G.S.), U.S.A. This RM was exten-

sively used four decades ago because it was recommended as *the* RM for all studies of lunar rocks, *i.e.*, researchers reporting data on lunar rocks had to evaluate their data quality by reporting the analysis of this particular RM. Since then, this RM has been widely used in geochemical laboratories, including isotope laboratories; most Nd isotope studies, even today, report $^{143}Nd/^{144}Nd$ in this RM. However, because BCR-1 is no longer available for distribution by the U.S.G.S., it is now replaced by BCR-2 (a sample collected at the same site as BCR-1).

The individual data for BCR-1 for Sm, Nd, and $^{143}Nd/^{144}Nd$ were compiled from Gladney *et al.* (1990) and are presented in supplementary Tables A62, A63, and A64, respectively. No attempt was made to complement them with more recent data on this widely used RM because our main aim was to compare the performance of the multiple-test method with the "two standard deviation" (2s) method used by Gladney *et al.* (1990) to process their compiled data. Although Gladney *et al.* (1990) is a relatively old compilation reference, this is the latest one available on this particular RM (BCR-1) in published literature.

The results of the application of the multiple-test method, along with those from the 2s method, are also summarized in Table 4. For Sm, the same number (10) of outliers were detected by both methods (multiple-test versus 2s) whereas for Nd the multiple-test method detected more outliers (22) than the 2s method (11). However, note that the multiple-test method was applied at the strict confidence level of 99% (and not at the less strict level of 95%, which would, in theory, correspond to the 2s method, and is likely to detect more outliers than the 99% level).

Nevertheless, the statistically correct procedure for applying discordancy tests to such analytical datasets (*e.g.*, Sm and Nd data obtained by different types of analytical methods) would be to: (i) construct statistical samples for each method-group and process them separately for outliers using the multiple-test method; (ii) apply ANOVA ("ANalysis Of VAriance") test to decipher any statistically significant differences among different method-groups, and isolate a particular group if significantly different from the remaining ones at a predetermined confidence level; (iii) combine the data from different method-groups showing no significant differences and process them again for possible outliers; and (iv) calculate the final statistical parameters from the normally distributed final data. We suggest that ANOVA be applied at the strict confidence level of 99% (see Verma, 2005, for the respective tabulated critical values). The results from such a statistically correct procedure are also presented for Sm and Nd in rows marked by an asterisk (*) in Table 4. Gladney *et al.* (1990) did not present such method group-based results for their 2s method although they did so for individual techniques. A clear advantage of the multiple-test method as compared to the 2s method is observed for detecting outlying observations for both Sm and Nd concentration data if we compare our results (see rows marked by an asterisk in Table 4) with the results of "all-groups" presented by Gladney *et al.* (1990).

We applied the F-test and Student-t test to the two sets of Sm and Nd concentration data in order to evaluate if there were significant differences between the final results obtained by the multiple-test method and the 2s method (Table 4). Significant differences (at 95% confidence level for both Sm and Nd, and even at 99% confidence level for Nd) in standard deviation or variance were observed between these two sets of data when statistically correct procedure involving method groups was used for the multiple-test method (see rows marked by an asterisk in Table 4). Variance estimates of Sm and Nd data processed by the multiple-test method were significantly lower than those obtained by the 2s method. Such results will have important implications for instrumental calibrations (using weighted regression techniques such as those used by Guevara *et al.*, 2005) and data quality assessments (using significance tests). More details can be found in Verma (2005). The comparison of the Nd isotope data is discussed below.

First of all, it is very important to note that for $^{143}Nd/^{144}Nd$ from California Institute of Technology (CalTec)-type laboratories that normalize, during data acquisition, the Nd isotopic composition to $^{148}Nd/^{144}Nd = 0.243082$ (see DePaolo and Wasserburg, 1976, for details on CalTec-type laboratories), the data have to be converted according to the following Equation (1), in order to make them consistent with the numerous laboratories around the world that are Lamont-type and use for normalization $^{146}Nd/^{144}Nd = 0.7219$ (see O'Nions *et al.*, 1977, for more details on Lamont-type laboratories).

$$\left(^{143}Nd/^{144}Nd\right)_{Lamont} = \frac{0.512638}{0.511836}\left(^{143}Nd/^{144}Nd\right)_{CalTec} \quad (1)$$

Almost concurrently with the CalTec and Lamont laboratories, Richard *et al.* (1976) from the University of Paris also discovered the utility of the Sm-Nd isotope systematics in Earth Sciences, but they used $^{143}Nd/^{146}Nd$ instead of $^{143}Nd/^{144}Nd$ to show the usefulness of their work.

The existence of two types of laboratories was probably one of the main reasons, among others, why the CalTec researchers introduced the ε_{Nd} notation (for the definition of ε_{Nd}, see DePaolo and Wasserburg, 1976, 1977). Although the actual values of $^{143}Nd/^{144}Nd$ from these two different types of laboratories are drastically different (see the nine lower values in the compilation by Gladney *et al.*, 1990, that are totally distinct from the rest of the compiled data; these values are identified by an asterisk and listed in Table A64 *after* their conversion according to Equation (1) and therefore, do not significantly differ from the rest of the data), the use of ε_{Nd} makes the Nd isotope data from these two types of laboratories directly comparable and fully consistent. It has so happened that the Lamont-type laboratories have become much more numerous than the CalTec-type laboratories (for example, see the compilation by Gladney *et al.*, 1990, in which only nine values, out of a total of 102, were from the CalTec-type laboratories; see also Table A64 in which

CalTec-type data are identified by an asterisk).

This essential conversion for handling $^{143}Nd/^{144}Nd$ from these two types of laboratories was not recognized by Gladney *et al.* (1990) as a requirement to statistically process Nd isotopic data, which resulted in erroneous processing of the isotope data (Table 4). Thus, their 2s method should have certainly, but erroneously, rejected the CalTec-type data (nine data; see Table 4) as outliers although this was not clearly specified by these authors. The multiple-test method, on the other hand, showed that no outlier is present in this dataset at the strict confidence level of 99%.

Verma and Quiroz-Ruiz (2006b) extensively commented on the shortcomings of the 2s method and used a rock RM peridotite JP-1 from Japan to show that just one multiple-outlier test N3 in its three variants (k=2,3,4) applied at 95% confidence level, performed better than the 2s method for detecting outliers. Here, using different datasets (Sm, Nd and $^{143}Nd/^{144}Nd$ in BCR-1), we conclude that the multiple-test method exemplified in this work can be advantageously used in future to arrive at the final statistical parameters in such interlaboratory studies and the probably statistically erroneous 2s method should be abandoned. Such outlier methods based on "fixed" multiples of standard deviation have also been recently criticized by Hayes *et al.* (2007).

Example 3. Other examples of outlier tests applicable in Earth Sciences

We now briefly comment on the need of using the above multiple-test method in numerous geoscience studies. Verma and Quiroz-Ruiz (2006b) already applied the multiple-test method to oxygen isotope data for the Los Azufres hydrothermal system reported by Torres-Alvarado (2002). They also pointed out other studies where the multiple-test method would be useful.

Similarly, this multiple-test method with the new critical values can be successfully applied to process and better interpret: (i) effective weight, variation index and other groundwater data discussed by El-Naqa *et al.* (2006); (ii) univariate and bivariate data of ammonites documented by López-Palomino *et al.* (2006), the latter (bivariate) data using studentized residuals from the regression; (iii) bivariate data for naturally fractured reservoirs from Mexico presented by Miranda-Martínez *et al.* (2006); (iv) data acquisition stage of mass spectrometric instrumentation used for ^{40}Ar-^{39}Ar (Molina-Garza and Ortega-Rivera, 2006) or K-Ar geochronology (Solé *et al.*, 2007), including Rb-Sr, Pb-Pb, Sm-Nd, and Re-Os geochronology or isotope geology (see for example, Wang *et al.*, 1998; Dougherty-Page and Bartlett, 1999, who used just one Dixon test); (v) geochemical data on granitic xenoliths and rocks presented and compiled by Corona-Chávez *et al.* (2006); (vi) microprobe data recently reported by Colombo *et al.* (2007); (vii) chemical data of inactive tailings from the Santa Barbara mineral zone, Chihuahua, documented by Gutiérrez-Ruiz *et al.* (2007); (viii) chemical data for recent and historic tailings of a Pb-Zn-Ag skarn deposit analyzed by Méndez-Ortiz *et al.* (2007); (ix) geochemical and stable isotope data for sedimentary rocks obtained by Nagarajan *et al.* (2007, in press); (x) cation and anion composition data for underground water reported by Ramos-Leal *et al.* (2007); (xi) geotechnical variables for oil prospect exploration decision making discussed by Salleh *et al.* (2007); (xii) major and trace element data for metabasic volcanic rocks presented by Shekhawat *et al.* (2007) and for topaz-bearing rhyolites documented by Rodríguez-Ríos *et al.* (2007); and (xiii) mineral composition data for metagabbroic rocks reported by Cruz-Gámez *et al.* (2007).

We note that the discordant outliers, if present, are of much value to further understand the geological processes provided they are not caused solely by analytical uncertainty. The remaining normally distributed data (after eliminating outlying observations as judged by the applied discordancy tests) can then be used for correctly calculating the central tendency or location (mean) and dispersion or scale (standard deviation) parameters (see Verma, 2005, for more details).

As an example of the application of multiple-test method, we can mention that Colombo *et al.* (2007) applied five single-outlier tests (N1, N2, N7, N8 and N9, with their seven test variants) to ascertain the absence of outliers in their geochemical data before calculating the mean and standard deviation values. We suggest that for small datasets such as those mentioned in this section, the multiple-test method could consist of applying consecutively all nine single-outlier tests (N1, N2, N4–k=1, N7-N10, N14, and N15), with their 13 test variants, to detect possible outlier(s). The multiple-outlier tests (k=2 to k=4 types), with 20 test variants, could additionally be used for larger datasets such as those obtained in interlaboratory studies of RMs.

In summary, therefore, we emphasize that the multiple-test method proposed by Verma (1997) and exemplified in our paper is a recommended procedure to process experimental data under the assumption that the data are drawn from a normal distribution and departure from this assumption due to any contamination or presence of discordant outliers can be properly handled by tests N1 to N15 (all 15 tests with their 33 variants, or only those selected for this purpose).

APPLICABILITY AND PERFORMANCE OF DISCORDANCY TESTS

There has been considerable confusion regarding the applicability of discordancy tests. Miller and Miller (2000, p. 54-57), among others, have expressed the view that Dixon tests are applicable to only small data sets, without actually providing any reasons for this limitation. Surprisingly, these authors intuitively refer to Dixon tests as a Dixon test (Dixon´s Q test). Dixon´s Q test (N7 by Dixon, 1951; N8 by King, 1953; see Table 1) is said to be "valid" for small n =

3 to 7 (Miller and Miller, 2000, p. 54). Unfortunately, this kind of view has plagued the literature in chemistry. Dixon (1951) presented approximate critical values independently for all tests and for *n* up to 30. Then, why should Dixon tests be limited to *n* = 3 to 7? It is true that Dixon tests are especially vulnerable to possible masking effects (Dixon, 1950; Barnett and Lewis, 1994), but this refers to their power and *not* their applicability. In fact, the power of tests as inferred by Dixon (1950) could have been seriously affected by the approximate nature of the critical values by Dixon (1951) – see the large differences between these values and those simulated by Verma and Quiroz-Ruiz (2006a), or those presented in the present work. This performance evaluation (Dixon, 1950) should also be considered rather incomplete from the modern statistical point of view presented by Barnett and Lewis (1994) and Hayes and Kinsella (2003). Therefore, the performance of Dixon and other tests should await further detailed work, which is already in progress by Verma and collaborators.

Nevertheless, statisticians specializing in outlier theory (*e.g.*, Barnett and Lewis, 1994, in their authoritative book, p. 218-234) have pointed out no such "applicability" limitations of single-outlier tests, including Dixon tests, the only limitation being the availability of critical values (see *e.g.*, Verma, 1997; Velasco *et al.*, 2000; Verma and Quiroz-Ruiz, 2006a, 2006b) and the efficiency of discordancy tests. Barnett and Lewis (1994, p. 126) have, in fact, suggested that single-outlier tests should be classified as consecutive tests, and therefore, these tests can be used for identifying multiple (*i.e.*, more than one) outliers (see p. 125-127 in Barnett and Lewis, 1994). The null and alternate hypotheses can thus be repeatedly postulated to test a given dataset for several outliers using single-outlier tests in a consecutive way. Of course, multiple-outlier tests or block procedures (see Barnett and Lewis, 1994) could be more suited for large datasets, but more work is needed to quantitatively evaluate the performance of all single- and multiple-outlier tests. For this reason we believe that new, precise and accurate critical values should be generated not only for small *n* but also for very large *n*, as has been accomplished in the present work. We plan to address these questions of utmost importance in our future work; the first paper in this direction is already in preparation by S.P. Verma, L. Díaz-González, and R. González-Ramírez.

The performance of a discordancy test is an important "quality" factor that should be properly evaluated (Barnett and Lewis, 1994; Velasco and Verma, 1998; Velasco *et al.*, 2000; Hayes and Kinsella, 2003; Efstathiou, 2006). During the last decade, Barnett and Lewis (1994, p. 131) pointed out that much remains to be done in the area of performance of the various available procedures in the presence of outliers. Even today, this statement seems to be true. Thus, appropriate statistical procedures for handling and interpretation of experimental data should be the focus of future work. Masking and swamping effects are also important factors that require special attention and evaluation (Barnett and Lewis, 1994). Finally, because the performance of discordancy tests is not properly known at present although some empirical results were reported by Velasco and Verma (1998) and Velasco *et al.* (2000). Verma and collaborators (Verma, 1997, 1998, 2005; Verma *et al.*, 1998; Guevara *et al.*, 2001; Velasco-Tapia *et al.*, 2001; Verma and Quiroz-Ruiz, 2006a, 2006b; this work) have proposed and practiced the use of the multiple-test method and *not* just some selected test(s). The computation of new, precise and accurate critical values as carried out in the present work should facilitate in future to empirically better evaluate the performance of discordancy tests than attempted by Velasco *et al.* (2000).

CONCLUSIONS

We have used our established and well-tested Monte Carlo-type simulation procedure for generating new, precise and accurate critical values for 15 discordancy tests with 33 test variants for sample sizes up to 1000. For the first time in the published literature, these critical values ($\bar{x}$), along with their individual uncertainty estimates ($se_{\bar{x}}$) as well as new interpolation equations obtained by ANN, are also tabulated. These new critical values will be very useful in many diverse fields of science and engineering, including in quality control in Earth Sciences. Specific examples are presented to highlight the use of these new critical values for quality control. The multiple-test method outlined in the present work seems to perform better than both the box-and-whisker plot and the "two standard deviation" (2s) methods used for processing interlaboratory data on RMs for quality control purposes. Much work is still needed to evaluate the performance of discordancy tests. The new critical values for all samples sizes up to 1000 simulated and interpolated in this work should certainly facilitate the performance evaluation.

ACKNOWLEDGEMENTS

This third manuscript was prepared as a result of the invitation to the first author extended during 2006 by the editor-in-chief Susana Alaniz-Álvarez. We are also grateful to Constantinos E. Efstathiou and an anonymous reviewer for a highly positive evaluation of our work. In spite of this appreciation, both of them pointed out several shortcomings in our earlier manuscript. Their critical comments helped us significantly improve our presentation.

APPENDIX A. SUPPLEMENTARY DATA

Tables A1-A64 can be found at the journal web site <http://satori.geociencias.unam.mx/>, in the table of contents of this issue (electronic supplement 25-1-01) .

REFERENCES

Barnett, V., Lewis, T., 1994, Outliers in Statistical Data. Third edition: Chichester, John Wiley, 584 p.

Bugner, E., Rutledge, D.N., 1990, Modelling of statistical tables for outlier tests: Chemometrics and Intelligent Laboratory Systems, 9(3), 257-259.

Colombo, F., Pannunzio-Miner, E.V., Gay, H.D., Lira, R., Dorais, M.J., 2007, Barbosalita y lipscombita en Cerro Blanco, Córdoba (Argentina): descripción y génesis de fosfatos secundarios en pegmatitas con triplita y apatita: Revista Mexicana de Ciencias Geológicas, 24 (1), 120-130.

Corona-Chávez, P., Reyes-Salas, M., Garduño-Monroy, V.H., Israde-Alcántara, I., Lozano-Santa Cruz, R., Morton-Bermea, O., Hernández-Álvarez, E., 2006, Asimilación de xenolitos graníticos en el campo volcánico Michoacán-Guanajuato: el caso de Arócutin Michoacán, México: Revista Mexicana de Ciencias Geológicas, 23(2), 233-245.

Cruz-Gámez, E.M., Maresch, W.V., Cáceres-Govea, D., Balcázar, N., 2007, Significado de las paragénesis de anfíboles en metagabros relacionados con secuencias de margen continental en el NW de Cuba: Revista Mexicana de Ciencias Geológicas, 24 (3), 318-327.

DePaolo, D.J., Wasserburg, G.J., 1976, Nd isotopic variations and petrogenetic models: Geophysical Research Letters, 3(5), 249-252.

DePaolo, D.J., Wasserburg, G.J., 1977, The source of island arcs as indicated by Nd and Sr isotopic studies: Geophysical Research Letters, 4(10), 465-468.

Dixon, W.J., 1950, Analysis of extreme values: Annals of Mathematical Statistics, 21(4), 488–506.

Dixon, W.J., 1951, Ratios involving extreme values: Annals of Mathematical Statistics, 22(1), 68-78.

Dougherty-Page, J.S., Bartlett, J.M., 1999, New analytical procedures to increase the resolution of zircon geochronology by the evaporation technique: Chemical Geology, 153(1-4), 227-240.

Dybczyński, R., Polkowska-Motrenko, H., Samczynski, Z., Szopa, Z., 1998, Virginia tobacco leaves (CTA-VTL-2) - new Polish CRM for inorganic trace analysis including microanalysis: Fresenius Journal of Analytical Chemistry, 360(3-4), 384-387.

Efstathiou, C.E., 2006, Estimation of type I error probability from experimental Dixon's "Q" parameter on testing for outliers within small size data sets: Talanta, 69(5), 1068–1071.

El-Naqa, A., Hammouri, N., Kioso, M., 2006, GIS-based evaluation of groundwater vulnerability in the Russeifa area, Jordan: Revista Mexicana de Ciencias Geológicas, 23(3), 277-287.

Farre M., Martinez E., Hernando M.D., Fernandez-Alba A., Fritz J., Unruh E., Mihail O., Sakkas V., Morbey A., Albanis T., Brito F., Hansen P.D., Barcelo D., 2006, European ring exercise on water toxicity using different bioluminescence inhibition tests based on Vibrio fischeri, in support to the implementation of the water framework directive: Talanta, 69(2), 323-333.

Gabrovská, D., Rysová J., Filová, V., Plicka, J., Cuhra, P., Kubík, M., Barsová, S., 2006, Gluten determination by gliadin enzyme-linked immunosorbent assay kit: Interlaboratory study: Journal of AOAC International, 89(1), 154-160.

Gill, U., Covaci, A., Ryan, J.J., Emond, A., 2004, Determination of persistent organohelogenated pollutants in human hair reference material (BCR 397): an interlaboratory study: Analytical and Bioanalytical Chemistry, 380(7-8), 924–929.

Gladney, E.S., Jones, E.A., Nickell, E.J., Roelandts, I., 1990, 1988 compilation of elemental concentration data for USGS basalt BCR-1: Geostandards Newsletter, 14(2), 209-359.

Gladney, E.S., Jones, E.A., Nickell, E.J., Roelandts, I., 1991, 1988 compilation of elemental concentration data for USGS DTS-1, G-1, PCC-1, and W-1: Geostandards Newsletter, 15(2), 199–396.

Gladney, E.S., Jones, E.A., Nickell, E.J., 1992, 1988 compilation of elemental concentration data for USGS AGV-1, GSP-1 and G-2: Geostandards Newsletter, 16(2), 111–300.

Govindaraju, K., Potts, P.J., Webb, P.C., Watson, J.S., 1994, 1994 Report on Whin Sill dolerite WS-E from England and Pitscurrie microgabbro PM-S from Scotland: assessment by one hundred and four international laboratories: Geostandards Newsletter, 18(2), 211–300.

Govindaraju, K., Potts, P.J., Webb, P.C., Watson, J.S., 1995, Correction to "1994 Report on Whin Sill dolerite WS-E from England and Pitscurrie microgabbro PM-S from Scotland: assessment by one hundred and four international laboratories": Geostandards Newsletter, 19(1), 97.

Graybeal, D.Y., DeGaetano, A.T., Eggleston, K.L., 2004, Improved quality assurance for historical hourly temperature and humidity: development and application to environmental analysis: Journal of Applied Meteorology, 43(11), 1722–1735.

Grubbs, F.E., 1950, Sample criteria for testing outlying observations: Annals of Mathematical Statistics, 21(1), 27-58.

Grubbs, F.E., 1969, Procedures for detecting outlying observations in samples: Technometrics, 11(1), 1-21.

Guevara, M., Verma, S.P, Velasco-Tapia, F., 2001, Evaluation of GSJ intrusive rocks JG1, JG2, JG3, JG1a, and JGb1 by an objective outlier rejection statistical procedure: Revista Mexicana de Ciencias Geológicas, 18(1), 74-88.

Guevara, M., Verma, S.P., Velasco-Tapia, F., Lozano-Santa Cruz, R., Girón, P., 2005, Comparison of linear regression models for quantitative geochemical analysis: Example of X-ray fluorescence spectrometry: Geostandards and Geoanalytical Research, 29(3), 271-284.

Gutiérrez-Ruiz, M., Romero, F.M., González-Hernández, G., 2007, Suelos y sedimentos afectados por la dispersión de jales inactivos de sulfuros metálicos en la zona minera de Santa Bárbara, Chihuahua, México: Revista Mexicana de Ciencias Geológicas, 24(2), 170-184.

Hassoun, M.H., 1995, Fundamentals of artificial neural networks: Massachusetts Institute of Technology, London, England, 511 p.

Hayes, K., Kinsella, A., 2003, Spurious and non-spurious power in performance criteria for tests of discordancy: The Statistician, 52(1), 69-82.

Hayes, K., Kinsella, A., Coffey, N., 2007, A note on the use of outlier criteria in Ontario laboratory quality control schemes: Clinical Biochemistry, 40 (3-4), 147-152.

Haykin, S., 1999, Neural network, a comprehensive foundation: New York, McMillan, 842 p.

In't Veld, P.H., 1998, The use of reference materials in quality assurance programmes in food microbiology laboratories: International Journal of Food Microbiology, 45(1), 35–41.

King, E.P., 1953, On some procedures for the rejection of suspected data: Journal of American Statistical Association, 48(263), 531–533.

Langton, S.D., Chevennement, R., Nagelkerke, N., Lombard, B., 2002, Analysing collaborative trials for qualitative microbiological methods: accordance and concordance: International Journal of Food Microbiology, 79(3), 175–181.

López-Palomino, R.I., Villaseñor-Martínez, A.B., Olóriz-Sáez, F., 2006, Primer registro del género Vinalesphinctes (Ammonitina) en el Oxfordiano de México: significación bioestratigráfica y consideraciones paleobiográficas en el Jurásico Superior de América: Revista Mexicana de Ciencias Geológicas, 23(2), 162-183.

Lozano, R., Bernal, J.P., 2005, Assessment of eight new geochemical reference materials for XRF major and trace element analysis: Revista Mexicana de Ciencias Geológicas, 22(3), 329-344.

Méndez-Ortiz, B.A., Carrillo-Chávez. A., Monroy Fernández. M.G., 2007, Acid rock drainage and metal leaching on mine waste material (tailings) from a Pb-Zn-Ag skarn deposit: Environmental assessment through static and kinetic laboratory tests: Revista Mexicana de Ciencias Geológicas, 24(2), 161-169.

Miller, J.N., Miller, J. C., 2000, Statistics and Chemometrics for Analytical Chemistry: Essex, England, Prentice Hall, 271 p.

Miranda-Martínez, M.E. Oleschko, K., Parrot, J.F., Castrejón-Vacio, F., Taud, H., Brambila-Paz, F., 2006, Porosidad de los yacimientos naturalmente fracturados: una clasificación fractal: Revista Mexicana de Ciencias Geológicas, 23(2), 199-214.

Molina-Garza, R.S., Ortega-Rivera, A., 2006, Chronostratigraphy and paleomagnetism of the Balsas Group in the Tuzantlán-Copalillo basin, northern Guerrero state, Mexico: Revista Mexicana de

Ciencias Geológicas, 23(2), 215-232.

Nagarajan, R., Madhavaraju, J., Nagendra, R., Armstrong-Altrin, J.S., Moutte, J., 2007, Geochemistry of Neoproterozoic shales of the Rabanpalli Formation, Bhima Basin, Northern Karnataka, southern India: implications for provenance and paleoredox conditions: Revista Mexicana de Ciencias Geológicas, 24(2), 150-160.

Nagarajan, R., Sial, A.N., Armstrong-Altrin, J.S., Madhavaraju, J., Nagendra, R., in press, Carbon and oxygen isotope geochemistry of Neoproterozoic limestones of the Shahabad Formation, Bhima Basin, Karnataka, southern India: Revista Mexicana de Ciencias Geológicas.

Okamoto, K., Yoshinaga, J., Morita, M., 1996, Biological and environmental reference materials from the National Institute for Environmental Studies (Japan): Mikrochimica Acta, 123(1): 15-21.

O'Nions, R.K., Hamilton, P.J., Evensen, N.M., 1977, Variations in $^{143}Nd/^{144}Nd$ and $^{87}Sr/^{86}Sr$ ratios in oceanic basalts: Earth and Planetary Science Letters, 34(1), 13-22.

Papadakis, I., Van Nevel, L., Harper, C., Aregbe, Y., Taylor, P.D.P., 2007, IMEP-12: trace elements in water; objective evaluation of the performance of the laboratories when measuring quality parameters prescribed in the European Directive 98/83/EC: Accreditation and Quality Assurance, 12 (2), 105-111.

Patriarca, M., Chiodo, F., Castelli, M., Corsetti, F., Menditto, A., 2005, Twenty years of the Me.Tos. project: an Italian national external quality assessment scheme for trace elements in biological fluids: Microchemical Journal, 79(1-2), 337–340.

Ramos-Leal, J.A., Durazo, J., González–Morán, T., Juárez-Sánchez, F., Cortés-Silva, A., Johannesson, K.H., 2007, Evidencias hidrogeoquímicas de mezclas de flujos regionales en el acuífero de la Muralla, Guanajuato: Revista Mexicana de Ciencias Geológicas, 24(3), 293-305.

Richard, P., Shimizu, N., Allègre, C.J., 1976, $^{143}Nd/^{146}Nd$, a natural tracer: an application to oceanic basalts: Earth and Planetary Science Letters, 31, 269-278.

Rodríguez-Ríos, R., Aguillón-Robles, A., Leroy, J.L., 2007, Evolución petrológica y geoquímica de un complejo de domos topacíferos en el campo volcánico de San Luis Potosí (México): Revista Mexicana de Ciencias Geológicas, 24(3), 328-343.

Rorabacher, D.B., 1991, Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level: Analytical Chemistry, 63(2), 139-146.

Salleh, S.H., Rosales, E., Flores de la Mota, I., 2007, Influence of different probability based models on oil prospect exploration decision making: A case from Southern Mexico: Revista Mexicana de Ciencias Geológicas, 24 (3), 306-317.

Sang, H.Q., Wang, F., He, H.Y., Wang, Y.L., Yang, L.K., Zhu, R.X., 2006, Intercalibration of ZBH-25 biotite reference material utilized for K-Ar and ^{40}Ar-^{39}Ar age determination: Acta Petrologica Sinica, 22(12), 3059-3078.

Santoyo, E., Guevara, M., Verma, S.P., 2006, Determination of lanthanides in international geochemical reference materials by reversed-phase high performance liquid chromatography: An application of error propagation theory to estimate total analysis uncertainties: Journal of Chromatography A, 1118(1), 73-81.

Shekhawat, L.S., Pandit, M.K., Joshi D.W., 2007, Geology and geochemistry of low grade metabasic volcanic rocks from Salumber area in the Palaeoproterozoic Aravalli Supergroup, NW India: Journal of Earth System Science, 116(6), 511–524.

Solé, J., Salinas, J.C., González-Torres, E., Cendejas Cruz, J.E., 2007, Edades K/Ar de 54 rocas ígneas y metamórficas del occidente, centro y sur de México: Revista Mexicana de Ciencias Geológicas, 24(1), 104-119.

Torres-Alvarado, I.S., 2002, Chemical equilibrium in hydrothermal systems: the case of Los Azufres geothermal field, Mexico: International Geology Review, 44(7), 639-652.

Velasco, F., Verma, S.P., 1998, Importance of skewness and kurtosis statistical tests for outlier detection and elimination in evaluation of Geochemical Reference Materials: Mathematical Geology, 30(1), 109-128.

Velasco, F., Verma, S.P., Guevara, M., 2000, Comparison of the performance of fourteen statistical tests for detection of outlying values in geochemical reference material databases: Mathematical Geology, 32(4), 439-464.

Velasco-Tapia, F., Guevara, M., Verma, S.P., 2001, Evaluation of concentration data in geochemical reference materials: Chemie der Erde-Geochemistry, 61(1), 69–91.

Verma, M.P., 2004, A revised analytical method for HCO^{3-} and $CO3^{2-}$ determinations in geothermal waters: an assessment of IAGC and IAEA interlaboratory comparisons: Geostandards and Geoanalytical Research, 28(3), 391–409.

Verma, S.P., 1997, Sixteen statistical tests for outlier detection and rejection in evaluation of international geochemical reference materials: example of microgabbro PM-S: Geostandards Newsletter: Journal of Geostandards and Geoanalysis, 21(1), 59–75.

Verma, S.P., 1998, Improved concentration data in two international geochemical reference materials, USGS basalt BIR-1 and GSJ peridotite JP-1) by outlier rejection: Geofísica Internacional, 37(3), 215–250.

Verma, S.P., 2005, Estadística Básica para el Manejo de Datos Experimentales: Aplicación en la Geoquímica (Geoquimiometría): México, D. F., Universidad Nacional Autónoma de México, 186 p.

Verma, S.P., Quiroz-Ruiz, A., 2006a, Critical values for six Dixon tests for outliers in normal samples up to sizes 100, and applications in science and engineering: Revista Mexicana de Ciencias Geológicas, 23(2), 133-161.

Verma, S.P., Quiroz-Ruiz, A., 2006b, Critical values for 22 discordancy test variants for outliers in normal samples up to sizes 100, and applications in science and engineering: Revista Mexicana de Ciencias Geológicas, 23(3), 302-319; with electronic tables available at <http://satori.geociencias.unam.mx/23-3.htm>.

Verma, S.P., Orduña-Galván, L.J., Guevara, M., 1998, SIPVADE: A new computer programme with seventeen statistical tests for outlier detection in evaluation of international geochemical reference materials and its application to Whin Sill dolerite WS-E from England and Soil-5 from Peru: Geostandards Newsletter: Journal of Geostandards and Geoanalysis, 22(2), 209–234.

Verma, S.P., Díaz-González, L., Sánchez-Upton, P., Santoyo, E., 2006, OYNYL: A new Computer Program for Ordinary, York, and New York least-Squares linear regressions: WSEAS Transactions on Environment and Development, 2(8), 997-1002.

Villeneuve, J.-P., de Mora, S.J., Cattini, C., Carvalho, F.P., 1999, World-wide and regional intercomparison for the determination of organochlorine compounds, petroleum hydrocarbons, and sterols in sediment sample IAEA-408: Vienna, Austria, International Atomic Energy Agency, Marine Environment Laboratory, 80 p.

Villeneuve, J.-P., de Mora, S.J., Cattini, C., 2002, World-wide and regional intercomparison for the determination of organochlorine compounds and petroleum hydrocarbons in sediment sample IAEA-417: Vienna, Austria, International Atomic Energy Agency, Analytical Quality Control Services, 136 p.

Wang, X.-D., Söderlund, U., Lindh, A., Johansson, L., 1998, U-Pb and Sm-Nd dating of high-pressure granulite- and upper amphibolite facies rocks from SW Sweden: Precambrian Research, 92(4), 319–339.

Xia, Q.W., Hendrickson, E.L., Zhang, Y., Wang, T.S., Taub, F., Moore, B.C., Porat, I., Whitman, W.B., Hackett, M., Leigh, J.A., 2006, Quantitative proteomics of the Archaeon Methanococcus maripaludis validated by microarray analysis and real time PCR: Molecular & Cellular Proteomics, 5 (5), 868-881.

Manuscript received: June 22, 2007
Corrected manuscript received: August 10, 2007
Manuscript accepted: August 12, 2007