



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN
INGENIERÍA**

CENTRO DE INVESTIGACIÓN EN ENERGÍA

**EFICIENCIAS RELATIVAS DE 15 PRUEBAS DE
DISCORDANCIA CON 33 VARIANTES Y EVALUACIÓN
DE DIAGRAMAS DE CLASIFICACIÓN PARA ROCAS
ALTERADAS Y APLICACIÓN EN EL CAMPO
GEOTÉRMICO DE LOS AZUFRES, MICHOACÁN,
MÉXICO.**

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN INGENIERÍA

ENERGÍA – GEOTERMIA

P R E S E N T A

ROSALINDA GONZÁLEZ RAMÍREZ

TUTOR:

DR. SURENDRA P. VERMA

TEMIXCO, MORELOS

Septiembre 2009





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DEDICATORIAS

A Dios por mostrarme su amor al permitirme culminar una meta más en mi vida.

A mis amados padres **Martín† y Pachita†** por darme la vida y ser un ejemplo de trabajo, honestidad y perseverancia, luchando siempre por mantener una familia unida.

A **Armando** por todo el amor, confianza y paciencia que siempre me ha brindado y enseñarme que tengo oportunidad de vivir intensamente cada día sin complicarme por pequeñeces.

A mis queridos hermanos **Vicente, Saúl, Estela, Araceli, Maria Luisa e Irma†**, quien en los momentos difíciles tienen una palabra de aliento que me ayuda a continuar.

A mis adorables sobrinos: **Ángel, Saúl, Israel, Martín, Eduardo, Juan Carlos, Azucena, Omar, Enrique, Ángeles, Ariana, Alejandro, Rogelio, Fernando, Arturo, e Irma**. Y a sus hermosos descendientes. Por todo el amor desinteresado y sincero que me han dado.

A la **familia Flores-Rosas** por la hospitalidad, amistad y cariño que siempre recibí.

AGRADECIMIENTOS

Mi agradecimiento al Dr. Armando Parra García por todo el apoyo y motivación durante el transcurso de este trabajo. Me invitaste a conocer el CIE-UNAM e inicié esta gran aventura de la ciencia e ingeniería. Gracias...

Mi gratitud especial al Lic. Enrique Alejandro García, por la amistad y confianza que depositó en mí, apoyándome siempre para culminar este doctorado.

Mi agradecimiento especial a mi director de Tesis, Dr. Surendra P. Verma por regalarme largas horas de su valioso tiempo para guiarme y formarme durante el proyecto doctoral, así como también por su brillante dirección para culminar esta tesis. Gracias doctor por ayudarme a alcanzar esta meta.

Mi agradecimiento a los Doctores Edgar, R. Santoyo Gutiérrez, Ignacio S. Torres Alvarado, Eduardo González Partida, y John S. Armstrong Altrin, por haber aceptado ser miembros del Jurado Doctoral y por sus recomendaciones que enriquecieron el contenido de este trabajo.

Al Dr. Rodolfo Rodríguez Ríos† por su apoyo y valiosa colaboración en este proyecto doctoral.

Al Mtra. Mirna Guevara García por sus asesorías y apoyo en el laboratorio de fluorescencia de rayos-X.

A los Doctores, Jorge A. Andaverde Arredondo y Lorena Díaz González por su asesoría académica y apoyo incondicional.

A la M. en I. Alejandra Álvarez del Castillo Moctezuma, por sus asesorías técnicas y apoyo desinteresado y amable en todo momento.

A los Ingenieros Alfredo Quiroz-Ruíz y María Abdelaly Rivera Gómez por sus asesorías en programas de computación.



Agradezco también al Centro de Investigación de Energía de la Universidad Nacional Autónoma de México (CIE-UNAM), el tiempo y el espacio facilitado para la realización de este proyecto. Así como a las personas que con su amistad, tiempo y apoyo contribuyeron en la realización de este trabajo.

Mi agradecimiento al programa de becas CONACYT por su apoyo económico durante mis estudios de doctorado.

A la Dirección General de Educación Tecnológica Industrial (DGETI) por todo el tiempo que me otorgó para la realización y culminación del doctorado.

CONTENIDO

Índice de Figuras	iv
Índice de Tablas	v
Resumen	1
Abstract	3
INTRODUCCIÓN.	
Planteamiento del problema y justificación	5
Objetivo general	7
Objetivos particulares	7
Organización de la tesis	8
CAPÍTULO 1: BASES DE DATOS DE 35 MATERIALES DE REFERENCIA GEOQUÍMICA (MRG)	
Introducción	11
1.1 Compilación de MRG	11
1.2 Descripción Breve de la Base de Datos	13
CAPÍTULO 2: EVALUACIÓN DE 15 PRUEBAS DE DISCORDANCIA CON 33 VARIANTES	
Introducción	15
2.1 Programa de cómputo utilizado para la evaluación	15
2.2 Criterio de eficiencia relativa (REC) de las pruebas de discordancia	17
2.3 Criterio relativo global de datos desviados (ROC) de las pruebas de discordancia	17
2.4 Resultados y análisis de la evaluación estadística de 15 pruebas de discordancia con 33 variantes	18
CAPÍTULO 3: PRUEBAS DE SIGNIFICANCIA	
Introducción	21
3.1 Aplicación de las Pruebas F -Fisher y t -Student	21
3.2 Aplicación de la prueba de ANOVA	25
CAPÍTULO 4: APLICACIÓN ESTADÍSTICA A LAS BASES DE DATOS MRG Y SU UTILIDAD EN LA CALIBRACIÓN DE FLUORESCENCIA DE RAYOS-X (FRX)	
Introducción	27
4.1 Procesamiento estadístico de las bases de datos	27
4.1a Procesamiento estadístico de la muestra W-2	29
4.1b Procesamiento estadístico de la muestra JA-1	33
4.2 Calibración del equipo de fluorescencia de rayos-X (FRX)	39

CAPÍTULO 5: COMPARACIÓN DE MÉTODOS DE PREPARACIÓN DE PASTILLAS PRENSADAS Y DISCOS FUNDIDOS EN FLUORESCENCIA DE RAYOS -X (FRX)

Introducción.	41
5.1 Base de datos y su procesamiento	41
5.2 Comparación de resultados de DODESYS con la literatura	42
5.3 Comparación de resultados de fluorescencia de rayos-X (FRX) obtenidos en pastillas prensadas y discos fundidos	43

CAPITULO 6: EVALUACIÓN DE DIAGRAMAS DE CLASIFICACIÓN PARA ROCAS ALTERADAS

Introducción.	45
6.1 Base de datos mundial	45
6.2 Diagramas de Clasificación para rocas alteradas	46
6.2.1 Diagrama de Clasificación para rocas alteradas de Winchester & Floyd (1977)	47
6.2.2 Diagramas de Clasificación para rocas alteradas de Hastie et al. (2007)	48
6.3 Metodología para la evaluación de diagramas para rocas alteradas	48
6.4 Evaluación de diagramas de clasificación para rocas alteradas de Winchester & Floyd (1977) y Hastie (2007)	49
6.4.1 Diagrama de Winchester & Floyd (1977)	49
6.4.2 Diagramas de Clasificación de Hastie et al. (2007)	50

CAPÍTULO 7: APLICACIÓN DE DIAGRAMAS DE WINCHESTER Y FLOYD PARA LA CLASIFICACIÓN DE ROCAS ALTERADAS EN EL CAMPO GEOTÉRMICO DE LOS AZUFRES MICH., MÉXICO

Introducción	53
7.1 Base de datos del campo geotérmico de Los Azufres, Michoacán México.	53
7.2 Diagramas de clasificación TAS y diagrama de Winchester & Floyd (1977) del campo geotérmico Los Azufres Mich., México	54
7.3 Discusión de resultados del diagrama W&F de clasificación para rocas alteradas	55
7.3.1 Aplicación del diagrama de clasificación W&F utilizando rocas Frescas	55
7.3.2 Aplicación del diagrama W&F utilizando rocas Alteradas	58

CONCLUSIONES Y TRABAJOS FUTUROS 63

REFERENCIAS 67

APÉNDICES 73

Apéndice 1. Verma, S.P, Díaz-González, L., and González-Ramírez R. (2009) Relative efficiency of single-outlier discordancy tests for processing geochemical data on reference materials and application to instrumental calibrations by a weighted least-squares linear regression model. *Geostandards and Geoanalytical Research*, 33 (1), 29-48.

Apéndice 2. González-Ramírez R., Díaz-González, L., Verma S.P. (2009) Eficiencia relativa de 15 pruebas de discordancia con 33 variantes aplicadas al procesamiento de datos geoquímicas. Revista Mexicana de Ciencias Geológicas, 26 (2), 501-515

Apéndice 3. Verma, S.P., González-Ramírez R., Rodríguez-Rios R. (2009a) Comparison of two sample preparation methods in x-ray fluorescence spectrometry for the analysis of Ni and Cr. Geostandards and Geoanalytical Research (enviado: 10 de mayo, 2009; revisado: 10 de agosto, 2009; se espera respuesta del editor).

Apéndice 4. Verma S.P., Rodríguez-Rios, R., González-Ramírez R. (2009b) Statistical evaluation of classification diagrams for altered igneous rocks. Turkish Journal of Earth Sciences (enviado: 22 de enero, 2009; aceptado con cambios menores: 3 de Julio, 2009; revisado: 15 de julio, 2009; disponible en línea de la revista: 15 de agosto, 2009).

INDICE DE FIGURAS

Fig. 4.1	Diagrama de flujo de la metodología estadística aplicada a los elementos mayores de los MRG de la base de datos utilizada en el Apéndice 1 (Verma et al., 2009).	29
Fig. 4.2	Gráfica de la muestra de andesita JA-1 después de depurar los datos desviados.	37
Fig. 4.3	Gráfica de la muestra de JA-1 con el grupo 5 depurado.	38
Fig. 7.1	Evaluación estadística del diagrama de clasificación TAS, usando rocas frescas intermedias (andesita y andesita basáltica).	56
Fig. 7.2	Evaluación estadística del diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) usando rocas frescas intermedias.	56
Fig. 7.3	Evaluación estadística del diagrama de clasificación TAS usando rocas frescas ácidas (riolitas).	57
Fig. 7.4	Evaluación del diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) usando rocas frescas ácidas.	58
Fig. 7.5	Evaluación estadística del diagrama de clasificación TAS usando rocas alteradas intermedias (andesita, andesita basáltica y traquiandesita).	60
Fig. 7.6	Evaluación estadística del diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) usando rocas alteradas intermedias.	60
Fig. 7.7	Evaluación estadística del diagrama de clasificación TAS usando rocas alteradas ácidas (riolitas y dacitas).	61
Fig. 7.8	Evaluación del diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) usando rocas alteradas ácidas.	62

INDICE DE TABLAS

Tabla 1.1	Una parte de la base de datos capturada en Excel.	12
Tabla 3.1	Datos para aplicar el estadístico de prueba de <i>F</i> -Fisher (tomados de la Tabla 2 del Apéndice 2; González-Ramírez et al., 2009).	22
Tabla 3.2	Prueba <i>t</i> -Student aplicada a los resultados de pruebas de discordancia que evalúan el valor más pequeño en un conjunto de datos organizado de manera ascendente.	24
Tabla 4.1	Resultados de la carpeta Report Data Summary de W-2.	30
Tabla 4.2	Resultados de la carpeta Report Test Data Summary de W-2.	31
Tabla 4.3	Resultados de la carpeta Performance de W-2.	31
Tabla 4.4	Ubicación del dato desviado de W-2.	32
Tabla 4.5	Muestra depurada de datos discordantes de W-2.	32
Tabla 4.6	Resultados de la carpeta Report Data Summary de JA-1.	34
Tabla 4.7	Resultados de la carpeta Report Test Data Summary de JA-1.	34
Tabla 4.8	Resultados de la carpeta Performance de JA-1.	35
Tabla 4.9	Muestra depurada de datos discordantes de JA-1.	36
Tabla 4.10	Concentraciones de elementos mayores para las andesitas JA-1 y JA-2, procedentes de Japón.	40
Tabla 7.1	Evaluación del diagrama de Nb/Y–Zr/TiO ₂ (Winchester & Floyd, 1977) utilizando rocas frescas del campo geotérmico de los Azufres Mich., México, comparado con la clasificación de rocas volcánicas de la IUGS.	55
Tabla 7.2	Evaluación del diagrama de Nb/Y–Zr/TiO ₂ (Winchester & Floyd, 1977) utilizando rocas alteradas del campo geotérmico de los Azufres Mich., México, comparado con la clasificación de rocas volcánicas de la IUGS.	59

RESUMEN

Las pruebas de discordancia son una herramienta estadística útil en las ciencias e ingenierías. Se han realizado compilaciones de datos geoquímicas, a los cuales se les han aplicado tratamientos estadísticos con la finalidad de conocer la eficiencia de las pruebas de discordancia. Sin embargo, una de las limitantes encontradas ha sido la falta de disponibilidad de valores críticos para muestras relativamente grandes, por ejemplo, tamaño de muestra arriba de 30 para las pruebas de Dixon. Por otra parte, para realizar esta evaluación, es preferible contar con una amplia base de datos que permita evaluar todas las pruebas de discordancia en diferentes tamaños de muestras, lo cual no se había logrado en la literatura revisada. Estos antecedentes motivaron a realizar un estudio completo de las pruebas de discordancia, partiendo de la compilación de una extensa base de datos geoquímicos y la incorporación, por primera vez, de los valores críticos para muestras de gran tamaño obtenidos recientemente en el Centro de Investigación de Energía de la Universidad Nacional Autónoma de México (CIE-UNAM). Por lo tanto, una de las principales metas del trabajo doctoral fue la evaluación de 15 pruebas de discordancia con 33 variantes utilizando una base de datos geoquímicos que contiene 35 materiales de referencia geoquímica (MRG) procedentes de cuatros países (U.S.A., Canadá, Japón y Sudáfrica). Esta base de datos fue diseñada y creada para el desarrollo de la presente tesis (Capítulo 1).

En este proyecto de investigación, se obtuvo el orden de 15 pruebas de discordancia con sus 33 variantes en función de sus eficiencias (Capítulo 2; Apéndices 1 y 2). Como el resultado más sobresaliente, se puede mencionar que las pruebas de curtosis, coeficiente de asimetría y de Grubbs son más eficientes que las de Dixon debido a que los valores de eficiencia obtenidos al evaluar las pruebas de Dixon presentaron los valores más bajos que las otras pruebas.

Otro aspecto importante de la tesis fue aplicar la metodología estadística de pruebas de significancia (*t*-Student, *F*-Fisher y ANOVA) a los datos geoquímicos medidos en los MRG con diferentes grupos de métodos analíticos. Esto permitió evaluar o definir como hipótesis de esta parte del proyecto doctoral si es estadísticamente aceptable combinar los datos de elementos

mayores y traza obtenidos por diferentes métodos analíticos, o mantenerlos separados para obtener los parámetros de tendencia central y dispersión (Capítulo 3).

Una vez aplicadas las pruebas de discordancia, detectados los datos desviados en los MRG e interpretados los resultados de las pruebas de significancia, fue posible determinar las medidas de tendencia central y dispersión para elementos mayores (Apéndice 1; Verma et al., 2009). Estos resultados fueron aplicados en la calibración de fluorescencia de rayos-x (FRX) (Capítulo 4) usando una metodología de regresión lineal ponderada (RLP). Así mismo, se realizaron los cálculos de los parámetros estadísticos para los elementos traza Cr y Ni, que fue de utilidad para establecer la comparación estadística de los métodos de preparación de pastillas prensadas y discos fundidos utilizados en FRX (Capítulo 5; Apéndice 3; Verma et al., 2009a). Se demostró que ambas técnicas de preparación proporcionan resultados de tendencia central estadísticamente similares, mientras que los discos fundidos presentaron, en general, mayor dispersión que las pastillas prensadas (Apéndice 3; Verma et al., 2009a).

Dado que en los campos geotérmicos se presentan tanto las rocas frescas como alteradas, una de las tareas de exploración de estos campos es clasificar estas rocas. En el caso de muestras de rocas frescas, existe un consenso (IUGS; Unión Internacional de Ciencias Geológicas) de utilizar el diagrama TAS (álcalis totales contra sílice) y la norma CIPW. Sin embargo en el caso de las rocas alteradas, no se tienen establecidas las normas de clasificación. Por lo tanto, el objetivo final de este proyecto doctoral fue la evaluación de los diagramas de clasificación para rocas alteradas de Nb/Y-Zr/TiO₂ de Winchester & Floyd (1977) y de Co-Th de Hastie et al. (2007) utilizando datos de rocas frescas provenientes de muchas partes del mundo. Los resultados obtenidos de esta evaluación demuestran que se requieren nuevas propuestas de clasificación para rocas alteradas (Capítulo 6; Apéndice 4; Verma et al., 2009b). Como una aplicación, y por no contar con los datos de Co y Th, se evaluó solamente el diagrama de Winchester & Floyd (1977) para muestras de rocas colectadas en el campo geotérmico de Los Azufres, Mich., México (Capítulo 7). Esta aplicación en Los Azufres confirmó los resultados obtenidos en la evaluación del Capítulo 6.

ABSTRACT

Discordancy tests are a useful statistical tool in science and engineering. Geochemical data were compiled in a database, which was statistically processed to evaluate the efficiency of discordancy tests. However, one of the limitations encountered was the lack of availability of critical values for relatively large samples, for example, sample sizes above 30 for Dixon tests. Moreover, to perform this evaluation, it is preferable to have an extensive database to evaluate all discordancy tests with all sample sizes, which was not yet achieved in the published literature. These reasons motivated me to perform a complete study of discordance tests, based on the compilation of an extensive geochemical database and the incorporation for, the first time, of new, precise and accurate critical values for large samples obtained recently in the Energy Research Center of the National Autonomous University of México (CIE-UNAM). Therefore, one of the main goals of the doctoral work was the evaluation of 15 tests with 33 variants using a geochemical database for 35 geochemical reference materials (GRM) from four countries (U.S.A., Canada, Japan, and South Africa). This database was designed and created for the development of this thesis (Chapter 1).

In this research project, 15 discordancy tests with 33 variants were evaluated and arranged according to their efficiencies (Chapter 2, Appendices 1 and 2). As the most important result, it can be mentioned that the kurtosis, skewness and Grubbs tests are more efficient than Dixon tests. This is because the efficiencies obtained for Dixon tests were among the lowest values.

Another important aspect of the thesis was to apply the methodology of statistical significance tests (Student t, Fisher F and ANOVA) to geochemical data in GRM obtained by different analytical methods. This allowed us to define or evaluate the null hypothesis if it is statistically acceptable to combine data from major- and trace-elements from different analytical methods, or the alternate hypothesis if we should deal them separately for the estimation of central tendency and dispersion parameters (Chapter 3).

Once the discordancy tests were applied, the discordant outliers were detected in the database of the GRM and the results of significance tests were interpreted, it was possible to determine central tendency and dispersion parameters for major-elements (Appendix 1; Verma et al., 2009). These results were applied in the calibration of x-ray fluorescence (XRF) (Chapter 4) using a weighted linear regression (WLR) model. Similarly, these statistical parameters were also calculated for two trace elements Cr and Ni. This work was useful for statistical comparison of two sample preparation methods (fused beads and pressed powder pellets) used in XRF (Chapter 5; Appendix 3; Verma et al., 2009a). It was shown that both sample preparation methods provided statistically similar central tendency results for both Cr and Ni. However, the data obtained from the fused bead method presented greater dispersion than the pressed powder method (Appendix 3; Verma et al., 2009a).

Because both fresh and altered rocks are present in geothermal fields, one of the tasks of geothermal exploration is to classify these rocks. In the case of fresh rock samples, there is a consensus (IUGS, International Union of Geological Sciences) to use the TAS (total alkalis versus silica) diagram and standard CIPW norm. However in the case of altered rocks, there is no consensus for their classification, although some diagrams have been widely used. Therefore, the ultimate goal of this doctoral project was to assess one such widely used classification diagram (Nb/Y-Zr/TiO₂) by Winchester & Floyd (1977) for the classification of altered rocks and a recently proposed Co-Th diagram by Hastie et al. (2007). This was done using data for fresh rocks from many parts of the world. The results of this evaluation show that there is a need for new classification diagrams for altered rocks that could be consistent with the TAS classification (Chapter 6, Appendix 4; Verma et al., 2009b).

Finally, a direct application of classification diagrams for altered rocks to the geothermal field of Los Azufres, Michoacán, Mexico is reported. Because no Co and Th data have been reported from this field, only the diagram of Winchester & Floyd (1977) could be evaluated (Chapter 7). This application in Los Azufres confirmed the results obtained in the evaluation reported in Chapter 6.

INTRODUCCIÓN

Planteamiento del problema y justificación

El manejo e interpretación de datos experimentales forma una parte esencial de todas las ciencias e ingenierías (Verma, 2005). Desde el punto de vista estadístico, el manejo de estos datos consiste en la estimación de los parámetros de tendencia central y de dispersión de datos experimentales univariados. En la literatura se tiene una amplia información sobre métodos estadísticos para estimar parámetros de tendencia central y dispersión de datos univariados (Barnett y Lewis, 1994). Estos se distribuyen en dos tipos de metodologías generales (Verma, 2005): (a) métodos robustos, que utilizan criterios de estimación que se ven poco afectados por la presencia de valores desviados; y (b) métodos de valores desviados, que consisten en la aplicación de pruebas de discordancia para la detección y eliminación de valores desviados. Como una alternativa, se han empleado métodos de valores desviados basados en la desviación estándar (σ) de la población, como podría ser el método de dos desviaciones estándar; sin embargo, esta clase de métodos son estadísticamente erróneos (Barnett y Lewis, 1994). Por esta razón, en la presente tesis no se abordarán esta clase de métodos.

Cabe aclarar que la media aritmética y la desviación estándar, muy frecuentemente usados como parámetros de la tendencia central y dispersión, respectivamente, pertenecen a esta segunda categoría (métodos de valores desviados), por lo cual es de suma importancia que se apliquen metodologías estadísticamente correctas, previo al empleo de estos parámetros.

En esta tesis no se abordarán los métodos robustos. Más bien, se concentrará la atención en el grupo de métodos basados en valores desviados. Sin embargo, en esta categoría existe una gran diversidad de pruebas de discordancia, para las cuales se tiene solamente información limitada sobre sus eficiencias globales. Es importante, por lo tanto, utilizar la base de datos geoquímicos con el objeto de conocer estas eficiencias. Las pruebas de discordancia tienen una extensa aplicabilidad en diferentes campos de las ciencias geológicas y de otras áreas del conocimiento, incluyendo el control de calidad mediante materiales de referencia. Esto se debe a que, posterior a la aplicación de esta metodología, se pueden emplear la media y la desviación

estándar como parámetros estadísticos apropiados para describir la tendencia central y dispersión de datos, respectivamente. Otra razón para usar la metodología de pruebas de discordancia es que las pruebas de significancia, como F -Fisher, t -Student y ANOVA, se basan en la suposición de que los datos representan muestras normales, sin ninguna contaminación estadística (p. ej., Jensen et al. 1997).

Como existe una gran diversidad de pruebas de discordancia (Barnett y Lewis, 1994), consideramos pertinente evaluar 15 pruebas de discordancia con sus 33 variantes, partiendo de una extensa base de datos que contiene materiales de referencia geoquímica (MRG). El conocer las eficiencias relativas de cada una de las pruebas de discordancia ayudará a decidir si es viable aplicar un esquema basado en un número pequeño de ellas o, si es necesario, aplicar todas las disponibles como inicialmente lo propuso Verma (1997).

Los datos depurados de MRG, después de la aplicación de la metodología de pruebas de discordancia, se pueden utilizar en aplicaciones estadísticas de calibración de instrumentos o procedimientos analíticos como la fluorescencia de rayos-x (FRX). Así mismo, esta metodología de valores desviados sirve para evaluar la calidad de datos analíticos obtenidos por diferentes técnicas en los MRG ya que posteriormente, se pueden aplicar pruebas de significancia para evaluar hipótesis nula y alterna (Verma, 2005). Como una herramienta estadística adicional, se puede utilizar regresiones, tanto ordinarias como ponderadas, para la evaluación de métodos analíticos (Miller y Miller, 1993; Verma, 2005). En este contexto, para la FRX existen dos tipos de métodos de preparación de muestras – discos fundidos y pastillas prensadas. Sería útil evaluar estadísticamente estos métodos para la determinación de algunos elementos traza en rocas ígneas o sedimentarias.

Por otra parte, en el área de la geotermia se tiene la necesidad de clasificar tanto las rocas frescas como las alteradas como parte de los estudios de campos geotérmicos. Una actividad primaria en estos estudios es la nomenclatura de las rocas. Muchos de los campos geotérmicos se ubican en las zonas de rocas ígneas tanto volcánicas como plutónicas. Aunque para las rocas frescas existe un consenso (Unión Internacional de Ciencias Geológicas) para asignar nombres

(Le Bas et al. 1986; Le Bas, 2000), para rocas alteradas no se cuenta con esta información. Para ello, se han propuesto algunos diagramas (Winchester y Floyd, 1977; Hastie et al. 2007). Sin embargo, no se conoce todavía si estos diagramas proporcionan nombres a rocas alteradas que sean compatibles o consistentes con las rocas frescas. Por lo tanto, los diagramas de clasificación deben ser evaluados mediante bases representativas de datos mundiales. Una vez reconocida la eficacia de estos diagramas, es pertinente aplicarlas a un campo geotérmico dado que actualmente no existen otras opciones para nombrar a las rocas hidrotermalmente alteradas.

Objetivo general de la tesis doctoral

El objetivo general de este proyecto es determinar las eficiencias relativas de 15 pruebas de discordancia con 33 variantes; evaluar los diagramas de clasificación de rocas alteradas mediante la graficación de rocas frescas; y aplicar en el campo geotérmico de los Azufres, Mich., con la finalidad de comparar resultados al clasificar rocas frescas de un campo geotérmico de México haciendo uso de estos diagramas.

Objetivos particulares

- i. Compilar una base de datos geoquímicos mundiales de 35 MRG para evaluar la eficiencia de las pruebas de discordancia.
- ii. Evaluar las eficiencias relativas de 15 pruebas de discordancia con 33 variantes y conocer cuáles pruebas –sencillas y múltiples– son más eficientes para detectar datos desviados.
- iii. Comparar los métodos de preparación pastillas prensadas y discos fundidos para medir Ni y Cr (elementos traza) por FRX.
- iv. Evaluar los diagramas de Winchester y Floyd (1977) y Hastie et al. (2007), mediante la comparación gráfica con el diagrama TAS, utilizando datos geoquímicos de rocas frescas provenientes de bases de datos mundiales, con el fin de conocer si la nomenclatura es correcta en rocas alteradas.

- v. Aplicar los diagramas de clasificación para rocas alteradas analizadas a la fecha en el campo geotérmico de Los Azufres, Mich., México.

Organización de la tesis

En esta tesis se describen los aspectos más sobresalientes del proyecto doctoral. Es importante mencionar que debido a que se incluyen (en los Apéndices) los artículos generados en el proceso de esta investigación, se describen los capítulos en forma breve y concreta, asumiendo que los detalles específicos se describen en los artículos que fueron publicados, aceptados o enviados para su publicación.

A continuación, se describe la organización de la tesis.

En la **Introducción**, se describe la importancia de la evaluación de las pruebas de discordancia para conocer la eficiencia de cada una de ellas, así como el planteamiento del problema, justificación, objetivo general y metas de este proyecto.

En el **Capítulo 1** se dan a conocer los aspectos más relevantes de la creación y compilación de la base de datos de 35 MRG. El **Capítulo 2** describe la evaluación de 15 pruebas de discordancia con 33 variantes. El **Capítulo 3** presenta las pruebas de significancia que fueron aplicadas a la base de datos de MRG. El **Capítulo 4** muestra la aplicación estadística a la base de datos MRG y su utilidad en la calibración de fluorescencia de rayos-X (FRX). El **Capítulo 5** muestra la comparación de los métodos de preparación de pastillas prensadas y discos fundidos para ser medidos por FRX. El **Capítulo 6** da a conocer los resultados de la evaluación de diagramas de clasificación para rocas alteradas. El **Capítulo 7** presenta la aplicación de los diagramas de clasificación para rocas frescas y alteradas en el campo geotérmico de Los Azufres, Michoacán México. Así mismo, se presentan las **Conclusiones** finales de este trabajo, **Trabajos** futuros y las **Referencias**. Finalmente, se anexan **Apéndices** que contienen los siguientes trabajos:

Apéndice 1.

Contiene el trabajo (Verma et al., 2009) publicado en una revista arbitrada incluida en el Science Citation index (SCI). Se trata del siguiente trabajo:

Verma, S.P, Díaz-González, L., and González-Ramírez R. (2008)

Relative efficiency of single-outlier discordancy tests for processing geochemical data on reference materials and application to instrumental calibrations by a weighted least-squares linear regression model. *Geostandards and Geoanalytical Research*, 33 (1), 29-48.

Apéndice 2.

Contiene el trabajo (González-Ramírez et al., 2009) publicado en una revista arbitrada incluida en el Science Citation index (SCI).

González-Ramírez R., Díaz-González, L., Verma S.P. (2009)

Eficiencia Relativa de 15 pruebas de discordancia con 33 variantes aplicadas al procesamiento de datos geoquímicos. *Revista Mexicana de Ciencias Geológicas*, 26 (2), 501-515

Apéndice 3.

Contiene el trabajo (Verma et al., 2009a) enviado y revisado en una revista arbitrada incluida en el Science Citation index (SCI).

Verma, S.P., González-Ramírez R., Rodríguez-Rios R. (2009a)

Comparison of two sample preparation methods in x-ray fluorescence spectrometry for the analysis of Ni and Cr. *Geostandards and Geoanalytical Research* (enviado: 10 de mayo, 2009; revisado: 10 de agosto, 2009; se espera respuesta del editor).

Apéndice 4.

Contiene el trabajo (Verma et al., 2009b) aceptado a una revista arbitrada incluida en el Science Citation index (SCI).

Verma S.P., Rodríguez-Rios, R., González-Ramírez R. (2009b)

Statistical evaluation of classification diagrams for altered igneous rocks. *turkish journal of earth sciences* (enviado: 22 de enero, 2009; aceptado con cambios menores: 3 de Julio, 2009; revisado: 15 de julio, 2009; disponible en línea de la revista: 14 de agosto, 2009)

CAPÍTULO 1

BASE DE DATOS DE 35 MATERIALES DE REFERENCIA GEOQUÍMICA (MRG)

Introducción

La base de datos se formó por la recopilación exhaustiva de la composición química de 35 MRG provenientes de cuatro países: Canadá con 3 MRG preparados por *Energy, Mines and Resources*; E.U.A con 9 preparados por *U. S. Geological Survey* (USGS); Japón con 17 de preparados por *Geological Survey of Japan* (GSJ); y Sudáfrica con 6 preparados por *National Institute of Metallurgy* (NIM). La base contiene la concentración de 11 elementos mayores y 71 elementos traza, incluyendo elementos de tierras raras (REE; *rare-earth elements*). Los nombres y claves de materiales se encuentran en el Apéndice 1 (Verma et al., 2009). Este Apéndice presenta el trabajo publicado por Verma et al. (2009), en el cuál realicé la compilación de la base de datos y las pruebas de significancia. Cabe aclarar que este artículo también fue incluido en otra tesis doctoral (Díaz-González, 2008) quien realizó la parte del procesado de los datos por el programa DODESYS (Verma y Díaz-González, no-publicado, 2009).

Los datos fueron capturados en el software comercial *Statistica*® que es una herramienta estadística muy robusta para el manejo de gran cantidad de datos y permite exportar o importar datos en otros formatos como el *Excel*®.

1.1 Compilación de MRG

La creación y compilación de la extensa base de datos MRG fue una actividad fundamental del trabajo de investigación. La mayor parte de los datos se obtuvo de compilaciones de autores descritos en el Apéndice 1 (Verma et al., 2009), artículos publicados en revistas, e Internet. Inicialmente, se realizó la captura en el programa Excel debido a que era

posible escanear artículos en papel y facilitar así la compilación de una mayor cantidad de datos. Se diseñó un formato para cada uno de los MRG que contenía 301 columnas y 200 renglones. Cada renglón contiene datos referentes a la concentración de elementos mayores (SiO_2 , TiO_2 , Al_2O_3 , Fe_2O_3 , FeO , MnO , MgO , CaO , Na_2O , K_2O y P_2O_5) y traza (Ag, As, Au, B, Ba, Be, Bi, Br, C, Cd, Ce, Cl, Co, Cr, Cs, Cu, Dy, Er, Eu, F, Ga, Gd, Ge, Hf, Hg, Ho, I, In, Ir, La, Li, Lu, Mo, Nb, Nd, Ni, Os, Pb, Pd, Pr, Pt, Rb, Re, Rh, Ru, S, Sb, Sc, Se, Sm, Sn, Sr, Ta, Tb, Te, Th, Tl, Tm, U, V, W, Y, Yb, Zn y Zr), método analítico de medición y la referencia de la cual se había obtenido la información.

En la Tabla 1.1, que forma parte de la base de datos, se indica, como ejemplo, la concentración de Fe_2O_3 en porcentaje (%) siendo ésta 0.72, el método (M) para medir la concentración de Fe_2O_3 (MFe₂O₃), en este caso el método (M) es CALC que quiere decir calculado, y 92-0228JA corresponde a la referencia “RFe₂O₃” de la cual fue tomada la información. De igual forma, se presenta en la Tabla 1.1 la información acerca de FeO.

Tabla 1.1 Una parte de la base de datos capturada en Excel.

Fe ₂ O ₃	MFe ₂ O ₃	RFe ₂ O ₃	FeO	MFeO	RFeO
%			%		
0.72	CALC	92-0228JA	1.09	TITR	92-0228JA
0.73	CALC	92-0228JA	1.17	TITR	92-0228JA
0.78	CALC	92-0228JA	1.31	COLOR	92-0228JA
0.88	CALC	92-0228JA	1.31	VOLU	92-0228JA
0.88	CALC	92-0228JA	1.32	TITR	92-0228JA
0.90	CALC	92-0228JA	1.34	VOLT	92-0228JA
0.92	CALC	92-0228JA	1.36	TITR	92-0228JA
0.96	CALC	92-0228JA	1.36	COLOR	92-0228JA
0.97	CALC	92-0228JA	1.36	TITR	92-0228JA

La base de datos MRG fue transferida a una hoja del software comercial STATISTICA para tener un manejo adecuado de los datos. En este nuevo formato, se organizaron los datos en 102 columnas y 26050 renglones. En la parte superior de las columnas, se escribieron títulos de su

contenido como: nombre del material de referencia, nombre de la técnica utilizada en la medición para elementos mayores y traza, concentración en tanto por ciento (%) o partes por millón (ppm), referencias y comentarios.

1.2 Descripción breve de la base de datos

Para organizar los datos individuales de diferentes técnicas analíticas, se siguió la clasificación de los métodos analíticos en ocho grupos generales propuesta por Velasco-Tapia et al. (2001): (Grupo 1 – Gr1) Métodos químicos; (Grupo 2 – Gr2) Absorción atómica; (Grupo 3 – Gr3) Fluorescencia de rayos X; (Grupo 4 – Gr4) Espectrometría de emisión; (Grupo 5 – Gr5) Métodos nucleares; (Grupo 6 – Gr6) Espectrometría de masas; (Grupo 7 – Gr7) Métodos cromatográficos; y (Grupo 8 – Gr8) Miscelánea de métodos.

En la base de datos MRG, se registró la técnica analítica en función del grupo al que pertenecía de manera simplificada, es decir, se utilizaron siglas o grupos. Por ejemplo, si un elemento químico se había medido por absorción atómica se utilizó “AA”, o también “Gr2”; de igual manera, para la fluorescencia de rayos X se capturó “FRX” o “Gr3”.

Se compilaron 62254 datos de análisis de las concentraciones de 11 elementos mayores y 71 elementos traza (incluyendo lantánidos: REE), registrando además su respectiva *técnica analítica* utilizada en la medición del elemento y la *referencia* de la cual se compiló el dato.

El mayor número de datos en la base MRG corresponde a elementos mayores, debido a que por su mayor concentración, existe una amplia variedad de métodos analíticos para su medición (p. ej., métodos químicos, absorción atómica y FRX). Por otra parte, los elementos traza están presentes en cantidades pequeñas por lo que los métodos de medición son más sofisticados y costosos (p. ej., espectrometría de masas y métodos nucleares).

CAPÍTULO 2

EVALUACIÓN DE 15 PRUEBAS DE DISCORDANCIA CON 33 VARIANTES

Introducción

En este capítulo, se presentan los resultados del proceso de evaluación de 15 pruebas de discordancia con 33 variantes, para lo cual se realizó un estudio comparativo de 9 pruebas de discordancia sencilla con 13 variantes y 7 pruebas múltiples con 20 variantes (la prueba N4 pertenece a ambos tipos). Mayor detalles sobre estas pruebas se encuentran en los Apéndices 1; Verma et al., 2008 y 2. Este análisis se realizó en una extensa base de datos de 35 MRG que fue depurada estadísticamente para determinar los parámetros de criterio de eficiencia relativa (“Relative efficiency criterion” REC) y el criterio de datos desviados relativo (“Relative outlier criterion” ROC), que permiten evaluar la eficiencia global de cada una de las pruebas de discordancia.

2.1 Programa de cómputo utilizado para la evaluación

La base de datos MRG fue procesada mediante el software DODESYS (Discordant Outlier Detection and Elimination System), que es un software especializado para la identificación y en su caso la depuración de datos desviados mediante la aplicación de pruebas de discordancia a muestras estadísticas univariadas (Verma y Díaz González, no-publicado). Un dato llamado “outlier” es aquel que está ubicado en cualquier de los extremos de un arreglo ordenado de los datos u observaciones. Este dato puede o no ser discordante de acuerdo a las pruebas de discordancia (Barnett y Lewis, 1994; Verma, 2005). Cuando éste se detecta como discordante, se le denomina dato discordante o desviado (“discordant outlier”; Barnett y Lewis, 1994; Verma, 2009). De lo contrario, este dato será simplemente un “outlier”.

El software DODESYS tiene la novedad de que se basa en valores críticos nuevos más precisos y exactos simulados recientemente por Verma y Quiroz-Ruiz (2006a, 2006b, 2008) y Verma et al. (2008). DODESYS inicialmente lee un archivo de entrada en formato Excel y realiza una validación de los datos, si existen errores en los datos, se presentan en un cuadro de dialogo y deberán ser corregidos manualmente. En caso de que el archivo no presente errores se activa el menú “Statistical análisis” y aparecen tres opciones de aplicación de las pruebas de discordancia (i) “*Default process*”, esta opción se encarga de aplicar todas las pruebas de discordancia (33 variantes), organizando la aplicación dependiendo del tamaño de muestra; (ii) “*Single-outlier test*”, esta opción aplica las pruebas de discordancia sencillas (13 variantes) y son pruebas que evalúan un solo dato por iteración; y (iii) “*Multiple-outlier test*”, esta opción permite aplicar las pruebas de discordancia múltiples (20 variantes) que evalúan dos o más valores discordantes en una iteración.

Es importante señalar que la aplicabilidad de una prueba de discordancia depende de que la muestra estadística univariada cumpla con el número mínimo de datos que requiere la prueba.

Una vez procesada la base de datos, se generan, dos tipos de reportes “Report Data Summary” y “Report Test Summary”. El primero contiene el nombre de la muestra univariada, tamaño inicial de la muestra, media aritmética inicial, desviación estándar inicial, número de datos desviados identificados y el porcentaje de datos desviados. Con respecto a la muestra depurada, se reporta en este mismo archivo el número de datos final, valor mínimo y máximo, la media, la desviación estándar, los límites de confianza y la desviación estándar relativa. El segundo reporte, además del nombre de la muestra univariada, grupo y número de datos inicial, contiene información sobre la eficiencia de la metodología estadística, esto es, el número de pruebas de discordancia aplicadas y el nombre de las pruebas que fueron eficientes en la detección de datos desviados.

Por otra parte, DODESYS proporciona de manera adicional una carpeta “performance” de cada uno de los 35 MRG. En las carpetas se tienen archivos de 11 elementos

mayores y 71 elementos en función de los grupos analíticos, con la siguiente información: (i) el número de datos de la muestra estadística o “caso”, concepto definido en el Apéndice 1 (Verma et al., 2009); (ii) las pruebas que detectaron datos discordantes en elementos mayores y traza; (iii) las vueltas o revisiones que realizó el software, hasta no detectar y depurar datos desviados que reciben el nombre de “iteraciones”; y finalmente (iv) el número de datos desviados en la iteración y los valores de los datos desviados.

2.2 Criterio de eficiencia relativa (REC) de las pruebas de discordancia

El REC es un parámetro estadístico que expresa la eficiencia que tiene una prueba de discordancia para detectar y eliminar datos desviados en una muestra estadística. REC puede ser calculado por cada iteración o bien de manera global $(REC)_g$. Ver los Apéndices 1 y 2 para mayores detalles sobre la definición de REC para cada iteración $(REC)_1, (REC)_2$, etc. En realidad, la expresión matemática reporta el porcentaje de eficiencia que tiene la prueba para detectar datos desviados.

La metodología para evaluar REC consta de: (i) compilación de una extensa base de datos de MRG; (ii) el procesamiento de la base de datos con el software DODESYS, utilizando la opción *Default process* que aplica todas las pruebas de discordancia de manera continua; (iii) programar la información de los archivos *Performance*; para evaluar REC; y finalmente (iv) determinar la eficiencia de cada una de las pruebas de discordancia $(REC)_g$. Ejemplos detallados del cálculo del parámetro REC se presenta en la Tabla A1 del Apéndice 2 (González-Ramírez et al., 2009).

2.3 Criterio relativo global de datos desviados $(ROC)_g$ de las pruebas de discordancia

Una vez determinado el $(REC)_g$ fue importante contar con un parámetro estadístico que tome en cuenta tanto el número de datos desviados en casos denominados “exitosos” como el

total de los datos en los casos “aplicables” en la muestra estadística total. Por ello se formuló un nuevo criterio llamado criterio relativo global de datos desviados $(ROC)_g$. De esta manera, el $(ROC)_g$ fue introducido, por primera vez en el presente trabajo (Apéndice 2; González-Ramírez et al., 2009), con el fin de distinguir entre las eficiencias de las pruebas basadas en un gran número de muestras (el parámetro REC) y el porcentaje de valores discordantes detectados en una determinada muestra (el parámetro %Ot, ver el Apéndice 1; Verma et al., 2009).

Para la aplicación de $(ROC)_g$, se formaron cuatro categorías de pruebas (Tabla 1 del Apéndice 2; González-Ramírez et al., 2009). Por lo tanto, es posible conocer el número de datos iniciales y finales de una muestra estadística univariada. Para tener una mayor comprensión de la evaluación de ROC global $(ROC)_g$ ver la expresión matemática y el ejemplo que se muestra en el Apéndice 2 (González-Ramírez et al., 2009).

2.4 Resultados y análisis de la evaluación estadística de 15 pruebas de discordancia con 33 variantes

Las pruebas de discordancia fueron clasificadas en 14 categorías (Tabla 1 del Apéndice 2; González-Ramírez et al., 2009) para su evaluación, mediante la comparación de los valores de REC. El parámetro REC fue calculado para todas las categorías. Los resultados de estos cálculos se presentan en las Figuras A1-A11 en el suplemento electrónico del Apéndice 2 (González-Ramírez et al., 2009) y Figuras 1-3 del Apéndice 2 (González-Ramírez et al., 2009).

Es importante mencionar que la nomenclatura utilizada para las diferentes versiones de las pruebas de discordancia es la misma que en el Apéndice 2 (González-Ramírez et al., 2009). Como ejemplo en estos resultados para pruebas sencillas, p. ej., para la prueba N4, se dice N4U1 o N4L1 cuando se refiere a la versión “upper” o “lower”, respectivamente. Las pruebas múltiples se especifican de manera similar, p. ej., para la prueba N4 se tienen 6 versiones N4U2, N4L2, N4U3, N4L3, N4U4 y N4L4, dependiendo del número de datos que se están evaluando, siendo en este ejemplo 2, 2, 3, 3, 4 y 4, respectivamente.

Los resultados de la evaluación empírica de las 15 pruebas de discordancia con 33 variantes se obtuvieron mediante el parámetro $(REC)_g$. Las categoría 5 contiene la evaluación de todas las pruebas de discordancia sencilla y la categoría 13 las pruebas de discordancia múltiple. Finalmente, la categoría 14 está formada por ambas categorías (ver Apéndice 2; González-Ramírez et al., 2009 para mayores detalles).

Las pruebas sencillas se caracterizan por evaluar un sólo dato discordante: N1, N4, N9 y N10 que presentan dos versiones de evaluación - dato más alto (U) y más bajo (L); N2 (extremo); N7 (más alto); N8 (extremo); N14 (extremo, coeficiente de asimetría “*skewness*”); y N15 (extremo, coeficiente de exceso “*kurtosis*”). En la evaluación de $(REC)_g$ en todas las pruebas de discordancia, se contaron todos los casos “exitosos” en todas las iteraciones de una prueba determinada (incluyendo ambas versiones, más alto y más bajo). Se consideraron como casos “aplicables” el número inicial de casos (2220 casos con $n \geq 9$ datos cada uno).

Los resultados del parámetro REC global de la eficiencia de las pruebas de discordancia sencillas tiene el orden siguiente: $N15 > N4 \approx N1 > N2 \approx N14 > N10 \approx N9 > N8 > N7$.

La categoría 13 integrada por las pruebas de discordancia múltiples se caracterizó por evaluar de dos hasta cuatro datos a la vez, $k=2-4$. La evaluación de estas pruebas se realizó de la misma manera que en el caso de las pruebas sencillas. Para la comparación de las pruebas de esta categoría, se requirió el cálculo del parámetro de eficiencia relativa global $(REC)_g$ (ecuación 1 del Apéndice 1; Verma et al., 2009). Los valores de $(REC)_g$ de las pruebas múltiples (Figura 2 del Apéndice 2 ; González-Ramírez et al., 2009) mostraron que si comparamos las pruebas de discordancia múltiples para un determinado valor de k , se puede inferir el siguiente orden en la eficiencia relativa $(REC)_g$ de las pruebas: (i) pruebas que evalúan dos datos a la vez ($k = 2$): $N4U2+N4L2 > N13U2+N13L2 > N12U2+N12L2 > N11U2+N11L2 > N5UL > N3U2+N3L2 > N6UL$; (ii) pruebas que evalúan tres valores ($k = 3$): $N4U3+N4L3 > N3U3+N3L3$; y (iii) pruebas que evalúan cuatro valores ($k = 4$): $N4U4+N4L4 > N3U4+N3L4$.

Para la categoría 14 que incluye pruebas de discordancia sencillas y múltiples (Figura 3 del Apéndice 2; González-Ramírez et al., 2009), la secuencia de las eficiencias $(REC)_g$ fue la siguiente: $N15 > N4U+N4L \approx N1U+N1L \approx N4U2+N4L2 > N2 \approx N14 \approx N13U2+N13L2 \approx N12U2+N12L2 > N4U3+N4L3 \approx N11U2+N11L2 > N5UL \approx N4U4+N4L4 \approx N10U+N10L \approx N9U+N9L > N3U2+N3L2 > N8 > N7 > N6UL > N3U3+N3L3 > N3U4+N3L4$.

Una vez realizada la interpretación del parámetro REC, se procedió la determinación del parámetro $(ROC)_g$ (ecuación 2 del Apéndice 2; González-Ramírez et al., 2009). Este parámetro para una prueba se obtiene de dividir el número de datos desviados detectados por la prueba entre el número total de datos que contienen los 2220 casos “aplicables”, que corresponden a 41,821 datos geoquímicos individuales. Debido a que $(ROC)_g$ considera la relación de datos desviados entre el total de datos, es un parámetro totalmente libre de unidad, por lo que en una categoría se tienen pruebas con distintos k (1-4). Se formaron cuatro categorías de pruebas (Tabla 1 del Apéndice 2; González-Ramírez et al., 2009).

El resultado de evaluar $(ROC)_g$ para todas las pruebas de discordancia tiene el orden siguiente: $N4U4+N4L4 > N4U3+N4L3 > N4U2+N4L2 > N13U2+N13L2 > N11U2+N11L2 > N5UL > N3U2+N3L2 > N12U2+N12L2 > N15 > N3U3+N3L3 \approx N4U+N4L \approx N1U+N1L > N2 \approx N14 > N6UL > N3U4+N3L4 > N10U2+N10L2 > N9$.

Es importante hacer notar que el orden de $(REC)_g$ comparado con el de $(ROC)_g$ existen notables diferencias, p. ej., la prueba N15 que es la prueba que resultó con mayor eficiencia relativa en $(REC)_g$ mientras que con $(ROC)_g$ se encuentra aproximadamente después de la mitad del orden indicado. Esto significa que aun cuando la prueba N15 es una prueba muy eficiente, el número de datos desviados que detecta y depura no se compara con el número de datos desviados que elimina la prueba múltiple $N4U4+N4L4$.

CAPÍTULO 3

PRUEBAS DE SIGNIFICANCIA

Introducción

En este capítulo se abordarán las pruebas de significancia: *F*-Fisher, *t*-Student, y prueba de ANOVA, que son de gran importancia en el manejo de datos experimentales. Los libros de Anderson (1987), Gilbert (1987), Ebdon (1988), Taylor (1990), Gibbons (1994) Jensen et al. (2000), Miller y Miller (2000) contienen aspectos importantes de estas pruebas estadísticas, los cuales han sido resumidos por Verma (2002, 2005).

Como parte de la aplicación estadística a la base de datos, se realizó la evaluación de las pruebas de discordancia que consistió en llevar a cabo regresiones lineales al parámetro REC y el tamaño de grupos. Se empleó el programa OYNYL (Verma et al. 2006) para obtener los parámetros de las regresiones (pendiente e intercepto así como sus incertidumbres). Posteriormente se aplicaron las pruebas *F*-Fisher, *t*-Student (Jensen et al. 2000; Verma, 2005, 2009) y ANOVA a los datos de las pendientes de las regresiones.

3.1 Aplicación de las pruebas de *F*-Fisher, *t*-Student

En todas las ciencias e ingenierías se trabaja generalmente, con muestras estadísticas, dado que nunca se tiene acceso directamente a la población original. La población original es considerada teóricamente Gaussiana o normal para todos los datos experimentales. De esta afirmación, entonces, se plantearía la interrogante: “¿Si dos o más muestras estadísticas provienen de la misma población o de poblaciones parecidas?”. Para ello, se tendría que investigar: (1) si las muestras estadísticas tienen varianzas iguales o parecidas; y (2) si las

muestras estadísticas tienen medias iguales o parecidas. Para responder a estas interrogantes existen pruebas estadísticas como la F -Fisher, t -Student.

En la Tabla 2 del Apéndice 2 (González-Ramírez et al., 2009), para analizar el efecto que tiene el tamaño de la muestra estadística en el criterio de eficiencia relativa de la primera iteración $(REC)_1$ se combinó la regresión lineal con las pruebas de significancia. En la prueba F -Fisher se compararon las varianzas en las eficiencias de las pruebas, partiendo de una hipótesis nula (H_0) y otra alterna (H_1).

Como ejemplo en la tabla 3.1, presentamos la evaluación de las pruebas N1L y N10L (datos tomados de la Tabla 2 del Apéndice 2; González-Ramírez et al., 2009). Para la aplicación de la prueba F , la hipótesis H_0 es: “La pruebas N1L y N10L tienen las mismas varianzas de las eficiencias relativas” y la hipótesis alternativa H_1 es: “Las pruebas N1L y N10L tienen diferentes varianzas de las eficiencias relativas”

Tabla 3.1 Datos para aplicar el estadístico de prueba de F -Fisher (tomados de la Tabla 2 del Apéndice 2; González-Ramírez et al., 2009).

Prueba	n	$\nu_1(n_x - 1)$	$\nu_2(n_y - 1)$	b	s_b
N1L	18	17	---	0.815	0.207
N10L	18	---	17	0.787	0.181

$$F_{N1L-N10L} = \frac{(0.207^2)_{N1L}}{(0.181^2)_{N10L}} = 1.14412919 \ 6$$

(La fórmula del estadístico fue tomada de Verma, 2002, 2005).

Para grados de libertad: $\nu_{1N1L} = 17$ y $\nu_{2N10L} = 17$ obtenemos un valor crítico de F entre 3.31 y 3.16 a un nivel de confianza de 99%, dado que el valor calculado es significativamente menor a cualquier de estos dos valores, no hace falta estimar el valor exacto interpolado por las ecuaciones presentadas por Verma, 2009)

Como $1.144129196 < 3.31-3.16$, corroboramos la hipótesis H_0 y concluimos que las varianzas en la eficiencia relativa de las dos pruebas de discordancia (N1L y N10L) son estadísticamente iguales al 99% de nivel de confianza.

La determinación del estadístico de la prueba de la F -Fisher se realizó para toda la Tabla 2 del Apéndice 2 (González-Ramírez et al., 2009), llegando a conclusiones semejantes, es decir, no existen diferencias significativas entre las varianzas de las eficiencias relativas de las pruebas de discordancia.

Después de calcular la prueba F -Fisher, se aplicó la prueba t -Student (o la prueba t) que se utiliza para comparar dos medias, partiendo de que las pendientes (el parámetro b en la Tabla 3.1) tienen varianzas iguales, o sea que los valores de s_b (Tabla 3.1) no presentan diferencias significativas. Las pendientes de las regresiones representan las eficiencias relativas de las pruebas de discordancia. En la aplicación de la prueba t -Student la hipótesis nula (H_0) se planteó como: “Las pruebas N1L y N10L tienen las mismas eficiencias relativas” o que provienen de la misma población o poblaciones idénticas, mientras que la hipótesis alterna (H_1) indica “las pruebas N1L y N10L tienen diferentes eficiencias relativas”.

Para determinar el estadístico t -Student, primeramente se requiere determinar la desviación estándar combinada para N1L y N10L ($s_{N1L-N10L}$).

$$s_{N1L-N10L} = \sqrt{\frac{(18-2)(0.207)^2 + (18-2)(.181)^2}{(18+18-2)}} = 0.19427744 \text{ 7}$$

El siguiente paso es aplicar el estadístico para $t_{N1L-N10L}$

$$t_{N1L-N10L} = \frac{|0.815 - 0.787|}{S_{N1L-N10L} \left(\sqrt{\left(\frac{1}{18} + \frac{1}{18} \right)} \right)} = 0.427584371$$

(Las fórmula de los estadísticos aplicados se tomaron de Verma, 2002, 2005).

Para obtener los grados de libertad: $\nu=34$, obtenemos un valor crítico entre 2.750 y 2.724, a un nivel de confianza de 99%.

Como el valor del estadístico calculado es menor que el valor crítico, la hipótesis nula (H_0) se acepta y concluimos que para las pruebas de discordancia N1L y N10L no existe diferencia significativa entre sus eficiencias relativas. Por tanto, son estadísticamente iguales a un nivel de confianza del 99%.

Sin embargo, al aplicar la prueba t -Student en otras comparaciones de pruebas de discordancia, los resultados (Tabla 3.2) indican que si existe diferencia significativa entre las eficiencias relativas de las pruebas para los pares N1L-N9L y N4L-N9L. En estos casos, el valor crítico $\nu = 34$ para la t -Student está entre 2.750 y 2.724, a nivel de confianza de 99%. Para ambos pares de pruebas N1L-N9L y N4L-N9L, t calculado (2.765165575; Tabla 3.2) es mayor que el valor crítico. Entonces, la hipótesis nula H_0 se rechaza y la hipótesis H_1 se acepta, infiriendo que ambos pares de pruebas tienen diferentes eficiencias relativas.

Tabla 3.2 Prueba t -Student aplicada a los resultados de la aplicación de pruebas de discordancia que evalúan el valor más pequeño en un conjunto de datos organizado de manera ascendente

Prueba	n	b	s_b	Comparación de pruebas	s desviación estándar combinada	t de Student
N1L	18	0.815	0.207	N1L-N4L	0.207	0
N1L	18	0.815	0.207	N1L-N9L	0.203857594	2.765165575
N1L	18	0.815	0.207	N1L-N10L	0.19427744	0.427584371
N4L	18	0.815	0.207	N4L-N9L	0.203857594	2.765165575
N9L	18	0.627	0.201	N4L-N10L	0.19427744	0.427584371
N10L	18	0.787	0.181	N9L-N10L	0.191066606	2.515510215

Todas las combinaciones de pares posibles de la categoría 2 (Tabla 2 del Apéndice 2; González-Ramírez et al., 2009) presentados en la Tabla 3.2, fueron probadas con pruebas *F*-Fisher, *t*-Student. Al realizar las comparaciones entre los pares formados predomina la validez de la hipótesis nula (H_0) para cuatro pares (N1L-N4L, N1L-N10L, N4L-N10L y N9L-N10L). En estos casos, se concluye que para cuatro pares de las pruebas de discordancia de esa categoría no existe diferencia significativa entre sus eficiencias relativas, mientras que para dos pares (N1L-N9L y N4L-N9L) sí existe diferencia significativa a 99% de confianza. Todos los resultados de esta metodología se presentan en el Apéndice 2 (González-Ramírez et al., 2009).

3.2 Aplicación de la prueba de ANOVA

La prueba ANOVA para comparar tres o más muestras, se interpreta de manera similar que las pruebas *F*-Fisher y *t*-Student.. La hipótesis nula (H_0) se plantea como: “las muestras provienen de una misma población o una población común o poblaciones idénticas” y la hipótesis alterna (H_1) “las muestras provienen de poblaciones distintas”. La prueba de ANOVA se basa en encontrar si existe estadísticamente más variación entre las muestras que la variación dentro de ellas.

La prueba de significancia ANOVA fue aplicada a las pruebas de discordancia anteriormente descritas en la Tabla 3.2, bajo las siguientes premisas: para la hipótesis nula (H_0) “La pruebas N1L, N4L, N9L y N10L tienen las mismas eficiencias relativas” o que provienen de la misma población o poblaciones idénticas, mientras que la hipótesis alterna (H_1) es “las pruebas N1L, N4L, N9L y N10L tienen diferentes eficiencias relativas” o que no provienen de la misma población o poblaciones idénticas.

Para determinar ANOVA, se llevó a cabo el siguiente procedimiento:

- 1) Se estima la media global de la muestra.

$$\bar{x}_g = \frac{(18-1)(.815)^2 + (18-1)(.815)^2 + (18-1)(.627)^2 + (18-1)(.787)^2}{72} =$$

$$\bar{x}_g = 0.7607125$$

2) Calcular \hat{S}^2_B , la estimación de la varianza entre las muestras

$$\hat{S}^2_B = \frac{18(.815 - .7607125)^2 + 18(.815 - .7607125)^2 + 18(.627 - .7607125)^2 + 18(.787 - .7607125)^2}{4 - 1}$$

$$\hat{S}^2_B = 0.1467203$$

3) Calcular \hat{S}^2_W la varianza dentro de las muestras

$$\hat{S}^2_W = \frac{(18-1)(.207)^2 + (18-1)(.207)^2 + (18-1)(.201)^2 + (18-1)(.181)^2}{72 - 4}$$

$$\hat{S}^2_W = 0.03965082$$

4) Finalmente, se determina el valor del estadístico de la prueba de ANOVA

$$F = \frac{0.1467203}{0.03965082} = 3.700309232$$

(Las fórmulas de los estadísticos aplicados fueron tomados de Verma, 2002, 2005).

Para grados de libertad: $\nu_1 = 3$, $\nu_2 = 68$ obtenemos un valor crítico entre: 4.13 y 4.04 a un nivel de confianza del 99%. Como el valor del estadístico calculado (3.700309232) es menor que el valor crítico de 99% de confianza, la hipótesis H_0 se acepta, corroborando que no existe diferencia significativa entre las eficiencias relativas del conjunto de las cuatro pruebas de discordancia (N1L, N4L, N9L y N10L). En el caso de que la prueba ANOVA se aplicara a un nivel de confianza de 95% (con el valor crítico entre 2.76-2.72) y dado que el estadístico calculado (3.700309232) es mayor que el valor crítico, se concluiría que la hipótesis H_0 se rechaza y la alterna H_1 se acepta, corroborando que sí existe diferencia significativa entre las eficiencias relativas de las pruebas de discordancia a este nivel de confianza (95%).

CAPÍTULO 4

APLICACIÓN ESTADÍSTICA A LAS BASES DE DATOS MRG Y SU UTILIDAD EN LA CALIBRACIÓN DE FLUORESCENCIA DE RAYOS-X

Introducción

Además de evaluar las eficiencias relativas de las pruebas de discordancia (Capítulo 3; Apéndices 1 y 2), otro producto importante del procesado de la base de datos MRG (Capítulo 1) fue su utilidad en la calibración de fluorescencia de rayos-x (FRX) para los elementos mayores. Después de detectar y depurar los datos desviados en elementos mayores de cada una de las muestras de los 35 MRG por DODESYS, se aplicó la prueba de ANOVA seguido nuevamente por DODESYS. Cabe aclarar que para esta aplicación la versión de DODESYS no contemplaba la salida de un archivo de datos después de la depuración, tampoco de un archivo de valores de datos desviados. En este capítulo, presentaremos el procedimiento realizado para la obtención de las medidas de tendencia central y de dispersión de las concentraciones de elementos mayores en los MRG, útiles para realizar la calibración.

4.1 Procesamiento estadístico de las bases de datos

Una vez procesada la base de datos MRG para los elementos mayores se analizaron los reportes “Report Data Summary” y “Report Test Summary” para conocer: (i) el número de datos discordantes de cada uno de los 35 MRG; y (ii) las pruebas que resultaron eficientes para detectar datos desviados. Se procedió a abrir las carpetas “Performance” para depurar datos

desviados con la finalidad de lograr una distribución normal de los datos de concentraciones de elementos mayores.

El procedimiento consiste en los siguientes pasos:

- 1) Se eliminaron de manera manual los datos desviados de los 11 elementos mayores (que incluyen Fe^{+2} y Fe^{+3}) y Fe_2O_3 (T), en los 35 MRG; es decir, se revisaron en promedio 385 archivos de las carpetas “Performance”, centrando la atención en la información específica: (i) el nombre del material de referencia; (ii) nombre del elemento mayor que contenía datos desviados; y (iii) valores desviados que debían ser depurados.
- 2) Una vez depurados los datos discordantes en los archivos individuales, se construyeron bases de datos en Excel, para todos los elementos mayores de cada uno de los 35 MRG.
- 3) Se exportaron las bases de datos de Excel a Statistica y se aplicó la prueba ANOVA, bajo la hipótesis nula (H_0) “todas las muestras provienen de una misma población” y la hipótesis alterna (H_1) “las muestras provienen de poblaciones distintas”.

El proceso de aplicación de ANOVA se explica en la Figura 4.1 donde se presenta un diagrama simplificado de flujo. El resultado del estadístico llamado F (F ratio, ver p. 86 en Verma, 2005) se comparó con el valor crítico con el fin de determinar cuál de las dos hipótesis era válida. En caso de no encontrar válida la hipótesis nula (H_0), en otras palabras, cuando se encontraron diferencias significativas entre los grupos para un determinado elemento en un MRG, fue necesario determinar la causa de dicha discrepancia. Esto se hizo mediante el gráfico de las medias y la desviación estándar de los diferentes grupos de métodos analíticos a los cuales pertenecen los datos bajo estudio. El grupo o los grupos que se encontraron diferentes a los restos fueron excluidos de la muestra combinada de todos los grupos estadísticamente similares.

- 4) Finalmente, se formó un sólo archivo que contenía los resultados de los archivos de todos los elementos mayores después de haber aplicado ANOVA a los MRG y fueron procesada nuevamente por DODESYS.

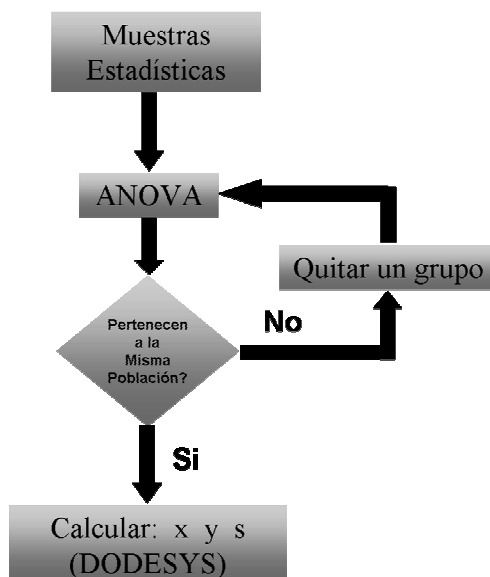


Figura 4.1 Diagrama de flujo de la metodología estadística aplicada a los elementos mayores de los MRG de la base de datos utilizada en el Apéndice 1 (Verma et al., 2009).

4.1.a) Procesamiento estadístico de la muestra W-2

Se muestra a continuación un ejemplo del procedimiento estadístico aplicado a la base MRG, para el material de referencia W-2, diabasa, procedente de U.S.A.

El primer paso fue trabajar con la primera parte del reporte proporcionado por DODESYS “Report Data Summary”, en el cual existe un reporte por cada material de referencia (Capítulo 2). En el ejemplo particular de W-2, el elemento de estudio será el Sílice (SiO_2).

En la Tabla 4.1 se presenta la primera parte de la información que proporciona DODESYS. En la parte superior contiene los títulos que identifican el contenido de cada una de las columnas: (i) Element corresponde a la fórmula del elemento en estudio; (ii) Group grupo de técnicas analíticas; (iii) nin número de datos de la muestra; (iv) xin la media de la muestra; (v) sin la desviación estándar de la muestra; y (vi) Ot (outliers) o datos desviados detectados. Para el

caso de SiO₂, la información estadística se presenta en la Tabla 4.1, en engrillado con el propósito de resaltar el grupo que presentó uno o más datos desviados.

Tabla 4.1 Resultados de la carpeta Report Data Summary de W-2.

Element	Group	nin	xin	sin	Ot
SiO ₂	1	7	52.45039	0.495556	0
SiO ₂	2	2	51.85749	0.786624	
SiO ₂	3	14	52.43511	0.94026	1
SiO ₂	4	5	53.05124	1.447605	0
SiO ₂	5	3	42.15917	15.1691	0

Un aspecto importante de este reporte es que informa en que grupos de la muestra estadística se detectaron datos desviados (outliers). El cero en Ot significa que no se detectaron datos desviados, mientras que el espacio vacío indica que no se tiene el número suficiente de datos para aplicar pruebas de discordancia. Para la aplicación de al menos una prueba de discordancia, se requiere que $n \geq 3$ datos (Tabla 1 del Apéndice 1; Verma et al., 2009). En el informe se reporta que la muestra W-2 únicamente detectó un dato desviado, que debe ser depurado previamente a la aplicación de la prueba ANOVA.

El siguiente paso fue abrir el reporte “Report Test Data Summary”. En este se indica el resultado de las pruebas que detectaron datos desviados o casos exitosos (Apéndice 1; Verma et al., 2009).

En la Tabla 4.2 los encabezados en la parte superior indican el contenido de cada una de las columnas: (i), Element corresponde a la fórmula del elemento en estudio; (ii) Gr grupo de técnicas analíticas; (iii) nin número de datos de la muestra; (iv) #Tests número de pruebas aplicadas; y (v) Tests successful in detecting outlier(s), proporciona el informe de las pruebas exitosas en la detección de datos desviados.

Tabla 4.2 Resultados de la carpeta Report Test Data Summary de W-2.

Element	Gr	nin	# Test	Tests successful in detecting outlier(s)
SiO ₂	1	7	0	
SiO ₂	2	2		
SiO ₂	3	14	13	N1L, N2, N4L1, N9L, N10L, N15.
SiO ₂	4	5	0	
SiO ₂	5	3	0	

Como se observa en el Tabla 4.3, para el elemento sílice SiO₂ en la muestra W-2 se detectó en el grupo 3 un valor desviado en una muestra de 14 datos. Se aplicaron 13 pruebas de discordancia y las pruebas que evalúan el valor más bajo (Lower; ver Tabla 1 del Apéndice 2; González-Ramírez et al., 2009) así como el valor extremo son las que detectaron el mismo valor desviado.

Tabla 4.3 Resultados de la carpeta Performance de W-2.

W - 2 Diabasa					
Gr = 3; n = 14					
N1 Lower			N9 Lower		
Iteration	Id Outliers	Valor	Iteration	Id Outliers	Valor
1	1	49.65397	1	1	49.65397
N2 Extreme			N10 Lower		
Iteration	Id Outliers	Valor	Iteration	Id Outliers	Valor
1	1	49.65397	1	1	49.65397
N4 (k=1) Lower			N15 Extreme		
Iteration	Id Outliers	Valor	Iteration	Id Outliers	Valor
1	1	49.65397	1	1	49.65397

En los resultados de la Tabla 4.3 se observa que la muestra W-2, el elemento SiO₂ en el grupo 3 detectó el valor desviados (49.65397). La depuración del dato desviado es Id Outliers=1, indica el lugar que ocupa el dato discordante en el ordenamiento de la muestra de datos, mostrado en la Tabla 4.4.

Tabla 4.4 Ubicación del dato desviado de W-2.

CONS.	NAME	Major MET	SiO ₂
1	W-2	XRF	49.653971
2	W-2	XRF	51.5579794
3	W-2	XRF	51.9216664
4	W-2	XRF	52.4778935
5	W-2	XRF	52.4992869
6	W-2	XRF	52.5634669
7	W-2	WXRF	52.6490404
8	W-2	WXRF	52.6704337
9	W-2	XRF	52.7560071
10	W-2	XRF	52.8415805
11	W-2	XRF	52.8415805
12	W-2	XRF	52.8629739
13	W-2	XRF	53.1410875
14	W-2	XRF	53.6545279

Después de eliminar el valor desviado se obtiene una media y desviación estándar diferente a la inicial en el grupo 3 en que se detectó y se depuró el dato desviado, como se observa en la tabla 4.5

Tabla 4.5 Muestra depurada de datos discordantes de W-2.

Group	Nout	x	s
1	7	52.45039	0.495556
2	2	51.85749	0.786624
3	13	52.64904	0.51344
4	5	53.05124	1.447605
5	3	42.15917	15.1691

Una vez depurado el dato desviado, se aplicó la prueba de ANOVA a los datos restantes mostrados en la Tabla 4.5, bajo la hipótesis nula (H_0) “las muestras (grupos) provienen de una

misma población” y como alterna (H1)”las muestras (grupos) provienen de poblaciones distintas”.

a) Se estimó la media global de la muestra.

$$\bar{x}_g = \frac{(7)(52.45039) + (2)(51.85749) + (13)(52.64904) + (5)(53.05124) + (3)(42.15917)}{30}$$

$$\bar{x}_g = 51.5679647$$

b) Se determinó \hat{S}^2_B , la estimación de la varianza entre las muestras

$$\hat{S}^2_B = \frac{7(52.45039 - 51.5679647)^2 + 2(51.85749 - 51.5679647)^2 + 13(52.64904 - 51.5679647)^2 + 5(53.05124 - 51.5679647)^2 + 3(42.15917 - 51.5679647)^2}{5-1}$$

$$\hat{S}^2_B = 74.3471403$$

c) Se determinó \hat{S}^2_W la varianza dentro de las muestras

$$\hat{S}^2_W = \frac{(7-1)(0.495556)^2 + (2-1)(0.786624)^2 + (13-1)(0.51344)^2 + (5-1)(1.447605)^2 + (7-1)(15.1691)^2}{30-5}$$

$$\hat{S}^2_W = 18.9536444$$

d) Finalmente se determinó el valor de la prueba de ANOVA

$$F = \frac{74.3471403}{18.9536444} = 3.92257757$$

Para grados de libertad: $\nu_1 = 4$; $\nu_2 = 25$ obtenemos un valor crítico de: 4.18 a un nivel de confianza del 99%. Como el valor del estadístico calculado es menor que el valor crítico, la hipótesis H_0 se Acepta, esto significa que las muestras o los grupos provienen de la misma población

4.1b) Procesamiento estadístico de la muestra JA-1

La muestra JA-1 corresponde al material de referencia andesita procedente de Japón. En esta muestra fueron detectados más de un dato desviado (outlier), siendo la razón por la cual se presenta este ejemplo.

1.- Conocer el reporte proporcionado por ODESYS “Report Data Summary” (Tabla 4.6).

Tabla 4.6 Resultados de la carpeta Report Data Summary de JA-1.

Element	Group	nin	xin	sin	Ot
SIO₂	1	21	63.90905	0.423838	1
SIO ₂	2	9	63.49857	1.255471	0
SIO₂	3	39	64.22733	0.61895	2
SIO ₂	4	6	63.98	0.551906	0
SIO ₂	5*	4	62.975	1.15	0
SIO ₂	6	2	63.89296	0.272891	
SIO ₂	8	1	64.06		

2.- El siguiente paso fue abrir el reporte “Report Test Data Summary” (Tabla 4.7), para conocer las pruebas de discordancia que habían detectado datos desviados.

Tabla 4.7 Resultados de la carpeta Report Test Data Summary de JA-1.

Element	Gr	nin	# Test	Tests successful in detecting outlier(s)
SIO ₂	1	21	13	N7U, N8, N9U, N10U.
SIO ₂	2	9	13	
SIO ₂	3	39	13	N1U, N2, N4U1, N7U, N8, N9U, N10U, N15.
SIO ₂	4	6	0	
SIO ₂	5	4	0	
SIO ₂	6	2		
SIO ₂	8	1		

3.- Posteriormente, de acuerdo con la carpeta “Performance” (Tabla 4.8) se observan los datos desviados para ser depurados. En el Grupo 1, el valor desviado 65.12, detectado por pruebas que evalúan al valor más alto y en el Grupo 3 los valores 66.66 y 62.3, detectados por pruebas que evalúan al valor mas alto y a los extremos.

Tabla 4.8 Resultados de la carpeta “Performance” de JA-1.

JA - 1 Andesita					
Gr = 1; n = 21			Gr = 3; n = 39		
N7 Upper			N1 Upper		
	Id			Id	
Iteration	Outliers	Valor	Iteration	Outliers	Valor
1	21	65.12	1	39	66.66
N8 Extreme			N2 Extreme		
	Id			Id	
Iteration	Outliers	Valor	Iteration	Outliers	Valor
1	21	65.12	1	39	66.66
N9 Upper			N15 Extreme		
	Id			Id	
Iteration	Outliers	Valor	Iteration	Outliers	Valor
1	21	65.12	2	1	62.3
N10 Upper			N15 Extreme		
	Id			Id	
Iteration	Outliers	Valor	Iteration	Outliers	Valor
1	21	65.12	1	39	66.66
N10 Upper			N15 Extreme		
	Id			Id	
Iteration	Outliers	Valor	Iteration	Outliers	Valor
1	21	65.12	2	1	62.3

Después de depurar los datos desviados, se aplicó la prueba de ANOVA a las muestras o grupos para el elemento SiO₂ bajo la hipótesis nula (H₀) “Las muestras (grupos) provienen de la misma población” y como alterna (H₁) “Las muestras (grupos) provienen de poblaciones distintas”.

En la Tabla 4.9 se observan los parámetros estadísticos después de haber depurado los datos discordantes de los grupos 1 y 3. Por otra parte, se optó por no considerar los grupos 6 y 8 dado que estos no cuentan con un mínimo tres datos para que puedan ser aplicadas las pruebas de discordancia.

Tabla 4.9 Muestra depurada de datos discordantes de JA-1

Group	Nout	x	s
1	20	63.8485	0.328718
2	9	63.498569	1.25547139
3	37	64.213673	0.36962427
4	6	63.98	0.55190579
5	4	62.975	1.15

Determinación de ANOVA:

a) Se estimó la media global de la muestra.

$$\bar{x}_g = \frac{(20-1)(63.8485)^2 + (9-1)(63.498569)^2 + (37-1)(64.213673)^2 + (6-1)(63.98)^2 + (4-1)(62.975)^2}{76}$$

$$\bar{x}_g = 63.9492503$$

b) Se determinó \hat{S}_B^2 , la estimación de la varianza entre las muestras

$$\hat{S}_B^2 = \frac{20(63.8485 - 63.9492503)^2 + 9(63.498569 - 63.9492503)^2 + 37(64.213673 - 63.9492503)^2 + 6(63.98 - 63.9492503)^2 + 4(62.975 - 63.9492503)^2}{5-1}$$

$$\hat{S}_B^2 = 2.10509476$$

c) Se determinó \hat{S}_W^2 la varianza dentro de las muestras

$$\hat{S}_W^2 = \frac{(20-1)(0.328718)^2 + (9-1)(1.25547139)^2 + (37-1)(0.36962427)^2 + (6-1)(0.55190579)^2 + (4-1)(1.15)^2}{76-5}$$

$$\hat{S}_W^2 = 0.353121380$$

d) Finalmente se determinó el valor de la prueba de ANOVA

$$F = \frac{2.10509476}{0.353121380} = 5.96139142$$

Para grados de libertad: $\nu_1 = 4$, $\nu_2 = 71$ obtenemos un valor crítico entre: 3.65 y 3.56 a un nivel de confianza del 99%. Como el valor del estadístico calculado es seguramente mayor que el valor crítico esperado de la interpolación (Verma, 2009), la hipótesis H_0 se rechaza, esto significa que existen diferencias significativas entre las muestras, por lo que se acepta la hipótesis alterna H_1 .

4.- Se procedió a graficar los datos depurados a fin de conocer cuál podía ser el grupo discordante. En la gráfica (Figura 4.2), se observa que el grupo 5 formado de 4 datos, presenta la media más pequeña con una alta desviación estándar más grande que el resto de los grupos.

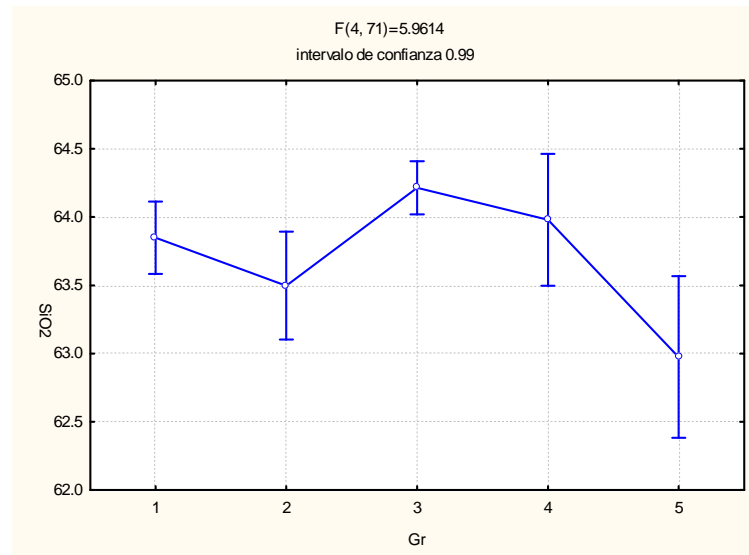


Figura 4.2 Gráfica de la muestra de andesita JA-1 después de depurar los datos desviados.

Debido a las diferencias cualitativas señaladas arriba en el grupo 5, se procedió aplicar nuevamente la prueba ANOVA a los otros grupos restantes (grupos 1-4) bajo la suposición de que el grupo 5 viene de otra población.

5.- Depurado provisionalmente el grupo 5 de la muestra de SiO₂, nuevamente se determinó la prueba de ANOVA y se comparó el valor del estadístico con el valor crítico.

$$\bar{x}_g = \frac{(20-1)(63.8485)^2 + (9-1)(63.498569)^2 + (37-1)(64.213673)^2 + (6-1)(63.98)^2}{72}$$

$$\bar{x}_g = 64.0033753$$

$$\hat{S}_B^2 = \frac{20(63.8485 - 64.0033753)^2 + 9(63.498569 - 64.0033753)^2 + 37(64.213673 - 64.0033753)^2 + 6(63.98 - 64.0033753)^2}{4-1}$$

$$\hat{S}_B^2 = 1.47093314$$

$$S^2_w = \frac{(20-1)(0.328718)^2 + (9-1)(1.25547139)^2 + (37-1)(0.36962427)^2 + (6-1)(0.55190579)^2}{72-4}$$

$$\hat{S}^2_w = 0.31498684$$

$$F = \frac{1.47093314}{0.31498684} = 4.66982418$$

ANOVA: F=4.6698241 valor crítico $\nu_1 = 3$, $\nu_2 = 68$ entre 4.13 y 4.04

Como se observa gráficamente en la Figura 4.3, el resultado del estadístico (4.6698241) es mayor que el valor crítico que estaría entre 4.13 y 4.04. Por lo tanto, el grupo 5 no resuelve el problema de diferencias significativa en el resto de los grupos. Por esta razón, el grupo 5 debe ser reincorporado en el conjunto de datos para seguir investigando la causa de las diferencias.

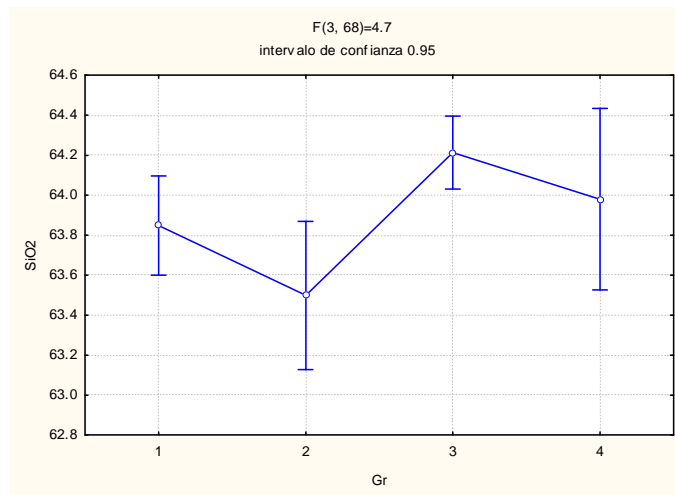


Figura 4.3 Gráfica de la muestra de JA-1 con el grupo 5 depurado.

6.- El siguiente paso fue incorporar el grupo 5 y depurar provisionalmente el grupo 3, en razón que es un grupo que presenta su media un poco mayor que los otros cuatro grupos.

$$\bar{x}_g = \frac{(20-1)(63.8485)^2 + (9-1)(63.498569)^2 + (6-1)(63.98)^2 + (4-1)(64.213673)^2}{39}$$

$$\bar{x}_g = 63.6983877$$

$$\hat{S}_B^2 = \frac{20(63.8485 - 63.6983877)^2 + 9(63.498569 - 63.6983877)^2 + 6(63.98 - 63.6983877)^2 + 4(64.213673 - 63.6983877)^2}{4 - 1}$$

$$\hat{S}_B^2 = 1.1263379$$

$$\hat{S}_W^2 = \frac{(20 - 1)(0.328718)^2 + (9 - 1)(1.25547139)^2 + (6 - 1)(0.55190579)^2 + (4 - 1)(1.15)^2}{39 - 4}$$

$$\hat{S}_W^2 = 0.57580635$$

$$F = \frac{1.1263379}{0.57580635} = 1.95610537$$

Para grados de libertad: $\nu_1 = 3$, $\nu_2 = 35$ obtenemos un valor crítico de: 4.4 a un nivel de confianza del 99%. Como el valor del estadístico calculado es menor que el valor crítico, la hipótesis H_0 se acepta, esto significa que no existen diferencias significativas entre las muestras, dado que, según la aplicación de ANOVA, todos los grupos, menos el grupo 3, provienen de la misma población. Por lo tanto, en este caso específico, los datos del grupo 3, siendo bastante numerosos (37 datos) y coherente (desviación estándar relativamente más pequeña que cualquier otro grupo; Figura 4.3), deberán ser tratados en forma separada, con el fin de reportar las estadísticas basadas solamente en este grupo. Por otra parte, el resto de los grupos se pueden combinar para volver a tratar por DODESYS y reportar las estadísticas finales.

El procedimiento descrito anteriormente se realizó para los elementos mayores de los 35 MRG, lo que representó una tarea laboriosa en esta fase del trabajo. Finalmente se aplicaron nuevamente las pruebas de discordancia mediante el software DODESYS.

Actualmente DODESYS proporciona la salida de un archivo de datos después de la depuración y otro de valores de datos desviados lo que minimiza esta tarea.

4.2 Calibración del equipo de Fluorescencia de rayos-x (FRX)

En la Tabla 4.10, se presentan las concentraciones de los elementos mayores para dos MRG procedentes de Japón. Estos resultados se obtuvieron después del tratamiento de

depuración de datos desviados, seguido de la aplicación de ANOVA y el reprocesamiento de los MRG con DODESYS, y fueron usados para la calibración descrita en el Apéndice 1 (Verma et al., 2009).

La información geoquímica referente a la calibración obtenida con regresión lineal ordinaria y regresión lineal ponderada constituye datos no-publicados de Guevara y Verma, y por lo tanto, no se presentan en la presente tesis.

Tabla 4.10 Concentraciones de elementos mayores para las andesitas JA-1 y JA-2, procedentes de Japón.

Elemento	JA-1			JA-2		
	n	x	S	n	x	S
SiO ₂	74	64.0450247	0.45257924	47	56.6105638	0.59961642
TiO ₂	79	0.85780045	0.03339527	46	0.67562273	0.01378847
Al ₂ O ₃	87	15.2105663	0.33876721	47	15.5080104	0.30641635
Fe ₂ O ₃ T	99	7.05728806	0.24457109	64	6.25101123	0.15592941
MnO	92	0.1570844	0.00992077	60	0.10743375	0.00613705
MgO	83	1.5743799	0.08277909	51	7.60630317	0.33585137
CaO	85	5.70503079	0.15898859	54	6.31269016	0.16668002
Na ₂ O	88	3.87063908	0.13858288	65	3.13649135	0.12424727
K ₂ O	84	0.78427877	0.04765096	65	1.81099654	0.05209899
P ₂ O ₅	76	0.16343665	0.01999043	43	0.14969583	0.01499951

Los resultados de la calibración del equipo de FRX con elementos mayores se encuentran en el Apéndice 1 (Verma et al., 2009) donde se observa una mejoría por la aplicación de regresión lineal ponderada.

CAPITULO 5

COMPARACIÓN DE MÉTODOS DE PREPARACIÓN DE PASTILLAS PRENSADAS Y DISCOS FUNDIDOS EN FLUORESCENCIA DE RAYOS-X (FRX).

Introducción

El método de medición por espectrometría de rayos-X es una de las técnicas más utilizadas tanto para elementos mayores como de elementos traza, para lo cual existen dos métodos de preparación de muestras (Verma et al. 1992). En este capítulo, se describe brevemente los resultados estadísticos obtenidos al comparar cuantitativamente los métodos de preparación de pastillas prensadas y discos fundidos, medidos por la técnica de FRX. Para ello, se utilizaron 14 materiales de referencia procedentes de Japón, Sudáfrica y Francia, así como también muestras de rocas ígneas (85) y sedimentarias (3) de México. Para mayores detalles ver el Apéndice 3 (Verma et al., 2009a).

5.1 Base de datos y su procesamiento

Este estudio utilizó información contenida en la base de datos MRG (Capítulo 1). Los datos fueron previamente procesados con el software DODESYS con la finalidad de detectar y eliminar datos desviados y obtener de manera confiable las medidas de tendencia central y dispersión de los elementos cromo (Cr) y níquel (Ni). Se procesaron 10 MRG de nuestra base de datos procedentes de Japón y Sudáfrica. Otros dos MRG fueron procesados con anterioridad por otros autores (Guevara et al. 2001; Marroquín-Guerra et al. 2009) utilizando el mismo procedimiento que el usado en el presente trabajo. Adicionalmente, la síntesis de datos geoquímicas de otros dos fue tomada directamente de la literatura anterior, para los cuales fueron usados otros tipos de metodologías estadísticas (mayores detalles en el Apéndice 3; Verma et al., 2009a). De esta manera, se utilizaron un total de 14 MRG.

Es importante señalar que la opción de “*Default process*” en el software DODESYS usado para este capítulo, está organizada de la siguiente manera: para n hasta 9 datos se aplican solamente las pruebas sencillas; para $n=10-19$, además de las sencillas, se aplican pruebas múltiples tipo $k=2$; para $n=20-29$ y, se aplican pruebas múltiples tipo $k=2$ y $k=3$; y para $n>29$ todas las 33 variantes.

Con el propósito de realizar el estudio comparativo entre el método de preparación de pastillas prensadas y discos fundidos para su análisis por FRX, también se utilizó una base de datos que contiene muestras de rocas ígneas y sedimentarias de México (Verma, compilación no-publicada). De igual manera, los elementos Cr y Ni en los 14 MRG fueron analizados por el método de FRX (ver el Apéndice 3; Verma et al., 2009a).

5.2 Comparación de resultados de DODESYS con la literatura

En la comparación de los métodos de preparación de muestras para FRX, se realizó previamente un análisis estadístico de las medias y desviación estándar de los MRG. Para ello se compararon los valores en la literatura que han aplicado el método de las dos desviaciones estándar (2s) o tres desviaciones estándar (3s) con los valores procesado por DODESYS, siguiendo la metodología propuesta por Verma (1997, 1998).

Se compararon los límites de confianza de la media de 95% de los elementos Cr y Ni de la literatura con los obtenidos en este trabajo (Figura 1a del Apéndice 3; Verma et al., 2009a). Se observó que los límites de confianza determinados en este trabajo son más pequeños que los reportados en la literatura ya que casi todos los datos se grafican arriba de la línea diagonal que representa valores iguales de los dos ejes. Por otra parte, comparando las medias de las concentraciones (Figura 1b del Apéndice 3; Verma et al., 2009a) aunque para muchos MRG las diferencias son relativamente pequeñas, para tres casos, estas diferencias son mucho más grandes (mayor al 10%).

Este estudio demuestra nuevamente que la metodología basada en pruebas múltiples proporciona mejor resultados estadísticos para los MRG.

5.3 Comparación de resultados de fluorescencia de rayos-x (FRX) obtenidos en pastillas prensadas y discos fundidos

Es conocido que para medir elementos traza el método de preparación de pastillas prensadas es más utilizado que el de discos fundidos, mientras que para elementos mayores los discos fundidos son más usados (Verma et al. 1992). Sería útil saber si los discos fundidos puedan proporcionar valores de concentración confiables para algunos elementos traza como Cr y Ni, además de los elementos mayores. Para ello, fue necesario comparar estadísticamente los resultados de Cr y Ni en discos fundidos con pastillas prensadas. Los resultados de esta comparación se presentan en las Figuras 2 y 3 del Apéndice 3 (Verma et al., 2009a) para los MRG y las rocas de México, respectivamente.

Las medias de concentraciones obtenidas por los dos métodos de preparación de MRG demuestran diferencias pequeñas (Figura 2a, b y d del Apéndice 3; Verma et al., 2009a; en general, menor al 10%), mientras que los límites de confianza de 95% para los discos fundidos son mayores que para pastillas prensadas (Figura 2c, Apéndice 3; Verma et al., 2009a). Los resultados para las rocas de México, por lo general, son similares a los MRG, excepto cuando se trata de bajas concentraciones de Cr y Ni (menor a 50 ppm; Figura 3a-c, del Apéndice 3; Verma et al., 2009a). Estas bajas concentraciones se presentan para rocas mexicanas ricas en sílice (Figura 3d, Apéndice 3; Verma et al., 2009a).

Estos resultados confirman que los dos métodos de preparación se pueden emplear para la determinación de Cr y Ni cuando se trata de una aplicación de rutina en Ciencias de la Tierra, aunque se debe recordar que las pastillas prensadas proporcionan datos con menores límites de confianza que los discos fundidos.

CAPÍTULO 6

EVALUACIÓN DE DIAGRAMAS DE CLASIFICACIÓN PARA ROCAS ALTERADAS

Introducción

Existen reglas establecidas por la Unión Internacional de Ciencias Geológicas (en inglés, International Union of Geological Sciences: IUGS) para clasificar las rocas ígneas frescas. La asignación de los nombres de las rocas depende de sus características químicas y mineralógicas. La clasificación de rocas ígneas volcánicas se basa en el uso del diagrama TAS (total álcalis *versus* sílica; Le Bas et al. 1986) y la norma CIPW (propuesta por Cross, Iddings, Pirsson and Washington., 1931). Sin embargo, la IUGS no previó las especificaciones para clasificar rocas alteradas. En este capítulo, se presenta la evaluación de los diagramas de Winchester & Floyd (1977) y Hastie (2007) propuestos para clasificar rocas alteradas. Para mayores detalles ver el Apéndice 4 (Verma et al., 2009b).

6.1 Base de datos mundial

La extensa base de datos mundial utilizada para evaluar los diagramas de clasificación para rocas alteradas es una compilación no-publicada de Verma. Se cuenta con datos de diferentes ambientes tectónicos (arco de isla, “rift” continental, isla oceánica y cresta mid-oceánica, así como del Cinturón Volcánico Mexicano (en inglés Mexican Volcanic Belt). Esta base de datos consiste de las siguientes clasificaciones de rocas o magmas: (i) 1934 muestras de magmas ácidos (dacitas (D), riolitas (R), traquitas (T) y, traquidacitas (TD)); (ii) 2082 muestras de magmas básicos (basaltos alcalinos (B,alk), basanitas (BSN), basalto traquiandesita (BTA), foiditas (FOI), fonotefritas (PHT), picritas (PIC), tefritas (TEP) y tefrifonolitas (TPH)); (iii) 12870 muestras de basaltos subalcalinos (B, Subalk); (iv) 1085 muestras de magmas intermedios alcalinos (basaltos traquiandesita (BTA), traquiandesita (TA) y tefrifonolita (TPH)); (v) 1925

muestras de magmas intermedios (andesitas (A)); (vi) 2123 muestras de basaltos alcalinos (B, alk); y (vii) 923 muestras de magmas ultrabásicos (tefritas (TEP), foiditas (FOI) y basanitas (BSN)).

La información que utilizamos en estos archivos fue: (i) nombre de la roca; (ii) clasificación (TAS y general) y (iii) datos geoquímicos de los elementos: SiO_2 , Na_2O y K_2O (todos ellos ajustados por el programa SINCLAS; Verma et al. 2002), Nb, Y, Zr, TiO_2 , Co y Th.

6.2 Diagramas de Clasificación para rocas alteradas

Los diagramas de clasificación para rocas alteradas fueron propuestos inicialmente por Floyd & Winchester (1975, 1978) y Winchester & Floyd (1976, 1977). Sus diagramas binarios se caracterizan por tres tipos de propuestas: (i) elemento-elemento, p. ej., Zr- TiO_2 ; (ii) elemento-relación elemento, p. ej., Nb/Y- SiO_2 ; y (iii) relación elemento-relación elemento, p. ej., Nb/Y-Zr/ TiO_2 . Estas propuestas para clasificar rocas alteradas presentan cierto parecido al diagrama TAS, excepto que los elementos sílice (SiO_2) y álcalis ($\text{Na}_2\text{O} + \text{K}_2\text{O}$) son sustituidos por elementos inmóviles aunque en algunos de ellos el sílice sigue presente.

El diagrama elegido para la evaluación de rocas alteradas utilizando rocas frescas fue el de Nb/Y vs Zr/ TiO_2 (Winchester & Floyd, 1977). Las razones de esta elección han sido expresadas ampliamente en el Apéndice 4 (Verma et al., 2009b). En el diagrama de Winchester & Floyd (1977), el elemento sílice (SiO_2), se sustituye por la relación Nb/Y y los álcalis ($\text{Na}_2\text{O} + \text{K}_2\text{O}$) por Zr/ TiO_2 .

El otro diagrama evaluado en el Apéndice 4 (Verma et al., 2009b) fue el propuesto recientemente por Hastie et al. (2007). Es un diagrama binario para rocas alteradas, especialmente para rocas de arco, utilizando a los elementos torio (Th) vs cobalto (Co) como elementos inmóviles.

Para la evaluación los diagramas Nb/Y-Zr/TiO₂ y Co-Th fueron construidos usando todo tipo de rocas ígneas descritas en la base de datos. Para mayor información ver Apéndice 4 (Verma et al., 2009b).

Es conocido que el diagrama TAS no es recomendado para clasificar rocas volcánicas alteradas. Sin embargo, para rocas frescas la nomenclatura del diagrama TAS parece ser aceptable en forma universal. Es importante mencionar que las base de datos que utilizamos para evaluar los diagramas de Winchester & Floyd (1977) y el diagrama de Hastie et al. (2007) corresponde a rocas frescas (ver la siguiente sección). Para este tipo de rocas, el diagrama TAS nos proporciona la clasificación que nos sirve de punto de referencia para evaluar los diagramas propuestos para las rocas alteradas. Por tal motivo, se procedió a evaluar con una base de datos de rocas frescas. Si estos diagramas nos proporcionaran los mismos nombres de TAS con un alto porcentaje de éxito, podríamos concluir que los diagramas funcionan correctamente para rocas frescas. Por consecuencia, podrían utilizarse en la clasificación de rocas alteradas. De lo contrario, si la clasificación de rocas frescas fuera con porcentaje bajos, podríamos concluir que los diagramas no funcionan y existe la necesidad de proponer nuevos diagramas para este propósito.

6.2.1 Diagrama de Clasificación para rocas alteradas de Winchester & Floyd (1977)

Winchester & Floyd presentaron una decena de diagramas para clasificar rocas alteradas, algunas de los cuales son: (i) Zr-P₂O₅; (ii) Zr-TiO₂; (iii) Ce-Zr/TiO₂; (iv) Ga-Zr/TiO₂; (v) Zr/TiO₂-SiO₂; Nb/Y-SiO₂; (vii) Y/Nb-TiO₂; (viii) Zr/P₂O₅-TiO₂; (ix) Nb/Y-Zr/TiO₂; (x) Nb/Y-Ga/Sc; y (xi) Zr/P₂O₅-Nb/Y (vii) Y/Nb-TiO₂; (viii) Zr/P₂O₅-TiO₂; (ix) Nb/Y-Zr/TiO₂; y (x) Zr/P₂O₅-Nb/Y.

Los nombres asignados a las rocas frescas de la base de datos fueron acordes con las reglas de la IUGS. El diagrama Nb/Y - Zr/TiO₂ de (Winchester & Floyd, 1977) tiene escala logarítmica, y en él se observa que la relación de Nb/Y se incrementa de regiones sub-alcálicas a alcálicas, mientras que la de Zr/TiO₂ se incrementa de regiones básicas a ácidas.

El diagrama de Nb/Y–Zr/TiO₂ (Winchester & Floyd, 1977) es utilizado ampliamente desde hace más de 30 años hasta el día de hoy como ha sido demostrado en el Apéndice 4 (Verma et al., 2009b).

6.2.2 Diagramas de clasificación para rocas alteradas de Hastie et al. (2007)

El diagrama propuesto por Hastie et al. (2007) proporciona las coordenadas de las fronteras entre la separación en los diferentes campos. Existen tres regiones que los autores proponen los nombres de las rocas: (i) basalto (B); (ii) andesita basáltica con andesita (BA/A); y (iii) dacita con riolita (D/R) que incluye latita y traquita (dacita-riolita-latita-traquita). Se pueden notar traslapes en las regiones cuando este diagrama se compara con el diagrama TAS, en el cual se cuenta con una docena de regiones separadas.

Cada una de estas tres regiones del diagrama de Hastie et al. (2007) está subdividida en serie toleítica, calc-alcalina y alto K₂O soshonítica, lo cual no fue evaluado por no corresponder a la nomenclatura de IUGS.

Un aspecto importante es que para esta evaluación se utilizaron los mismos ambientes tectónicos o provincias volcánicas dado que Hastie et al. (2007) afirmaban que su diagrama es más útil para el ambiente de arco.

6.3 Metodología para la evaluación de diagramas para rocas alteradas

La metodología de evaluación de los diagramas para rocas alteradas, inicia con: (i) la clasificación de las muestras de rocas frescas en el diagrama TAS que les asignó un nombre de acuerdo a la IUGS. Este diagrama divide a las rocas en: (a) ultrabásicas; (b) básicas; (c) intermedias y; (d) ácidas según contenido de sílice (Le Bas et al. 1986). La nomenclatura en el diagrama TAS está basada en un sistema que incluye un nombre de raíz con un calificativo adicional, p. ej., el nombre de basalto debe ir acompañado por alcalino o sub-alcalino.

Como siguientes pasos del proceso, se realizó lo siguiente: (ii) Separación de las bases de datos de rocas frescas por ambientes tectónicos; (iii) Separación de muestras por tipo de roca en las bases compiladas; (iii) Clasificaron IUGS por tipo de roca en la base de datos, para que pueda ser evaluado cada roca en el nuevo diagrama y determinado a qué tipo corresponde; (v) Contabilización del número de muestras de cada tipo de roca evaluada para determinar los números totales que corresponden a cada categoría; (vi) Cálculos estadísticos de la información mediante porcentajes de cada nuevo tipo de roca en términos del tipo de roca original que ha sido evaluado; y (vii) Reporte de los resultados en figuras y tablas.

6.4 Evaluación de diagramas de clasificación para rocas alteradas de Winchester & Floyd (1977) y Hastie et al. (2007)

6.4.1 Diagrama de Winchester & Floyd (1977).

Los resultados de la evaluación del diagrama para rocas alteradas de Winchester & Floyd (1977), se muestran en la Tabla 1 del Apéndice 4 (Verma et al., 2009b).

En la parte superior de la tabla se observan los nombres asignados a las rocas de acuerdo a la clasificación de Winchester & Floyd, y en la columna izquierda los asignados para clasificar rocas volcánicas de acuerdo a IUGS de tal forma que se pueda establecer la comparación. La tabla fue dividida en dos partes, en la parte superior se tienen los nombres de las rocas que coinciden en la clasificación IUGS, como Winchester & Floyd y dónde aparece la línea de división se incluyen rocas ultrabásicas y ácidas no incluidas en la clasificación de Winchester & Floyd.

El tipo de roca que presentó menor porcentaje en la clasificación fue la riolita (16.6%), por lo que se considera mal clasificada en el diagrama de Winchester & Floyd (1977) y, el porcentaje más alto lo obtuvo el basalto alcalino (73.8). De manera general, se puede decir que las rocas mejor clasificadas fueron: (i) basalto alcalino; (ii) andesita y; (iii) fonolita (73.8-76%).

Sin embargo, el porcentaje obtenido para la clasificación de basalto sub-alcalino, basanita, traquiandesita, dacita, y riolita son muy bajos (16.6%-32.0%), por lo que no puede ser considerada como una clasificación aceptable. Los otros nueve tipos restantes de rocas incluidos en la clasificación de IUGS (segunda parte de la tabla) no se clasificaron con un nombre coherente de acuerdo al esquema propuesto por Winchester & Floyd (2007). De estos resultados, podemos concluir que no es recomendable el uso de este diagrama, por lo que surge la necesidad de un esquema adecuado de la clasificación para rocas alteradas.

6.4.2 Diagramas de Clasificación de Hastie et al. (2007).

En la Tabla 2 del Apéndice 4 (Verma et al., 2009b), se muestra la comparación de la clasificación de las rocas frescas de acuerdo a la IUGS con la de Hastie et al. (2007). La Tabla 2 fue dividida en dos partes: (i) en la parte superior se ubican las rocas que pueden ser comparadas directamente con la IUGS; (ii) en la parte inferior se encuentran las rocas que sólo tienen nombres asignados de acuerdo a la IUGS, pero no por Hastie et al. (2007).

La base de datos utilizada se redujo porque únicamente se emplearon las rocas de arco de isla y el Cinturón Volcánico Mexicano. Por otra parte, no fue posible realizar comparaciones directas, dado que, p ej., en IUGS los basaltos se definieron como basalto-subalcalino y en Hastie et al. (2007) se encuentran distribuidos en B de tres tipos: (i) Thol; (ii) CA; y (iii) SHO. Esto hace difícil establecer el porcentaje de clasificación correcta, por lo que se optó por estimar la mala clasificación de las rocas. Como ejemplo, si en el diagrama TAS un conjunto de muestras se clasifican como BA, la mala clasificación estaría cuando las muestras caen en el campo de B o de D/R en el diagrama de Hastie et al. (2007).

De acuerdo a lo anterior, los porcentajes reportados para rocas mal clasificadas en el diagrama de Hastie et al. (2007) se presentaron en tres categorías: (i) los más bajos se ubicaron entre 1.2-8%; (ii) los que se consideraron en término medio (12.8-24%); y (iii) los porcentajes más altos de 25.3-93%.

Es difícil de evaluar la clasificación IUGS de rocas, cuyos nombres no se encuentran en la clasificación de Hastie et al. (2007), como es el caso de la segunda parte de la Tabla 2 del Apéndice 4 (Verma et al., 2009b). Sin embargo, es de esperarse que las rocas en el último grupo de esta tabla están claramente mal clasificadas cuando se ubican como basalto o dacita/riolita; esto representa un porcentaje relativamente alto (34.0-70.2%) de mala clasificación. Para mayor información ver el Apéndice 4 (Verma et al., 2009b).

CAPÍTULO 7

APLICACIÓN DEL DIAGRAMA DE WINCHESTER Y FLOYD PARA LA CLASIFICACIÓN DE ROCAS ALTERADAS EN EL CAMPO GEOTÉRMICO DE LOS AZUFRES, MICHOACÁN, MÉXICO

Introducción

El campo geotérmico de Los Azufres está ubicado en la parte central del cinturón volcánico Mexicano (en inglés MVB). Representa el segundo campo geotérmico más grande de México, dada su importancia ha sido motivo de estudio de diferentes autores entre muchos (Portugal et al. 1984; Santoyo et al., 1991; González-Partida et al. 2003; Verma et al. 2005). Por ello se eligió como campo de estudio para evaluar el diagrama para rocas alteradas de Winchester y Floyd (1977).

En este capítulo se aplicará el diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977), a muestras de rocas frescas y alteradas del campo geotérmico de Los Azufres, Michoacán, México. El diagrama de Hastie et al. (2007) para rocas alteradas descrito en el capítulo anterior no fue posible aplicarlo, debido a que no se tenía información geoquímica completa de los elementos Co y Th necesarios para aplicar el diagrama.

7.1 Bases de datos del campo geotérmico de Los Azufres.

La base de datos geoquímica del campo geotérmico de Los Azufres, Michoacán, México, utilizada para evaluar el diagrama de clasificación para rocas alteradas Winchester y Floyd (1977), se formó de 19 muestras de rocas frescas (ácidas e intermedias) y 86 muestras de rocas alteradas (ácidas e intermedias). Sin embargo, pocos trabajos (Dobson_Mahood, 1985; Torres-

Alvarado et al. 1987 y 2007; Verma et al. 2005; Pandarinath et al. 2008) presentan datos completos para usarlos en el diagrama W&F.

Para aplicar el diagrama de TAS y de Winchester y Floyd (1977), la información utilizada de estos archivos fue: (i) nombre de la roca; (ii) clasificación (TAS y general); y (iii) datos geoquímicos de los elementos SiO₂, Na₂O, K₂O, Nb, Y, Zr, y TiO₂. La base de datos fue procesada previamente por el programa de computación SINCLAS (Verma et al. 2002) que ajustó las concentraciones de SiO₂, Na₂O y K₂O con base seca y al 100%.

En la base de datos no se encontraron rocas básicas con los datos necesarios de elementos traza para el diagrama Winchester y Floyd (1977).

7.2 Diagramas de clasificación TAS vs diagrama W&F del campo geotérmico Los Azufres

Las muestras de rocas procedentes del campo geotérmico de Los Azufres separadas previamente en muestras de rocas frescas (ácidas e intermedias) y alteradas (ácidas e intermedias). Se clasificaron de acuerdo a la IUGS en: (i) ácidas frescas (16 riolitas (R)); (ii) intermedias frescas (1 andesitas (A) y 2 andesitas basálticas (BA)); (iii) ácidas alteradas (25 riolitas (R) y 19 dacitas (D)); (iv) intermedias alteradas (33 andesitas (A), 8 andesitas basálticas (BA) y 1 traquiandesita (TA)).

Una síntesis de los resultados de comparación de la clasificación de rocas en el diagrama TAS con el diagrama de Winchester y Floyd (1977) se presenta en las Tablas 7.1 y 7.2. En estas tablas y por simplicidad, las columnas de tipo de rocas que no tienen muestras de Los Azufres, han sido eliminadas. Se obtuvieron las gráficas de las Figuras 7.1-7.8.

En la aplicación de los diagramas TAS y Winchester y Floyd (1977) se siguió la misma metodología descrita en el capítulo anterior y el Apéndice 4 (Verma et al., 2009b).

7.3 Discusión de resultados del Diagrama Winchester y Floyd (1977) de clasificación para rocas alteradas

7.3.1 Aplicación del diagrama de clasificación Winchester y Floyd (1977) utilizando rocas frescas

En la Tabla 7.1, se presentan los resultados de las 19 muestras de rocas frescas (incluye rocas ácidas e intermedias) del campo geotérmico de los Azufres. En las Figuras 7.1 del diagrama TAS y Figura 7.2 diagrama de Winchester y Floyd (1977) se muestra la comparación en la clasificación de 3 rocas frescas intermedias. La andesita (A) se clasificó como una roca de traslape del campo Andesita/Basalto en el diagrama Winchester y Floyd (1977) y las 2 andesitas basálticas como basaltos sub-alkalinos (B, Sub-Alk; Figura 7.2).

Tabla 7.1. Evaluación del diagrama de Nb/Y–Zr/TiO₂ (Winchester & Floyd, 1977), utilizando rocas **frescas** del campo geotérmico de Los Azufres, Michoacán, México, comparado con la clasificación de rocas volcánicas de la IUGS.

IUGS classification	Número total de ejemplos (%)	Figura #	Número de ejemplos clasificados (% de ejemplos clasificados) acorde al diagrama Nb/Y–Zr/TiO ₂ (Winchester & Floyd, 1977)						
			Basalto sub-alkalino	Andesita/Basalto	Andesita	Traqui-andesita	Comendita/Pantelerita	Riodacita/Dacita	Riolita
Andesita	1 (100)	7.1, 7.2		1					
Riolita	16 (100)	7.3, 7.4			1 (6)	2(13)	5(31)	5 (31)	3(19)
Andesita basáltica	2 (100)	7.1, 7.2	2						

Los números en negrita y cursiva indican la correcta clasificación, los números en cursiva indican el tipo de roca no incluido en el diagrama de clasificación de W&F.

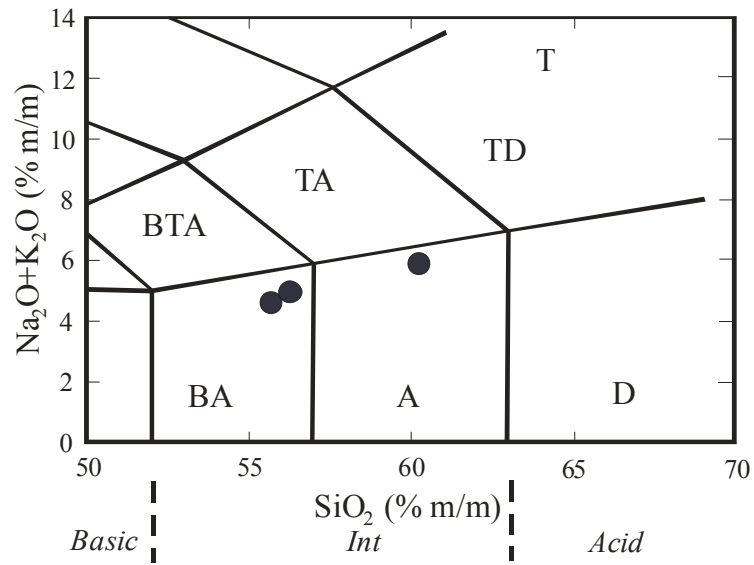


Figura 7.1 Evaluación estadística del diagrama de clasificación TAS, usando rocas frescas intermedias (andesita y andesita basáltica).

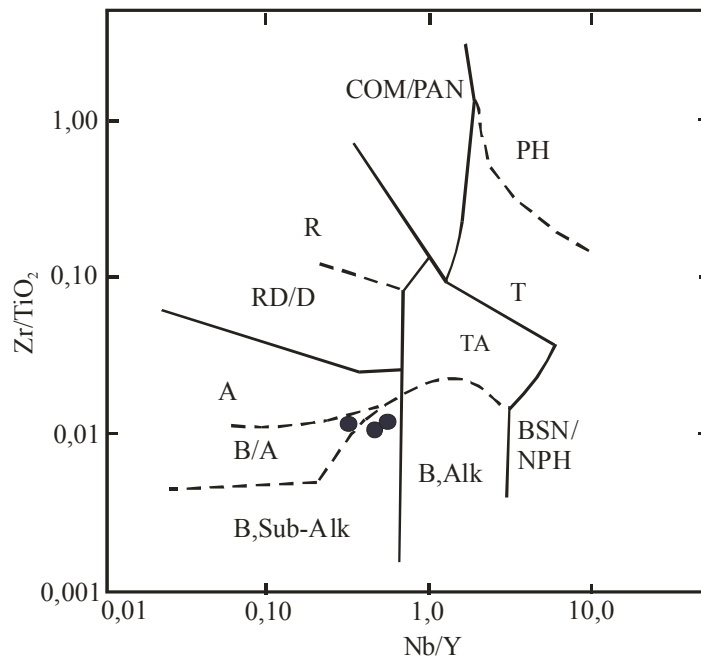


Figura 7.2 Evaluación estadística del diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) usando rocas frescas intermedias.

Por otra parte, las 16 muestras de rocas frescas ácidas en del diagrama TAS (Figura 7.3) que comprende a las riolitas se clasificaron en el diagrama Winchester y Floyd (1977) (Figura 7.4) como: (i) andesitas (A; 6.3%); (ii) traquiandesita (TA; 12.5%); (iii) comendita/pantelerita (COMEN/PAN; 31.2%); (iv) riodacita/dacita (RD/D; 31.2%); y (v) riolita (R; 18.8%). Solamente, 3 de las 16 muestras de riolitas se clasifican correctamente.

Se observa que los porcentajes obtenidos al clasificar las muestras de rocas frescas ácidas, mal clasificadas en el diagrama de Winchester y Floyd (1977) son semejantes a los obtenidos en la evaluación descrita en el capítulo VI y Apéndice 4 (Verma et al., 2009b), por lo que nuevamente se confirma la necesidad de la creación de un esquema adecuado para clasificar rocas alteradas.

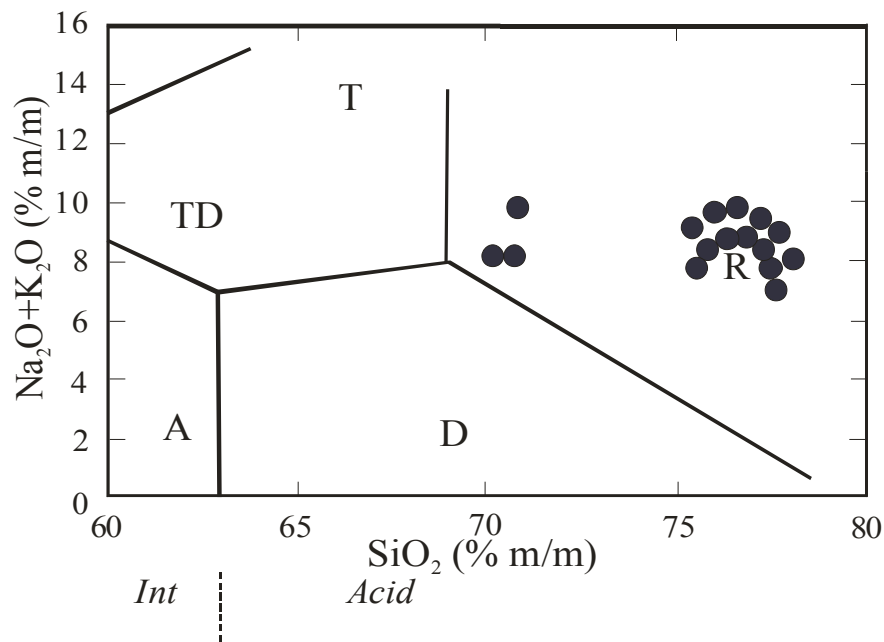


Figura 7.3 Evaluación estadística del diagrama de clasificación TAS usando rocas frescas ácidas (riolitas).

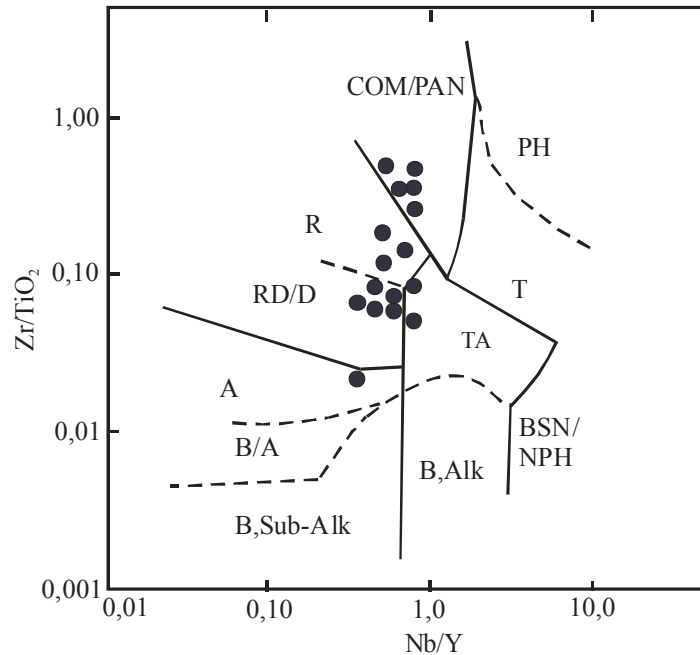


Figura 7.4 Evaluación del diagrama de clasificación para rocas alteradas de Winchesteer y Floyd (1977) usando rocas frescas ácidas.

7.3.2 Aplicación del diagrama de clasificación Winchesteer y Floyd (1977) utilizando rocas alteradas

Se aplicó el diagrama de clasificación para rocas alteradas de Winchesteer y Floyd (1977) a 86 muestras de rocas alteradas (intermedias y ácidas) del campo geotérmico de Los Azufres, Michoacán, México. Los resultados de esta aplicación se presentan en la Tabla 7.2. y Figuras 7.5-7.8.

En el diagrama TAS (Figura 7.5) (clasificación IUGS) se graficaron previamente las muestras de rocas alteradas aunque la clasificación no se debe utilizar como referencia para otros diagramas. De las 86 muestras alteradas, 42 fueron rocas alteradas intermedias clasificadas en 33 andésitas (A), 8 andesitas basálticas (BA) y 1 traquiandesita (TA). Sin embargo en el diagrama de Winchesteer y Floyd (1977) (Figura 7.6) las 33 andesitas (A) se clasificaron como: andesita (A; 72.7%); rocas de traslape en el campo basalto/andesita (B/A; 18.2%); en el campo de riodacita/dacita (RD/D; 6.1%); y como basalto sub-alcalino (B,

Subalk; 3.0%). La única muestra de traquiandesita (TA) se clasificó como basalto sub-alkalino (B, Subalk). Finalmente, de las 8 andesitas basálticas, 4 se clasificaron en el campo de traslape basalto/andesita (B/A) y 3 en basalto sub-alkalino (B, Subalk) y, 1 andesita (A).

Tabla 7.2. Evaluación del diagrama de Nb/Y–Zr/TiO₂ (Winchester & Floyd, 1977) utilizando rocas **alteradas** del campo geotérmico de Los Azufres, Michoacán, México, comparado con la clasificación de rocas volcánicas de la IUGS.

IUGS clasificación	Número total de ejemplos (%)	Figura #	Número de ejemplos clasificados (% de ejemplos clasificados) acorde al diagrama Nb/Y– Zr/TiO ₂ (Winchester & Floyd, 1977)							
			Basalto alkalino	Basalto sub- alkalino	Andesita/ Basalto	Andesita	Traqui- andesita	Comendita/ Pantelerita	Riodacita/ Dacita	Riolita
Andesita	33 (100)	7.5, 7.6		1 (3.0)	6 (18.2)	24 <i>(72.7)</i>			2 (6.1)	
Traquiandesita	1 (100)	7.5, 7.6	1							
Dacita	19 (100)	7.7, 7.8		1 (5.3)		5 (26.3)	2 (10.5)		10 <i>(52.6)</i>	1 (5.3)
Riolita	25 (100)	7.7, 7.8				2 (8.0)	3 (12.0)	11 (44.0)	4 (16.0)	5 <i>(20.0)</i>
Andesite basáltica	8 (100)	7.5, 7.6		3 (38)	4 <i>(50)</i>	1 (12)				

Los números en negrita y cursiva indican la correcta clasificación; los números en cursiva indican el tipo de roca no incluido en el diagrama de clasificación de W&F.

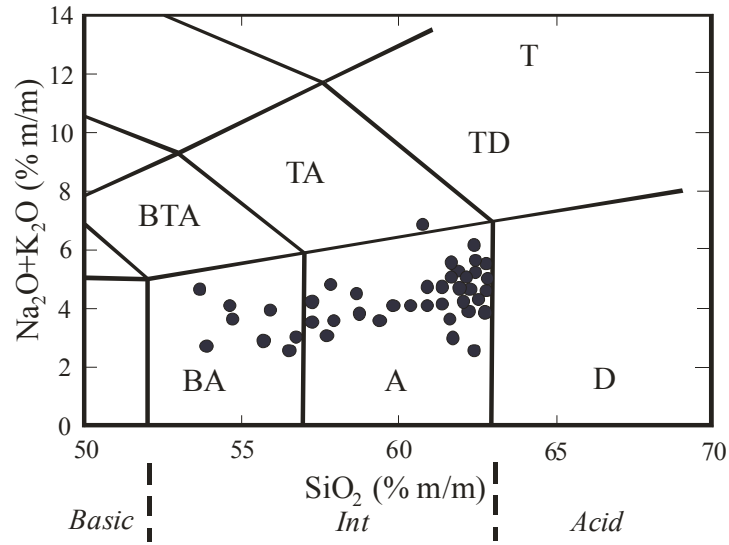


Figura 7.5 Evaluación estadística del diagrama de clasificación TAS usando rocas alteradas intermedias (andesita, andesita basáltica y traquiandesita).

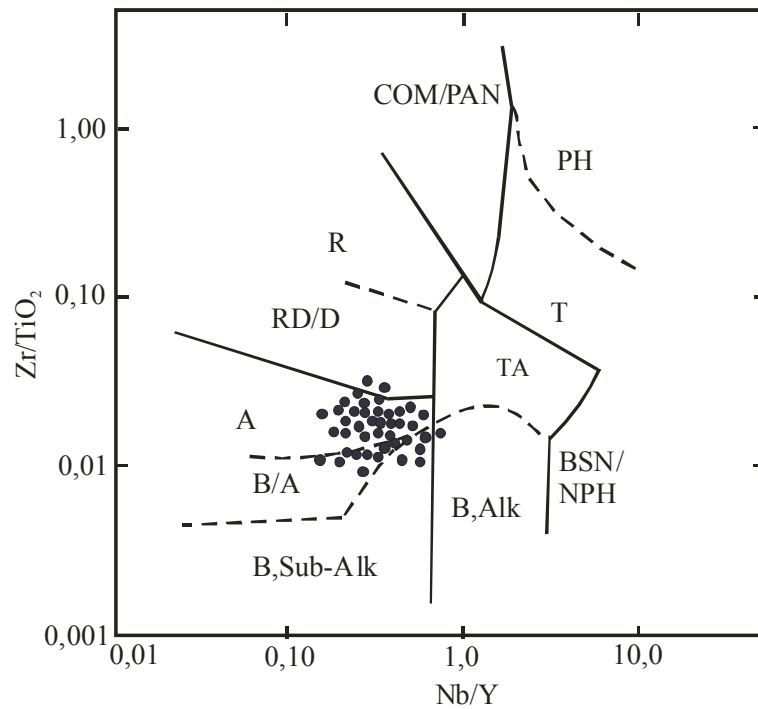


Figura7.6 Evaluación estadística del diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) usando rocas alteradas intermedias.

Las 44 rocas alteradas ácidas fueron divididas, según el diagrama TAS (Figura 7.7), como 19 dacitas y 25 riolitas. Las dacitas y riolitas, en el diagrama W&F (Figura 7.8), se clasificaron como A (26.3%), RD/D (52.6%), TA (10.5%), R (5.3%) y B, Subalk (5.3%), mientras que las riolitas como R (20.0%), A (8.0%), RD/D (16.0%), TA (12.0%); COMEN/PAN (44.0%).

De los resultados obtenidos en este capítulo, concluimos que el diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) no es confiable para clasificar rocas frescas. Por consecuencia, no debe aplicarse en rocas alteradas, estableciendo nuevamente que existe la necesidad de proponer nuevos diagramas para la clasificación de rocas alteradas.

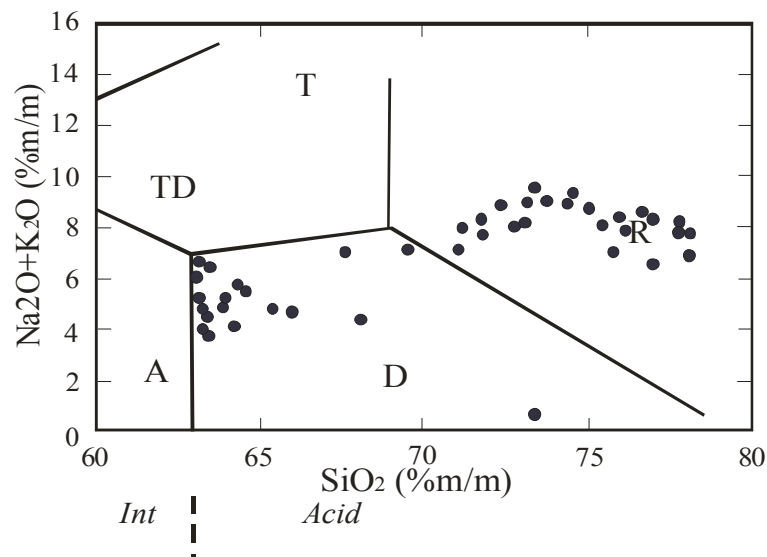


Figura 7.7 Evaluación estadística del diagrama de clasificación TAS usando rocas alteradas ácidas (riolitas y dacitas).

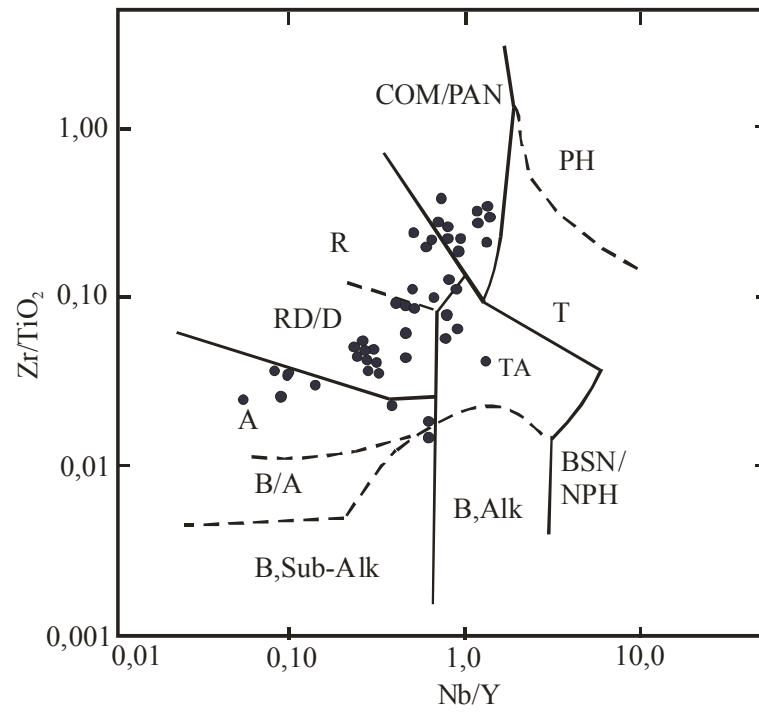


Figura 7.8. Evaluación del diagrama de clasificación para rocas alteradas de Winchester y Floyd (1977) usando rocas alteradas ácidas.

CONCLUSIONES Y TRABAJOS FUTUROS

En este proyecto doctoral se logró la compilación de una extensa base de datos con 35 materiales de referencia geoquímica (MRG), de 102 columnas (datos de elementos mayores y elementos traza) y 26,050 renglones (en total 62,254 datos geoquímicos). Mediante el tratamiento a la base de datos con el Software DODESYS se cumplieron importantes objetivos programados en el proyecto de investigación, llegando a las siguientes conclusiones:

1. En la determinación de las eficiencias $(REC)_g$ de 15 pruebas de discordancia con 33 variantes, se encontró que las pruebas tiene el siguiente orden: $N15 > N4U+N4L \approx N1U+N1L \approx N4U2+N4L2 > N2 \approx N14 \approx N13U2+N13L2 \approx N12U2+N12L2 > N4U3+N4L3 \approx N11U2+N11L2 > N5UL \approx N4U4+N4L4 \approx N10U+N10L \approx N9U+N9L > N3U2+N3L2 > N8 > N7 > N6UL > N3U3+N3L3 > N3U4+N3L4$.
2. En la regresión lineal determinada entre el criterio de eficiencia relativa de la primera iteración $(REC)_1$ y el tamaño de la muestra estadística, en combinación con las pruebas de significancia (*t*-Student, *F*-Fisher, y ANOVA), se demostró aumento de $(REC)_1$ con el tamaño de la muestra, así como también diferencias en $(REC)_1$ para algunas pruebas de discordancia.
3. Con la aplicación de pruebas de discordancia para detectar y en su caso depurar datos desviados a la base de datos, fue posible obtener medidas de tendencia central y dispersión, confiables para elementos mayores y traza. Los valores de elementos mayores fueron utilizados en la calibración del equipo de fluorescencia de rayos-x (FRX), logrando una calibración mediante la regresión lineal ponderada.
4. Las medidas de tendencia central y dispersión obtenidas para los elementos traza Cr y Ni de 10 MIRG, demostraron que los datos procesados mediante el método de valores desviados (aplicando pruebas de discordancia), proporcionan mejores límites de confianza al 95% que los publicados en la literatura por otros métodos estadísticos.
5. Al realizar la comparación estadística de los métodos de preparación de pastillas prensadas con el de disco fundidos se demostró que ambas técnicas de preparación proporcionan resultados de tendencia central similares. Sin embargo dispersión

diferente, dado que los límites de tendencia central determinados al 95% de confianza son mayores para discos fundidos, que para pastillas prensadas.

La otra parte esencial en este proyecto doctoral fue también la evaluación de los diagramas de clasificación para rocas alteradas de Winchester & Floyd (1977) y Hastie et al. (2007), utilizando para ello una base amplia de rocas frescas. En la evaluación de estos diagramas se encontró que:

6. No es confiable el uso de los diagramas de Winchester y Floyd (1977) y Hastie et al. (2007). Esto se debe a que la “clasificación correcta” en el diagrama de W&F resultó tener muy bajos porcentajes y la “mala clasificación” en el de Hastie et al. (2007) fue considerable.
7. Al aplicar el diagrama de clasificación de Winchester y Floyd (1977) a muestras de rocas frescas del campo geotérmico de Los Azufres Mich., México, se confirmó nuevamente que este diagrama no es confiable para la clasificación de rocas. La clasificación de rocas alteradas en este diagrama también fue poco coherente ya que las rocas de un determinado tipo (IUGS) se graficaron en distintos campos.

Tomando en consideración los logros obtenidos en el proyecto doctoral, percibo la necesidad de profundizar los estudios que permitan tener bases sólidas en el área estadística y en los campos geotérmicos. Por lo tanto, partiendo de esta inquietud se proponen las siguientes metas a realizar como trabajos futuros.

- i. Continuar con la captura de la base de datos MRG, para tener cada vez una base más completa y con datos actualizados. En la búsqueda electrónica realizada por las palabras clave, en muchas ocasiones aparecen documentos que no tienen la información deseada; de igual manera, otros documentos con datos relevantes no aparecen en dicha búsqueda. Hemos encontrado que muchos artículos de aplicación geoquímica sí contienen datos sobre MRG sin que se les pueda localizar en las búsquedas electrónicas. Se pueden encontrar los artículos pertinentes en forma manual y posteriormente, compilar estos datos. Si alguien localiza un artículo de los últimos cinco años con datos sobre algunos

de estos 35 MRG, sería de mucha ayuda contar con esta información con el objeto de actualizar la base de datos para futuras aplicaciones.

- ii. Determinar el $(REC)_g$ de las 15 pruebas de discordancia con sus 33 variantes.
- iii. Procesar la base de datos actualizada para conocer los parámetros estadísticos de tendencia central y dispersión para todos los elementos químicos.
- iv. Realizar la calibración del equipo de FRX, utilizando elementos traza.
- v. Comparar los métodos de preparación de pastillas prensadas y discos fundidos con un mayor número de elementos traza en el método de FRX.
- vi. Difundir en el sector productivo la importancia de los avances estadísticos en el ámbito científico, específicamente en el área de Control de Calidad de las empresas, así como también en el ámbito educativo.
- vii. Trabajar en una nueva propuesta para clasificar adecuadamente rocas alteradas, tomando en cuenta la naturaleza multi-variable de los componentes composicionales.

CONCLUSIONES Y TRABAJOS FUTUROS

En este proyecto doctoral se logró la compilación de una extensa base de datos con 35 materiales de referencia geoquímica (MRG), de 102 columnas (datos de elementos mayores y elementos traza) y 26,050 renglones (en total 62,254 datos geoquímicos). Mediante el tratamiento a la base de datos con el Software DODESYS se cumplieron importantes objetivos programados en el proyecto de investigación, llegando a las siguientes conclusiones:

1. En la determinación de las eficiencias $(REC)_g$ de 15 pruebas de discordancia con 33 variantes, se encontró que las pruebas tiene el siguiente orden: $N15 > N4U+N4L \approx N1U+N1L \approx N4U2+N4L2 > N2 \approx N14 \approx N13U2+N13L2 \approx N12U2+N12L2 > N4U3+N4L3 \approx N11U2+N11L2 > N5UL \approx N4U4+N4L4 \approx N10U+N10L \approx N9U+N9L > N3U2+N3L2 > N8 > N7 > N6UL > N3U3+N3L3 > N3U4+N3L4$.
2. En la regresión lineal determinada entre el criterio de eficiencia relativa de la primera iteración $(REC)_1$ y el tamaño de la muestra estadística, en combinación con las pruebas de significancia (t -Student, F -Fisher, y ANOVA), se demostró aumento de $(REC)_1$ con el tamaño de la muestra, así como también diferencias en $(REC)_1$ para algunas pruebas de discordancia.
3. Con la aplicación de pruebas de discordancia para detectar y en su caso depurar datos desviados a la base de datos, fue posible obtener medidas de tendencia central y dispersión, confiables para elementos mayores y traza. Los valores de elementos mayores fueron utilizados en la calibración del equipo de fluorescencia de rayos-x (FRX), logrando una calibración mediante la regresión lineal ponderada.
4. Las medidas de tendencia central y dispersión obtenidas para los elementos traza Cr y Ni de 10 MIRG, demostraron que los datos procesados mediante el método de valores desviados (aplicando pruebas de discordancia), proporcionan mejores límites de confianza al 95% que los publicados en la literatura por otros métodos estadísticos.
5. Al realizar la comparación estadística de los métodos de preparación de pastillas prensadas con el de disco fundidos se demostró que ambas técnicas de preparación proporcionan resultados de tendencia central similares. Sin embargo dispersión

diferente, dado que los límites de tendencia central determinados al 95% de confianza son mayores para discos fundidos, que para pastillas prensadas.

La otra parte esencial en este proyecto doctoral fue también la evaluación de los diagramas de clasificación para rocas alteradas de Winchester & Floyd (1977) y Hastie et al. (2007), utilizando para ello una base amplia de rocas frescas. En la evaluación de estos diagramas se encontró que:

6. No es confiable el uso de los diagramas de Winchester y Floyd (1977) y Hastie et al. (2007). Esto se debe a que la “clasificación correcta” en el diagrama de W&F resultó tener muy bajos porcentajes y la “mala clasificación” en el de Hastie et al. (2007) fue considerable.
7. Al aplicar el diagrama de clasificación de Winchester y Floyd (1977) a muestras de rocas frescas del campo geotérmico de Los Azufres Mich., México, se confirmó nuevamente que este diagrama no es confiable para la clasificación de rocas. La clasificación de rocas alteradas en este diagrama también fue poco coherente ya que las rocas de un determinado tipo (IUGS) se graficaron en distintos campos.

Tomando en consideración los logros obtenidos en el proyecto doctoral, percibo la necesidad de profundizar los estudios que permitan tener bases sólidas en el área estadística y en los campos geotérmicos. Por lo tanto, partiendo de esta inquietud se proponen las siguientes metas a realizar como trabajos futuros.

- i. Continuar con la captura de la base de datos MRG, para tener cada vez una base más completa y con datos actualizados. En la búsqueda electrónica realizada por las palabras clave, en muchas ocasiones aparecen documentos que no tienen la información deseada; de igual manera, otros documentos con datos relevantes no aparecen en dicha búsqueda. Hemos encontrado que muchos artículos de aplicación geoquímica sí contienen datos sobre MRG sin que se les pueda localizar en las búsquedas electrónicas. Se pueden encontrar los artículos pertinentes en forma manual y posteriormente, compilar estos datos. Si alguien localiza un artículo de los últimos cinco años con datos sobre algunos

de estos 35 MRG, sería de mucha ayuda contar con esta información con el objeto de actualizar la base de datos para futuras aplicaciones.

- ii. Determinar el $(REC)_g$ de las 15 pruebas de discordancia con sus 33 variantes.
- iii. Procesar la base de datos actualizada para conocer los parámetros estadísticos de tendencia central y dispersión para todos los elementos químicos.
- iv. Realizar la calibración del equipo de FRX, utilizando elementos traza.
- v. Comparar los métodos de preparación de pastillas prensadas y discos fundidos con un mayor número de elementos traza en el método de FRX.
- vi. Difundir en el sector productivo la importancia de los avances estadísticos en el ámbito científico, específicamente en el área de Control de Calidad de las empresas, así como también en el ámbito educativo.
- vii. Trabajar en una nueva propuesta para clasificar adecuadamente rocas alteradas, tomando en cuenta la naturaleza multi-variable de los componentes composicionales.

REFERENCIAS

Anderson R.L. (1987)

Practical statistics for analytical chemists, Van Nostrand Reinhold, 316 pp.

Barnett V. and Lewis T. (1994)

Outliers in statistical data, Chichester, John Wiley & Sons, 584 pp.

Cross, W., Iddings, J.P., L.V. and Washington, H.S. (1902)

A quantitative chemico-mineralogical classification and nomenclature of igneous rocks. *J. Geology*, 10, 555-690

Díaz-González L., (2008)

Calibración del geotermómetro de Na/K usando una base de datos geoquímicos mundial, mas representativa, técnicas geoquimométricas mejoradas y herramientas computacionales. Tesis de doctorado, Posgrado en Ingeniería (Energía), Universidad Nacional Autónoma de México (UNAM)

Dobson P.F. and Mahood G.A. (1985)

Volcanic stratigraphy of the Los Azufres geothermal area, Mexico. *Journal of Volcanology and Geothermal Research*, 25, 273-287.

Ebdon D. (1988)

Statistics in Geography, Oxford, Basic Blackwell, 232 p. pp.

Floyd P.A. and Winchester J.A. (1975)

Magma type and tectonic setting discrimination using immobile elements. *Earth and Planetary Science Letters*, 27, 211-218.

Floyd P.A. and Winchester J.A. (1978)

Identification and discrimination of altered and meta-morphosed volcanic rocks using immobile elements. *Chemical Geology*, 21, 291-306.

Gibbons R.D. (1994)

Statistical methods for groundwater monitoring, New York, John Wiley & Sons, Inc., 279 pp.

Gilbert R.O. (1987)

Statistical methods for environmental pollution monitoring, New York, Van Nostrand Reinhold, 319 pp.

González-Partida E., Levresse G. and Santoyo Gutierrez E. (2003)

Paleo-fluid and actual-fluid in the Los Azufres geothermal field, Central Mexico. *Journal of Geochemical Exploration*, 78-79, 67-70.

González-Ramírez R., Díaz-González, L., Verma S.P. (2009)

Eficiencia relativa de 15 pruebas de discordancia con 33 variantes aplicadas al procesamiento de datos geoquímicas. *Revista Mexicana de Ciencias Geológicas*, 26 (2), 501-515

Guevara M., Verma S.P. and Velasco-Tapia F. (2001)

Evaluation of GSJ intrusive rocks JG1, JG2, JG3, JG1a, and JGb1. *Revista Mexicana de Ciencias Geológicas*, 18, 74-88.

Hastie A.R., Kerr A.C., Pearce J.A. and Mitchell S.F. (2007)

Classification of altered volcanic island rocks using immobile trace elements: development of the Th-Co discrimination diagram. *Journal of Petrology*, 48, 2341-2357.

Jensen J.L., Lake L.W., Corbett P.W.M. and Goggin D.J. (1997)

Statistics for petroleum engineers and geoscientists, Upper Saddle River, Prentice-Hall, 390 pp.

Keskin M. (2007)

Eastern Anatolia: a hot spot in a collision zone without a mantle plume. in: G.R. Foulger and D.M. Jurdy (Eds.), *Plates, Plumes, and Planetary Processes*, Geological Society of America. Special Paper 430: 27 p.

Le Bas M.J. (2000)

IUGS reclassification of the high-Mg and picritic volcanic rocks. *Journal of Petrology*, 41, 1467-1470

Le Bas M.J., Le Maitre R.W., Streckeisen A. and Zanettin B. (1986)

A chemical classification of volcanic rocks based on the total alkali-silica diagram. *Journal of Petrology*, 27, 745-750.

Marroquín-Guerra., Velasco-Tapia, F and Díaz-González, L. (2009)

Evaluación estadística de Materiales de Referencia Geoquímica del Centre de Recherches Pétrographiques et Géochimiques (Francia) aplicando un esquema de detección y eliminación de valores desviados y su posible aplicación en el control de Calidad de datos geoquímicos. *Revista Mexicana de Ciencias Geológicas* (in press).

Miller J.C. and Miller J.N. (1993)

Estadística para química analítica, Wilmington, Delaware, U.S.A., Addition-Wesley Iberoamericana, 211 pp.

Miller J.N. and Miller J.C. (2000)

Statistics and chemometrics for analytical chemistry, Essex CM20 2JE, England, Prentice Hall, 271 pp.

Portugal-Marín E., Santoyo E., Verma S.P. and Nieva D. (1984)

Composicion química de fluidos geotérmicos, el campo de Los Azufres, Mich., Memorias de la Reunión de la Unión Geofísica Mexicana, La Paz, B.C. Sur.

Pandarinath K., Dulski P., Torres-Alvarado I.S. and Verma S.P. (2008)

Element mobility during hydrothermal alteration of rhyolitic rocks of the Los Azufres geothermal field, Mexico. *Geothermics*, 37, 53-72.

Santoyo E., Verma S.P., Nieva D. and Portugal E. (1991)

Variability in the gas phase composition of fluids discharged from Los Azufres geothermal field, Mexico. *Journal of Volcanology and Geothermal Research*, 47, 161-181.

Taylor J.K. (1990)

Statistical techniques for data analysis, Michigan, U.S.A, Lewis Publishers Inc., 200 pp.

Torres-Alvarado I.S. (1997)

Aplicación del modelado geoquímico en el estudio de la alteración hidrotermal en el campo geotérmico de Los Azufres, Mich. *Actas INAGEQ*, 3, 205-214.

Torres-Alvarado I.S., Pandarinath K., Verma S.P. and Dulski P. (2007)

Mineralogical and geochemical effects due to hydrothermal alteration in the Los Azufres geothermal field, Mexico. *Revista Mexicana de Ciencias Geológicas*, 24, 15-24.

Velasco-Tapia F., Guevara M. and Verma S.P. (2001)

Evaluation of concentration data in geochemical reference materials. *Chemie der Erde*, 61, 69-91.

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. The Journal of Geostandards and Geoanalysis*, 21, 59-75.

Verma S.P. (2005)

Estadística básica para el manejo de datos experimentales: aplicación en la Geoquímica (Geoquimiometría), México, D.F., UNAM, 186 pp.

Verma S.P. (2009)

Evaluation of polynomial regression models for the Student t and Fisher F critical values, the best interpolation equations from double and triple natural logarithm transformation of degrees of freedom up to 1000, and their applications to quality control in science and engineering. *Revista Mexicana de Ciencias Geológicas*, 26, 79-92.

Verma S.P. and 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, 133-161.

Verma S.P. and 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, 302-319.

Verma S.P. and Quiroz-Ruiz A. (2008)

Critical values for 33 discordancy test variants for outliers in normal samples for very large sizes of 1,000 to 30,000. *Revista Mexicana de Ciencias Geológicas*, 25, 369-381.

Verma S.P., Guevara M., Besch T. and Schulz-Dobrich B. (1993)

Elementos traza y mayores en muestras internacionales de referencia geoquímica por espectrometría de rayos X. *Geofísica Internacional*, 32, 311-319.

Verma S.P., Torres-Alvarado I.S. and Sotelo-Rodríguez Z.T. (2002)

SINCLAS: standard igneous norm and volcanic rock classification system. *Computers & Geosciences*, 28, 711-715.

Verma S.P., Quiroz-Ruiz A. and Díaz-González L. (2008)

Critical values for 33 discordancy test variants for outliers in normal samples up to sizes 1000, and applications in quality control in Earth Sciences. *Revista Mexicana de Ciencias Geológicas*, 25, 82-96.

Verma S.P., Díaz-González L. and González-Ramírez R. (2009)

Relative efficiency of single-outlier discordancy tests for processing geochemical data on reference materials. *Geostandards and Geoanalytical Research*, 33, 29-49.

Verma, S.P., González-Ramírez R., Rodríguez-Ríos R. (2009a)

Comparison of two sample preparation methods in x-ray fluorescence spectrometry for the analysis of Ni and Cr. *Geostandards and Geoanalytical Research* (enviado: 10 de mayo, 2009; revisado: 10 de agosto, 2009; se espera respuesta del editor).

Verma S.P., Rodríguez-Ríos, R., González-Ramírez R. (2009b)

Statistical Evaluation of Classification Diagrams for Altered Igneous Rocks. *Turkish Journal of Earth Sciences* (enviado: 22 de enero, 2009; aceptado con cambios menores: 3 de Julio, 2009; revisado: 15 de julio, 2009; disponible en línea de la revista: 14 de agosto, 2009).

Verma S.P., Besch T., Guevara M. and Schulz-Dobrich B. (1992)

Determination of twelve trace elements in twenty-seven and ten major elements in twenty-three geochemical reference samples by X-Ray fluorescence spectrometry. *Geostandards Newsletter*, 16, 301-309.

Verma S.P., Díaz-González L., Sánchez-Upton P. and 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, 997-1002.

Winchester J.A. and Floyd P.A. (1976)

Geochemical magma type discrimination: application to altered and metamorphosed basic igneous rocks. *Earth and Planetary Science Letters*, 28, 459-469.

Winchester J.A. and Floyd P.A. (1977)

Geochemical discrimination of different magma series and their differentiation products using immobile elements. *Chemical Geology*, 20, 325-343.

APÉNDICE 1

Verma, S.P, Díaz-González, L., and González Ramírez R. (2008)

Relative efficiency of single-outlier discordancy tests for processing geochemical data on reference materials and application to instrumental calibrations by a weighted least-squares linear regression model. *Geostandards and Geoanalytical Research*, 33 (1), 29-48.

Relative Efficiency of Single-Outlier Discordancy Tests for Processing Geochemical Data on Reference Materials and Application to Instrumental Calibrations by a Weighted Least-Squares Linear Regression Model

Surendra P. Verma*, Lorena Diaz-González and Rosalinda González-Ramírez

Departamento de Sistemas Energéticos, Centro de Investigación en Energía, Universidad Nacional Autónoma de México, Priv. Xochicalco s/no, Col Centro, Apartado Postal 34, Temixco 62580, Mexico

* Corresponding author e-mail: spv@cie.unam.mx

Numerous studies report geochemical data on reference materials (RMs) processed by outlier-based methods that use univariate discordancy tests. However, the relative efficiency of the discordancy tests is not precisely known. We used an extensive geochemical database for thirty-five RMs from four countries (Canada, Japan, South Africa and USA) to empirically evaluate the performance of nine single-outlier tests with thirteen test variants. It appears that the kurtosis test (N15) is the most powerful test for detecting discordant outliers in such geochemical RM databases and is closely followed by the Grubbs type tests (N1 and N4) and the skewness test (N14). The Dixon-type tests (N7, N8, N9 and N10) as well as the Grubbs type test (N2) depicted smaller global relative efficiency criterion values for the detection of outlying observations in this extensive database. Upper discordant outliers were more common than the lower discordant outliers, implying that positively skewed inter-laboratory geochemical datasets are more frequent than negatively skewed ones and that the median, a robust central tendency indicator is likely to be biased especially for small-sized samples. Our outlier-based procedure should be useful for objectively identifying discordant outliers in many fields of science and engineering and for interpreting them accordingly. After processing these databases by single-outlier discordancy tests and obtaining reliable estimates of central tendency and dispersion parameters of the geochemical data for the RMs in our database, we used these statistical data to apply a weighted least-squares linear regression (WLR) model for the major element determinations by X-ray fluorescence spectrometry and compared the WLR results with an ordinary

De nombreuses études présentent des données géochimiques sur les matériaux de références, données traitées par des méthodes statistiques éliminant des valeurs aberrantes avec des tests de discordance univariés. Mais l'efficacité de ces tests de discordance est mal connue. Nous avons travaillé sur une base de données géochimiques importante, reposant sur trente cinq RM provenant de quatre pays (Canada, Japon, South Africa et USA) afin d'évaluer empiriquement les performances de neuf tests et treize variations de ces neuf tests. Le test kurtosis (N15) se révèle être le test le plus puissant pour détecter les valeurs aberrantes dans cette base de données, suivi de près par les tests de type Grubbs (N1 et N4) et par le test sur la dissymétrie (N14). Les tests de type Dixon (N7, N8, N9 et N10) ainsi que le test de type Grubbs N2 donnent de plus petites valeurs du critère d'efficacité globale relative dans la détection de l'allure générale de cette importante base de données. Les valeurs aberrantes supérieures étaient plus fréquentes que les valeurs aberrantes inférieures: ceci implique que les données géochimiques inter-laboratoires biaisées vers le haut sont plus fréquentes que celles qui sont biaisées vers le bas et que la médiane, indicateur robuste d'une répartition centré, a des chances d'être biaisée tout spécialement pour les échantillons de petite taille. Notre protocole basé sur les valeurs aberrantes devrait être utile dans la recherche objective de valeurs aberrantes dans de nombreux domaines scientifiques et en ingénierie et leur interprétation correcte. Après avoir traité ces bases de données avec les tests de valeurs aberrantes discordantes et obtenu des estimations fiables de la tendance centrale et des paramètres de dispersion d'es

least-squares linear regression model. An advantage in using our outlier procedure and the new concentration values and uncertainty estimates for these RMs was clearly established.

Keywords: reference materials, robust methods, outlier-based methods, consecutive tests, Dixon tests, Grubbs tests, skewness, kurtosis, X-ray fluorescence spectrometry.

Received 22 Oct 07 — Accepted 11 Jul 08

Reference materials (RMs) are routinely used for quality control purposes in all sciences including the Earth sciences (e.g. Potts and Kane 1992, Guevailler et al. 1996, Dybczynski et al. 1998, Mahwar et al. 1998, Namiesnik and Zygmunt 1999, Vendemiato and Erzweiler 2001, Gabrievská et al. 2006, Santoyo et al. 2006, Hayes et al. 2007, Verma et al. 2008, Verma and Quiroz-Ruiz 2008). Traditionally, the calibration of an analytical instrument has been achieved through an ordinary least-squares linear regression (OLR) model. However, recently there has been an increasing trend to use weighted least-squares linear regression (WLR) models (Baumann 1997, Santoyo and Verma 2003, Sayago et al. 2004, Guevara et al. 2005, Aspero and González 2007, Staliopoulos and Stickel 2007, Tellinghuisen 2007), for which reliable estimates of both central tendency and dispersion parameters should be available for all RMs to be used in a given calibration.

To obtain these statistical estimates, inter-laboratory geochemical data on RMs have been processed traditionally by both robust and outlier-based methods. Robust methods have been used for processing RMs from France, but after the elimination of the so-called "gross" outliers (e.g., Govindaraju and Roelandts 1989). Inter-laboratory data for RMs from the USA (e.g., Gladney et al. 1992) and Japan (e.g., Imai et al. 1995) were processed by an outlier-based scheme using the so-called "two standard deviation" method (see Barnett and Lewis 1994, Verma 1998, and Verma et al. 2008 for more details on this class of statistical method). Similarly, the data for RMs from the International Atomic Energy Agency (IAEA) have also been evaluated

données géochimiques de notre base de données, nous avons utilisé les données statistiques pour appliquer un modèle de régression linéaire des moindres carrés pondérés (WLR) pour analyser les éléments majeurs par spectrométrie de fluorescence X et nous avons comparé ces résultats avec ceux obtenus par un modèle de régression linéaire de moindres carrés ordinaire (OLR). On montre finalement clairement le bénéfice obtenu avec notre procédure basée sur les valeurs aberrantes ainsi que l'intérêt des nouvelles estimations des concentrations des RM et de leurs incertitudes.

Mots-clés : matériau de référence, méthodes robustes, méthodes basées sur les valeurs aberrantes, tests consécutifs, tests de Dixon, tests de Grubbs, dissymétrie, kurtosis, spectrométrie de fluorescence X.

through some kind of the outlier-based scheme (e.g., Dybczynski et al. 1979, Verma 2004, Villeneuve et al. 2004). The results of the robust methods for microgabro RM-S and dolerite WS-E (Govindaraju et al. 1994) were compared with the outlier-based methods by Verma (1997) and Verma et al. (1998), respectively. The univariate discordancy tests, being an integral part of the outlier-based methods, have also been applied for processing RM databases by Guevara et al. (2001) and Velasco-Tapia et al. (2001).

According to the ISO Guide 35 (ISO 1989), however, an elaborate procedure involving several rounds of data collection from several laboratories should be carried out in order to certify RMs, and the detection and elimination of discordant outliers is not particularly encouraged. Nevertheless, for older RMs, such as those dealt with in the present work, the statistical method of detection and elimination of discordant outliers may be the only one available to "correctly" evaluate them. This is because the statistically correct application of some robust methods, such as the trimmed mean or Winsorised mean, requires the distribution of these outliers to be known a priori, and other methods, such as the median, trimean or Gastwirth's mean, demand a symmetrical distribution of the outlying observations (Barnett and Lewis 1994, Verma 2005). Furthermore, the outlier procedures can be used for other kinds of experimental data in many fields of science and engineering (Verma and Quiroz-Ruiz 2006a, b). The statistically correct identification of these discordant outliers will be useful for a better interpretation of the experimental data.

On the other hand, the robust method approach practiced for obtaining the central tendency parameter in some older RMs also relies on the standard deviation as the dispersion parameter (e.g., Govindaraju and Roelandts 1989), whose unbiased estimate very much depends on the absence of discordant outlier(s) that, if present, can seriously distort this estimate (Barnett and Lewis 1994, Verma 1997, 2005, Verma *et al.* 2008). For the purpose of identifying such discordant outliers, therefore, the present methodology, initially proposed by Verma (1997), but much improved by the availability of new, precise critical values (e.g., Verma *et al.* 2008), can now be advantageously applied.

The discordancy tests (single- as well as multiple-outlier types) originally proposed by both Dixon (1951) and Grubbs (1950, 1969) have been very popular and are still in wide use in different fields (e.g., Barnett and Lewis 1994, Li *et al.* 2003, Serbst *et al.* 2003, Forns *et al.* 2006, Gabrovská *et al.* 2006, Sang *et al.* 2006, Verma and Quiroz-Ruiz 2006a, b, 2008, Hayes *et al.* 2007, Verma *et al.* 2008). Several other discordancy tests are also available for this purpose, most notable among which are the skewness and kurtosis tests (Barnett and Lewis 1994, Verma 1997, Velasco and Verma 1998, Verma and Quiroz-Ruiz 2006a, b, 2008).

Velasco *et al.* (2000) compare the performance or efficiency of outlier tests using the rare earth element (REE) data for twenty-six RMs. However, at that time because of the unavailability of critical values for certain sample sizes (e.g., for $n > 30$ for Dixon tests N7 to N10, Dixon 1951; see Verma and Quiroz-Ruiz 2006a, or $n > 20$ for test N2; see Barnett and Lewis 1994, Verma and Quiroz-Ruiz 2006b; n being the total number of observations in a given univariate dataset), all tests could not be applied to all cases in their database by Velasco *et al.* (2000). This seriously limited the performance evaluation of the outlier tests.

The relative efficiency of these tests is important for deciding if the application of a smaller number of tests would suffice in an outlier-based scheme, or if all the available tests should be applied as originally suggested by Verma (1997). In this work we present the results of our empirical evaluation of the relative efficiency of all nine presently available single-outlier tests for normal univariate data with thirteen test variants (Barnett and Lewis 1994, Verma *et al.* 2008) using an extensive database (i.e., data for all major and trace elements, including the REE, in thirty-five RMs). It has been possible to use all RM data because precise and

accurate critical values (or percentage points) are now available for these tests for all values of n from 3, 4 or 5 (depending on the type of statistics) up to 30,000 (Verma and Quiroz-Ruiz 2006a, b, 2008, Verma *et al.* 2008). We also illustrate the advantage of using these central tendency and dispersion parameters derived from the present statistical methodology for the calibration of X-ray fluorescence spectrometry through a WLR regression model as compared to an OLR model commonly practiced in most analytical laboratories.

Database, tests evaluated and their categorisation

The data for all major and trace elements, along with the analytical techniques (classified according to Velasco-Tapia *et al.* 2001), were compiled for a total of thirty-five RMs from four countries: (i) Canada (Abbey (1979) and Gladney and Roelandts (1990) for MRG-1, SY-2 and SY-3); (ii) Japan (<http://riodb02.jbase.aist.go.jp/earthsci/welcome.html> for JA-1, JA-2, JA-3, JB-1, JB-1a, JB-2, JB-3, JF-1, JF-2, JG-1, JG-1a, JG-2, JG-3, JH-1, JP-1, JR-1, and JR-2); (iii) South Africa (Steele *et al.* (1972, 1978) for NIM-D, NIM-G, NIM-L, NIM-N, NIM-P and NIM-S); and (iv) USA (Flanagan (1986) for GSM-1; Gladney (1988) for BIR-1 and W-2; Gladney and Roelandts (1988) for BHVO-1, GLO-1 and RGM-1; Gladney *et al.* (1991) for DFS-1 and W-1; Gladney *et al.* (1992) for G-2).

The nine single-outlier tests with thirteen variants (Table 1) evaluated in this work are: N1 (N1_Upper and N1_Lower), N2 (N2_Extreme), N4 (N4k = 1_Upper and N4k = 1_Lower), N7 (N7_Upper), N8 (N8_Extreme), N9 (N9_Upper and N9_Lower), N10 (N10_Upper and N10_Lower), N14 (skewness N14_Extreme) and N15 (kurtosis N15_Extreme).

A "finite" statistical sample drawn from a normal population can be designated as: $x_1, x_2, x_3, \dots, x_n, x_{n-1}, x_n$. If these observations (e.g., analytical data) are arranged in an ascending order, the new array may be called $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n-2)}, x_{(n-1)}, x_{(n)}$. The statistical samples of analytical data for a given element in a RM were kept separately according to the analytical method used to determine them; these methods were categorised into eight method-groups following Velasco-Tapia *et al.* (2001). All of these individual cases of statistical samples constituted the "applicable" cases in our database of thirty-five RMs. For a particular chemical element in a RM, these statistical samples arising from different analytical methods were combined into a

Table 1.
Nine single-outlier discordancy tests with thirteen test variants for univariate normal samples (modified after Barnett and Lewis 1994, Verma and Quiroz-Ruiz 2006a, b)

Test code*	Value tested	Test statistic	Test significance	Applicability of test $n_{min} - n_{max}$		
				Literature pre-2006 (less precise critical values) ^a	Literature 2006 (more precise critical values) ^b	Literature 2008 (most precise critical values) ^c
N1	Upper $X_{(n)}$	$DN1_{(u)} = (x_{(n)} - \bar{x}) / s$	Greater	3-147	3-100	3-30,000
	Lower $X_{(1)}$	$DN1_{(l)} = (\bar{x} - x_{(1)}) / s$	Greater	3-147	3-100	3-30,000
N2 (two-sided)	Extreme $X_{(n)}$ or $X_{(1)}$	$DN2 = \text{Max} : \{ x_{(n)} - \bar{x} / s, (\bar{x} - x_{(1)}) / s \}$	Greater	3-20	3-100	3-30,000
N4	k=1 Upper $X_{(k)}$	$DN4_{(u)} = S^*_{(k)} / S^*$	Smaller	3-50	3-100	3-30,000
	k=1 Lower $X_{(k)}$	$DN4_{(l)} = S^*_{(k)} / S^*$	Smaller	3-50	3-100	3-30,000
N7	Upper $X_{(n)}$	$DN7 = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(1)})$	Greater	3-30	3-100	3-30,000
N8 (two-sided)	Extreme $X_{(n)}$ or $X_{(1)}$	$DN8 = \text{Max} : \{ x_{(n)} - x_{(n-1)} / (x_{(n)} - x_{(1)}), (x_{(n)} - x_{(1)}) / (x_{(n)} - x_{(1)}) \}$	Greater	4-30	4-100	4-30,000
N9	Upper $X_{(n)}$	$DN9_{(u)} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(1)})$	Greater	4-30	4-100	4-30,000
	Lower $X_{(1)}$	$DN9_{(l)} = (x_{(n)} - x_{(1)}) / (x_{(n-1)} - x_{(1)})$	Greater	4-30	4-100	4-30,000
N10	Upper $X_{(n)}$	$DN10_{(u)} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(1)})$	Greater	5-30	5-100	5-30,000
	Lower $X_{(1)}$	$DN10_{(l)} = (x_{(n)} - x_{(1)}) / (x_{(n-1)} - x_{(1)})$	Greater	5-30	5-100	5-30,000
N14	Extreme $X_{(n)}$ or $X_{(1)}$	$DN14 = [n^{1/2} \{ 2x_{(n)} (x_{(n)} - 2\bar{x}) / \{ 2x_{(n)} (x_{(n)} - 2\bar{x}) \} \}]$	Greater	5-1000	5-100	5-30,000
N15	Extreme $X_{(n)}$ or $X_{(1)}$	$DN15 = [n \{ 2x_{(n)} (x_{(n)} - 2\bar{x}) / \{ 2x_{(n)} (x_{(n)} - 2\bar{x}) \} \}]$	Greater	5-1000	5-100	5-30,000

* Test code (N series) is from Barnett and Lewis (1994); with N1, N2, and N4 are Grubbs type; N7 to N10 are Dixon type; N14 and N15 are respectively the skewness and kurtosis tests (both one two-sided). Other tests that are not identified as two-sided were classified as one-sided by Barnett and Lewis (1994). The symbols for test statistics $DN1_{(u)}$, $DN1_{(l)}$, $DN2$, etc. were proposed by Verma (2005) and used by Verma and Quiroz-Ruiz (2006a, b). The subscripts u and l are the upper (the highest) and lower (the lowest) observations, respectively. The test statistics are self explanatory except the statistics of the type n -sided sum of squares χ^2 'total sum of squares', for example, $S^*_{(k)} / S^*$ for test N4, $k = 1$, proposed by Grubbs (1939, 1949), which need some explanation. For an ordered array $X_{(1)} < X_{(2)} < \dots < X_{(n)}$, the S^* term is calculated using all data $S^* = \sum_{i=1}^n (x_{(i)} - \bar{x})^2$, where \bar{x} is the arithmetic mean ($\bar{x} = (\sum_{i=1}^n x_{(i)}) / n$), whereas $S^*_{(k)}$ is computed from the $(n-1)$ remaining data $X_{(1)}, X_{(2)}, \dots, X_{(n-1)}$, after eliminating the highest datum to be tested $X_{(n)}$ (use the subscript u in the term $S^*_{(u)}$) as follows: $S^*_{(u)} = \sum_{i=1}^{n-1} (x_{(i)} - \bar{x}_{(n-1)})^2$ where $\bar{x}_{(n-1)} = (\sum_{i=1}^{n-1} x_{(i)}) / (n-1)$.

^a For literature values see books by Barnett and Lewis (1994) and Verma (2005).
^b Verma and Quiroz-Ruiz (2006a, b) increased n_{max} to 100 by involving more precise and accurate critical values for all discordancy tests.
^c Verma et al. (2008) and Verma and Quiroz-Ruiz (2008) increased n_{max} to 30,000 for all discordancy tests 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 those presented by Verma and Quiroz-Ruiz (2006a, b).

Table 2.
Categorisation used for performance comparison of nine single-outlier discordancy tests with thirteen test variants for univariate normal samples

Test categorisation		Number of tests / test variants categorised	Test code *
#	Type		
Test category 1	Upper-outlier tests	5 / 5	N1_Upper, N4k = 1_Upper, N7_Upper and N10_Upper
Test category 2	Lower-outlier tests	4 / 4	N1_Lower, N4k = 1_Lower, N9_Lower and N10_Lower
Test category 3	Extreme-outlier tests	4 / 4	N2_Extreme, N8_Extreme, N14_Extreme and N15_Extreme
Test category 4	Combined Upper-Lower outlier test	4 / 8	N1, N4, N9 and N10 (Upper and Lower for all tests)
Test category 5	All single-outlier tests	9 / 13	N1, N2, N4, N7, N8, N9, N10, N14 and N15 (with Upper and Lower types if available)

* Note Test category 1 (5 tests or test variants) and Test category 2 (4 tests or test variants) include one-sided tests whereas Test category 3 shows two-sided tests (4 tests or test variants). Note the one-sided and two-sided classification of discordancy tests adopted here follows the recommendation of Barnett and Lewis (1994).

single statistical sample only after the discordancy tests (single-outlier type test, Table 1) and significance tests (ANOVA for three or more method-groups, or F and Student's *t* tests for two method-groups) demonstrated at the strict 99% confidence level that none of them had any discordant outliers and all were drawn from a single normal population or identical normal populations. On the contrary, the data from one or sometimes more analytical method-groups were not combined into this final statistical sample for estimating the central tendency and dispersion parameters.

The discordancy tests of a single-outlier type normally evaluate either an upper outlier $x_{(n)}$, e.g., test N1 (N1_Upper), a lower outlier $x_{(1)}$, e.g., test N1 (N1_Lower), or an extreme outlier $x_{(n)}$ or $x_{(1)}$ (whichever of the two is farther away from the central tendency parameter), e.g., test N2 (see Table 1), as a possible discordant outlier under a contamination model (more details on such models are given at the end of the next section).

The single-outlier discordancy tests can be applied consecutively to test as many outliers as a given test is capable of detecting as discordant (Barnett and Lewis 1994, Vema 1997, 1998, Vema et al. 2008). Thus, a given test, e.g., N1_Upper is applied to test the value $x_{(n)}$ and if this value is declared to be a discordant outlier in this first iteration ("iteration 1"), the same test is again applied (this is called consecutive testing by Barnett and Lewis 1994) to the remaining data for testing a second upper outlier; this represents "iteration 2". The iterations are thus numbered consecutively. The test is repeated until no more discordant outliers are identified.

For evaluating relative efficiencies of the single-outlier tests we classified these tests into five test categories as shown in Table 2. As an example, we cite the first category of tests (Test category 1), which included five tests: N1_Upper, N4k = 1_Upper, N7_Upper, N9_Upper and N10_Upper.

Test evaluation parameter: relative efficiency criterion (REC)

In the first iteration (iteration 1), the relative efficiency criterion (REC)₁ for a given test was defined as follows:

$$REC_1 = \frac{\text{number of "successful" cases in the first iteration}}{\text{number of "applicable" cases}} \times 100 \quad (1)$$

In the subsequent iterations (iteration 2 and onwards), the REC was defined as the general equation:

$$REC_i = \frac{\text{number of "successful" cases in the } i^{\text{th}} \text{ iteration}}{\text{number of "successful" cases in the } (i-1)^{\text{th}} \text{ iteration}} \times 100 \quad (2)$$

For example, for the second iteration (iteration 2), this means that (REC)₂ will be:

$$REC_2 = \frac{\text{number of "successful" cases in the second iteration}}{\text{number of "successful" cases in the first iteration}} \times 100 \quad (3)$$

The reason that the subsequent REC parameters were defined as above (e.g., equation 3) is that the "applicable" cases for the second iteration are the same as the "successful" cases for the first iteration.

Finally, the overall or global relative efficiency criterion (REC)_g for a given test was defined as follows:

$$REC_g = \frac{\sum_{i=1}^n \text{number of "successful" cases in an iteration}}{\text{initial number of "applicable" cases}} \times 100 \quad (4)$$

where the iteration number, varied from 1 to n (n being the final iteration considered for the REC calculations).

Thus, the overall REC parameter $(REC)_n$ takes into account the fact that a given test might identify a first outlier as discordant in iteration 1, but also a second one in iteration 2, and so on. Therefore, the number of discordant outliers should be summed up for this particular test to calculate $(REC)_n$ from equation (4), the number of initial "applicable" cases being the same as the "applicable" cases for the first iteration. Thus, if a given test detected an outlier in iteration 1 but failed to do so in iteration 2 for any of these "successful

cases, its overall $(REC)_n$ value would remain the same as that for iteration 1, i.e., for this particular test $(REC)_n = (REC)_1$. However, if a test also detected a discordant outlier in the second iteration, we should have for this test $(REC)_n > (REC)_1$.

For any REC to be statistically meaningful, we arbitrarily set a lower limit to the total number of applicable cases for each of the tests to be at least 30 in any given iteration, below which we did not calculate individual REC values for that particular iteration. We also note that theoretically the REC values are expected to be close to zero under the null hypothesis (H_0).

Table 3.
Database information applicable to define the relative efficiency criterion (REC) for sample size $n = 15$ data

#	Applicable cases			Successful cases (at 99% confidence level)		
				N1U	N1L	N2
1	BHFO-1, Cr, Gr4	JB-3, U, Gr6	NRO-1, Co, Gr8	DS-1, Al ₂ O ₃ , Gr1	JA-2, Ta, Gr5	DS-1, Al ₂ O ₃ , Gr1
2	BIR-1, Fe ₂ O ₃ , Gr5	JB-3, Fe ₂ O ₃ , Gr8	NRM-D, Cu, Gr2	DS-1, Pb, Gr2	JB-2, Ba, Gr6 †	DS-1, Pb, Gr2
3	DS-1, Al ₂ O ₃ , Gr1	JF-1, U, Gr5	NRM-D, Zn, Gr2	DS-1, Cr, Gr3 †	JG-1a, Lu, Gr6	DS-1, Cr, Gr3 †
4	DS-1, Pb, Gr2	JG-1a, MnO, Gr2	NRM-D, Ni, Gr2	JA-2, Ga, Gr3	JG-1a, Pr, Gr6	JA-2, Ga, Gr3
5	DS-1, Cr, Gr3	JG-1a, Na ₂ O, Gr2	NRM-D, MnO, Gr8	JB-2, Zn, Gr2	JF-1, Co, Gr5 †	JA-2, Ta, Gr5
6	DS-1, H ₂ OP, Gr8	JG-1a, Co, Gr5	NRM-G, Sr, Gr3	JB-2, Ga, Gr3	JF-1, MnO, Gr2 †	JB-2, Zn, Gr2
7	G-2, K ₂ O, Gr1	JG-1a, Ca, Gr5	NRM-L, Ba, Gr2	JB-2, Tb, Gr6 †	NRM-L, MnO, Gr8	JB-2, Ga, Gr3
8	G-2, TiO ₂ , Gr2	JG-1a, Eu, Gr5	NRM-L, MnO, Gr8	JB-3, Fe ₂ O ₃ , Gr8	NRM-S, Rb, Gr3	JB-2, Ba, Gr6 †
9	G-2, MgO, Gr4	JG-1a, Hf, Gr5	NRM-P, Sr, Gr2	JG-1a, Na ₂ O, Gr2	W-1, Ba, Gr2	JB-2, Tb, Gr6 †
10	G-2, Yb, Gr4	JG-1a, Lu, Gr5	NRM-P, Ni, Gr2	JG-1a, Hf, Gr5	W-1, K ₂ O, Gr4 ‡	JG-1a, Na ₂ O, Gr2
11	G-2, H ₂ OP, Gr8	JG-1a, Lu, Gr5	NRM-P, MnO, Gr8	JG-1, Rb, Gr6	W-2, Hf, Gr5	JG-1a, Hf, Gr5
12	JA-1, Na ₂ O, Gr5	JG-1a, Sm, Gr5	NRM-S, Rb, Gr3	NRM-D, Cu, Gr2	-	JG-1a, Lu, Gr6
13	JA-2, K ₂ O, Gr2	JG-1a, Dy, Gr6	NRM-S, Sr, Gr3	NRM-D, Zn, Gr2	-	JG-1a, Pr, Gr6
14	JA-2, MnO, Gr2	JG-1a, Er, Gr6	ST-2, Fe ₂ O ₃ , Gr2	NRM-G, Sr, Gr3	-	JG-1, Rb, Gr6
15	JA-2, Ga, Gr3	JG-1a, Eu, Gr6	ST-2, MnO, Gr3	NRM-P, Sr, Gr2 †	-	JF-1, Co, Gr5 †
16	JA-2, Eu, Gr5	JG-1a, Ho, Gr6	ST-3, Al ₂ O ₃ , Gr1	NRM-P, Ni, Gr3	-	JF-1, MnO, Gr2
17	JA-2, La, Gr5	JG-1a, Lu, Gr6	ST-3, Rb, Gr3	NRM-S, Sr, Gr3 ‡	-	NRM-D, Cu, Gr2
18	JA-2, Na ₂ O, Gr5	JG-1a, Lu, Gr6	ST-3, Sr, Gr8	ST-3, Sr, Gr8	-	NRM-G, Sr, Gr3
19	JA-2, Sm, Gr5	JG-1a, Pr, Gr6	W-1, Ba, Gr2	W-1, SiO ₂ , Gr4	-	NRM-L, MnO, Gr8 †
20	JA-2, Ta, Gr5	JG-1a, Tb, Gr6	W-1, Y, Gr3	W-2, Al ₂ O ₃ , Gr3	-	NRM-P, Sr, Gr2 †
21	JB-1a, Ca, Gr5	JG-1, Na ₂ O, Gr1	W-1, K ₂ O, Gr4	-	-	NRM-P, Ni, Gr3
22	JB-1a, Co, Gr5	JG-1, U, Gr2	W-1, SiO ₂ , Gr4	-	-	NRM-S, Rb, Gr3
23	JB-1a, Eu, Gr5	JG-1, Rb, Gr6	W-1, Al ₂ O ₃ , Gr5	-	-	NRM-S, Sr, Gr3 ‡
24	JB-1a, Lu, Gr5	JG-2, MnO, Gr2	W-1, Cu, Gr5	-	-	ST-3, Sr, Gr8
25	JB-1a, Sc, Gr5	JG-2, Na ₂ O, Gr2	W-1, Ho, Gr5	-	-	W-1, Ba, Gr2
26	JB-1a, Ta, Gr5	JG-2, Co, Gr5	W-1, Tm, Gr5	-	-	W-1, K ₂ O, Gr4 ‡
27	JB-1, Fe ₂ O ₃ , Gr5	JG-2, Er, Gr5	W-1, Er, Gr6	-	-	W-1, SiO ₂ , Gr4
28	JB-1a, Fe ₂ O ₃ , Gr8	JG-2, Sm, Gr5	W-2, Al ₂ O ₃ , Gr3	-	-	W-2, Al ₂ O ₃ , Gr3
29	JB-1, Pb, Gr2	JG-2, La, Gr5	W-2, CoO, Gr3	-	-	W-2, Hf, Gr5
30	JB-1, Nd, Gr5	JF-1, K ₂ O, Gr3	W-2, K ₂ O, Gr3	-	-	-
31	JB-1, Ta, Gr5	JF-1, Co, Gr5	W-2, Sr, Gr3	-	-	-
32	JB-1, Cd, Gr6	JF-1, MnO, Gr2	W-2, Ce, Gr5	-	-	-
33	JB-2, Zn, Gr2	JF-1, Eu, Gr5	W-2, Fe ₂ O ₃ , Gr5	-	-	-
34	JB-2, Ga, Gr3	JF-1, Na ₂ O, Gr5	W-2, Hf, Gr5	-	-	-
35	JB-2, Ba, Gr6	JF-1, Ce, Gr6	-	-	-	-
36	JB-2, Tb, Gr6	JF-2, Yb, Gr5	-	-	-	-

The abbreviations for method-groups (Gr1 to Gr8) are: Gr1 - classical methods; Gr2 - atomic absorption methods; Gr3 - X-ray fluorescence methods; Gr4 - emission spectrometry methods; Gr5 - nuclear methods; Gr6 - mass spectrometry methods; Gr7 - chromatography methods; Gr8 - miscellaneous methods. For more details see Velasco-Tapia et al. (2001). The application of these and other single-outlier tests to some "successful" cases also showed a second (identified by †) and a third discordant outlier (identified by ‡).

that all samples were drawn from a normal population without any "statistical" contamination. The alternate hypothesis (H_1) being tested is related to a suitable contamination model, according to which an extreme observation came from a different distribution (shifted mean, greater standard deviation or both) than the remaining observations in a given sample. Barnett and Lewis (1994) is an excellent source for more details on

these hypotheses and models. Finally, we emphasise that, following Verma (1998), the discordancy tests were applied to the statistical samples derived from a single analytical method-group at the strict confidence level of 99%.

An example of the REC calculations: We illustrate the meaning and calculation of the REC parameter for

Table 4. Individual concentration data to show the discordant outliers (application of tests N1U, N1L and N2 only at 99% confidence level) for sample size $n = 15$ data, along with the statistical information of these data as well as for all data (for the five elements in these five samples)

File name (ReferenceMaterial, Element, Methodgroup): Concentration data (major in %m/m, trace in $\mu\text{g g}^{-1}$)					
#	BHVO-1, Cr, Gr4 (No discordant outlier by example tests)	JG-1a, Eu, Gr6 (No discordant outlier by example tests)	DTS-1, Al_2O_3 , Gr1 (One upper discordant outlier by example tests)	JA-2, Ta, Gr5 (One lower discordant outlier by example tests)	JB-3, P_2O_5 , Gr8 (One upper discordant outlier by example tests)
1	230	0.309	0.130	0.10 (N1L, N2)	0.239
2	341	0.49	0.212	0.58	0.253
3	252	0.60	0.227	0.61	0.27
4	257	0.64	0.227	0.61	0.28
5	260	0.65	0.246	0.66	0.28
6	285	0.65	0.265	0.66	0.28
7	285	0.65	0.283	0.70	0.28
8	294	0.67	0.302	0.88	0.28
9	306	0.70	0.456	0.88	0.287
10	315	0.70	0.453	0.88	0.29
11	315.4	0.73	0.567	0.89	0.29
12	321.7	0.76	0.605	0.898	0.29
13	322	0.81	0.624	0.90	0.293
14	328	0.87	0.737	0.98	0.3
15	340	0.88	1.479 (N1U, N2)	0.98	0.35 (N1U)
Statistical information (these data)					
n_{obs}	15	15	15	15	15
\bar{x}_{obs}	290.1	0.674	0.451	0.747	0.2841
s_{obs}	34.8	0.143	0.338	0.228	0.024
$\%O_{\text{obs}}$	0	0	1	1	1
$\%O_{\text{el}}$	0	0	6.67	6.7	6.7
n_{el}	15	15	14	14	14
\bar{x}_{el}	290.1	0.674	0.378	0.793	0.2794
s_{el}	34.8	0.143	0.189	0.147	0.0162
$CI_{99\%}$	263.4-316.9	0.564-0.784	0.226-0.530	0.675-0.912	0.2664-0.2925
Statistical information (all data)					
n_{el}	47	39	80	22	65
\bar{x}_{el}	278.9	0.692	0.371	0.863	0.2836
s_{el}	63.7	0.094	0.254	0.425	0.0415
O_{el}	3	1	15	4	7
$\%O_{\text{el}}$	6.4	2.6	18.8	18.2	10.8
n_{el}	44	38	65	18	58
\bar{x}_{el}	292	0.702	0.269	0.764	0.2869
s_{el}	37.2	0.07	0.107	0.143	0.0231
$CI_{99\%}$	276.9-307.2	0.671-0.733	0.234-0.305	0.666-0.862	0.2788-0.2950

For method-groups see footnote Table 3. The statistical symbols are: n - number of analytical data; \bar{x} - mean; s - standard deviation; O_{el} - number of discordant outliers; $\%O_{\text{el}} = (O_{\text{el}}/n_{\text{el}})$; $100 \cdot CI_{99\%}$ - 99% confidence interval of the mean; the subscript obs and el refer to the statistics before and after, respectively, the process of discordant outlier detection and elimination; and the subscript el for statistical parameters refers to these data (tabulated above) for a particular method-groups in our database. For the complete data (combined data for all method-groups, the results are summarized after the application of all single-outlier tests. Note the ANOVA significance test was applied before combining the data from different method-groups (see text for more details). In statistical information one extra digit was purposely retained (see Berrington and Robinson 2003, for details).

a particular example of sample size $n = 15$ in our database. First, all cases of this particular sample size were identified in the database (see Table 3 for a synthesis of these sets). Thus, a total of 106 "applicable" cases existed for $n = 15$ in our database (Table 3); for example, the first "applicable" case listed in Table 3 is BHVO-1_Cr_Gr4, i.e., there were fifteen data for Cr (see Table 4 for the listing of these individual data) in BHVO-1 by analytical method-group # 4 - emission spectrometry methods; similarly, the second applicable case is BR-1_Fe₂O₃_Gr5 (Fe₂O₃ data for BR-1 by method-group # 5 - nuclear methods), and so on. For illustrative purposes, Table 3 also reports explicitly the results of the application of just three discordancy test variants (see Table 1 for more details on these test variants; the other variants for tests N3 to N15 were similarly applied, but were not included in Table 3 for the sake of simplicity): N1U (test N1 for the upper outlier), N1L (test N1 for the lower outlier), and N2 (test N2 for an extreme outlier). Table 3 lists all cases in which one or sometimes more discordant outliers were detected in 106 "applicable" cases; these cases were termed the "successful" cases for a particular discordancy test. Thus, during iteration 1 tests N1U, N1L and N2 detected one outlier in 20, 11, and 29 cases (Table 3), and using equation (1) above, the corresponding REC values (REC)₁ would be about 18.9, 10.4 and 27.4, respectively. The application of these and other single-outlier tests showed the presence of a discordant second outlier in some of these "successful" cases (identified by † in Table 3) and also a third discordant outlier in still fewer cases (identified by ‡, Table 3). For the cases with only one discordant outlier present, i.e., those cases not identified by † or ‡ in any of the columns corresponding to the "successful" cases in Table 3, we note that (REC)₂ = (REC)₁, of which a few selected cases are discussed below for the %DiscordantOutliers parameter.

%DiscordantOutliers (%Ot) parameter

Note that the REC parameter is not synonymous with %DiscordantOutliers (%otd in Verma 1997, or %O_d in Velasco-Tapia et al. 2001). For a given analytical method-group, the (%Ot)_{CG} parameter for a given element in a RM can be defined as:

$$(\%O_t)_{CG} = \frac{\text{total number of discordant outliers in an "applicable" case}}{\text{total number of initial data in that "applicable" case}} \times 100 \quad (5)$$

For the complete dataset for a given element in a RM, the %Ot can be similarly defined as:

$$\%O_t = \frac{\text{total number of discordant outliers in a "combined" case}}{\text{total number of initial data in that "combined" case}} \times 100 \quad (6)$$

Here, a "combined" case refers to all analytical data of an element in a RM, obtained from all analytical methods.

The major distinction between the REC and %Ot parameters is that the former refers to all "applicable" cases whereas the latter is for one particular "applicable" case, either for a single method-group (here referred to as (%Ot)_{CG}) or for a "combined" case of all method-groups (simply called %Ot). The REC parameter can only be defined for a statistically significant number of "applicable" cases (equations 1 to 4) whereas the %Ot parameter is generally meaningful for one particular "applicable" or "combined" case (equations 5 and 6).

If we were to consider only the three single-outlier tests (N1U, N1L and N2), the (%Ot)_{CG} would be only about 6.7% for all those cases that are listed in any of the last three columns (under the heading "successful cases", Table 3) but are not identified by the symbols † and ‡. For the latter, because two or three outliers, respectively, were identified as discordant by the present single-outlier test method, the (%Ot)_{CG} would be about 13.3% or 20.0%, respectively.

For all other "applicable" cases (those cases that are listed in the first three columns but are not present in the other columns could be termed, on the contrary, as "unsuccessful" cases although we did not use this term), the (%Ot)_{CG} for each of them will be exactly zero if we were to consider only these three single-outlier tests (Table 3).

An example of %Ot calculations: We used selected cases from Table 3 as examples to illustrate the difference between (%Ot)_{CG} or %Ot. In Table 4 we present five cases selected from Table 3, for which the (%Ot)_{CG} was either 0% (BHVO-1_Cr_Gr4 and JG-1_Eu_Gr6) or about 6.7% (DTS-1_Al₂O₃_Gr1, JA-2_Ta_Gr5, and JB-3_P₂O₅_Gr8). Table 4 also provides the statistical information for these data (individually listed) as well as for the complete data (not individually listed) for all methods for these elements. For the latter the %Ot parameter varied from 2.6 to 18.8% (for all iterations - see the final lowest part of Table 4).

Results

Relative efficiency criterion (REC)

Test category 1 (single upper-outlier test): The results are presented in Figure 1. Note that in iteration

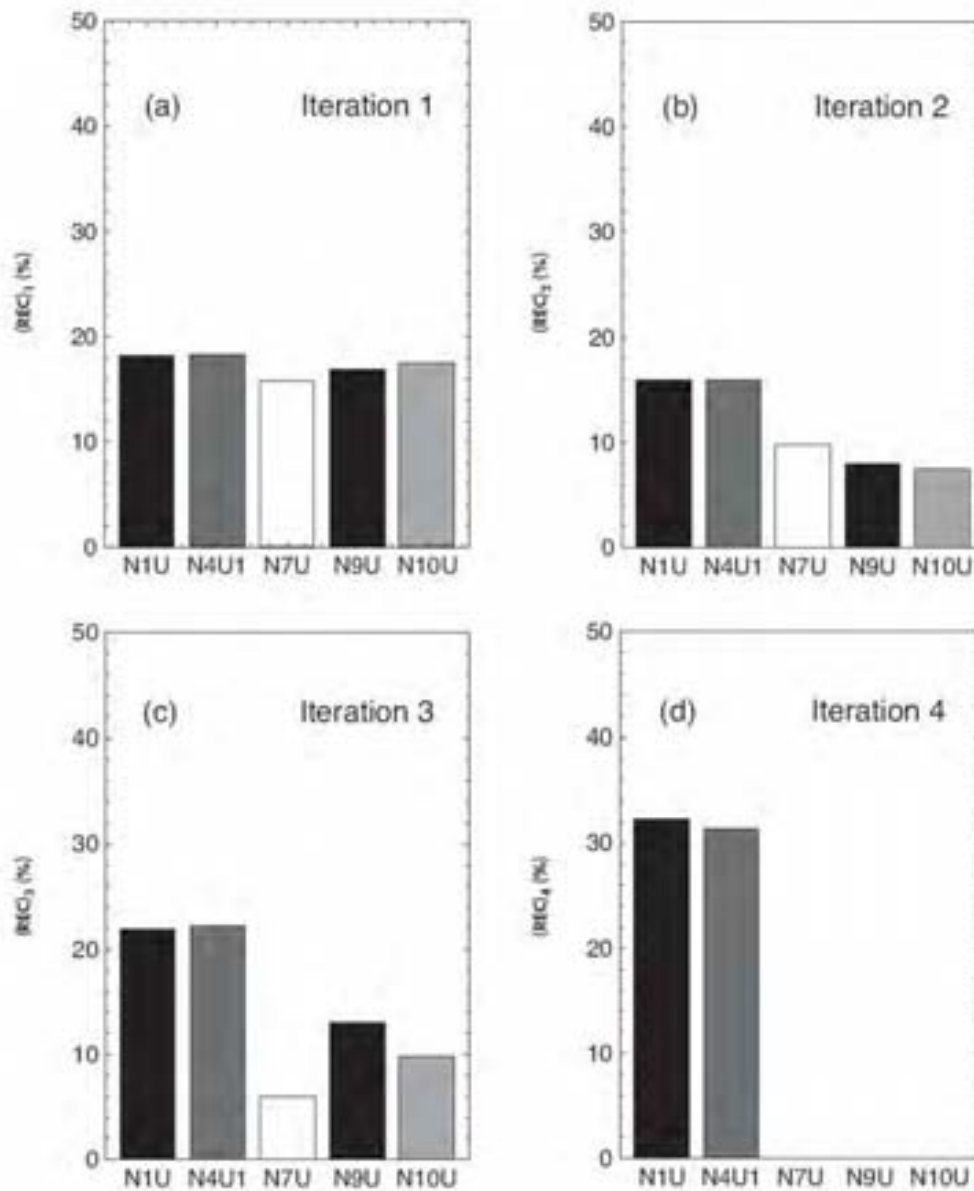


Figure 1. The relative efficiency criterion (REC) of the discordancy tests in Test category 1 (see Table 2 for categorisation). The five tests of the upper outlier type compared in Test category 1 are: N1U - N1 Upper, N4U1 - N4k = 1 Upper, N7U - N7 Upper, N9U - N9 Upper and N10U - N10 Upper. Note that in order to facilitate a better visual comparison of (REC)₁ to (REC)₄ values, the same y-scale (0-50%) was maintained throughout Figures 1-5. The number of cases to which a given test was applied is also presented explicitly in this and later figure captions. (a) Iteration 1 (5336 cases with sample sizes $n = 3$ for N1U, N4U1, and N7U; 4490 cases with sample sizes $n = 4$ for N9U; and 3889 cases with sample sizes $n = 5$ for N10U); (b) Iteration 2 (973 cases for N1U, 979 for N4U1, 846 for N7U, 759 for N9U, and 681 for N10U); and (c) Iteration 3 (155 cases for N1U, 157 for N4U1, 83 for N7U, 61 for N9U, and 51 for N10U).

If the tests were applied to a different number of cases depending on the minimum sample size applicable for a given test statistic (5336 cases of $n \geq 3$ for N1U, N4U1, and N7U; 4490 cases of $n \geq 4$ for N9U; and 3889 cases of $n \geq 5$ for N10U; see Table 1). We also emphasise that the $(REC)_1$ values were very similar (more importantly, they maintained the trends and differences between the $(REC)_1$ values as explained in this and the later sections) when all these tests were applied to the identical cases (3889 cases with sample sizes $n \geq 5$), implying that the overall $(REC)_1$ values did not strongly depend on whether all "applicable" cases (Iteration 1; Figure 1a) or the same cases were considered for all tests (the results of the latter are not shown graphically). This was also true for other test categories.

For Iteration 1 (Figure 1a), the Grubbs type tests (N1U and N4U1) showed slightly greater $(REC)_1$ values (18.2 and 18.3%) than the Dixon type tests (N7U, N9U and N10U; 15.9-17.5%). The $(REC)_2$ and $(REC)_3$ values (for iterations 2 and 3; Figure 1b, c) for

the Grubbs type tests (15.9-22.3%) were notably greater than the Dixon type tests (6.0-13.1%). For Iteration 4 (Figure 1d), a significant number (> 30) of "applicable" cases being the number of the "successful" cases for Iteration 3, persisted only for the Grubbs type tests (N1U and N4U1), for which the $(REC)_4$ values were also greater (31.4 and 32.4%) than for the earlier iterations. Discordant outliers were still detected for a few cases up to iteration 6 for these Grubbs type tests.

Test category 2 (single lower-outlier tests): The results are presented in Figure 2. For Iteration 1 (Figure 2a), the Grubbs type tests (N1L and N4L1) showed somewhat smaller $(REC)_1$ than the Dixon type tests (N9L and N10L; 9.1 and 9.7%). However, for Iteration 2 (Figure 2b), the $(REC)_2$ values for these two Grubbs type tests were significantly greater (10.9 and 11.5%) than those for the Dixon type tests (5.6 and 7.4%). For Iteration 3 (Figure 2c), similar $(REC)_3$ values (11.8 and 12.5%) were obtained for the Grubbs type tests (N1L and N4L1) whereas no outliers (0% REC) were detected by the Dixon type tests (N9L and N10L). Outliers

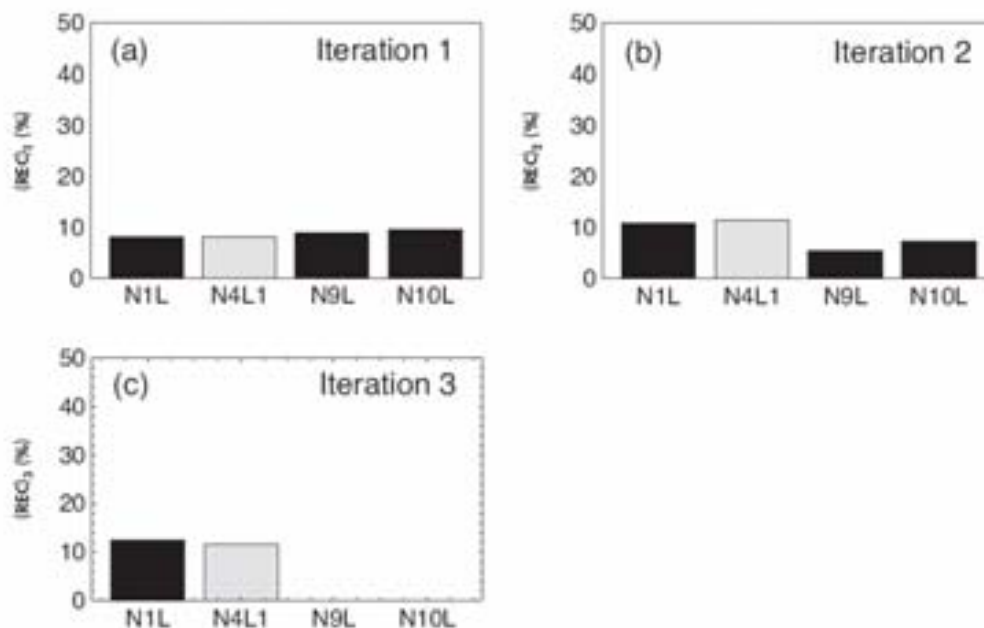


Figure 2. The relative efficiency criterion (REC) of the discordancy tests in Test category 2 (see Table 2 for categorisation). The four tests of the lower outlier type compared in Test category 2 are: N1L - N1 Lower, N4L1 - N4k = 1 Lower, N9L - N9 Lower and N10L - N10 Lower. (a) Iteration 1 (5336 cases for N1L and N4L1, 4490 for N9L; and 3889 for N10L); (b) Iteration 2 (439 cases for N1L, 444 for N4L1, 408 for N9L, and 377 for N10L); and (c) Iteration 3 (48 cases for N1L, 51 for N4L1, 23 for N9L, and 28 for N10L; note no outliers were detected as discordant by N9L and N10L in this iteration).

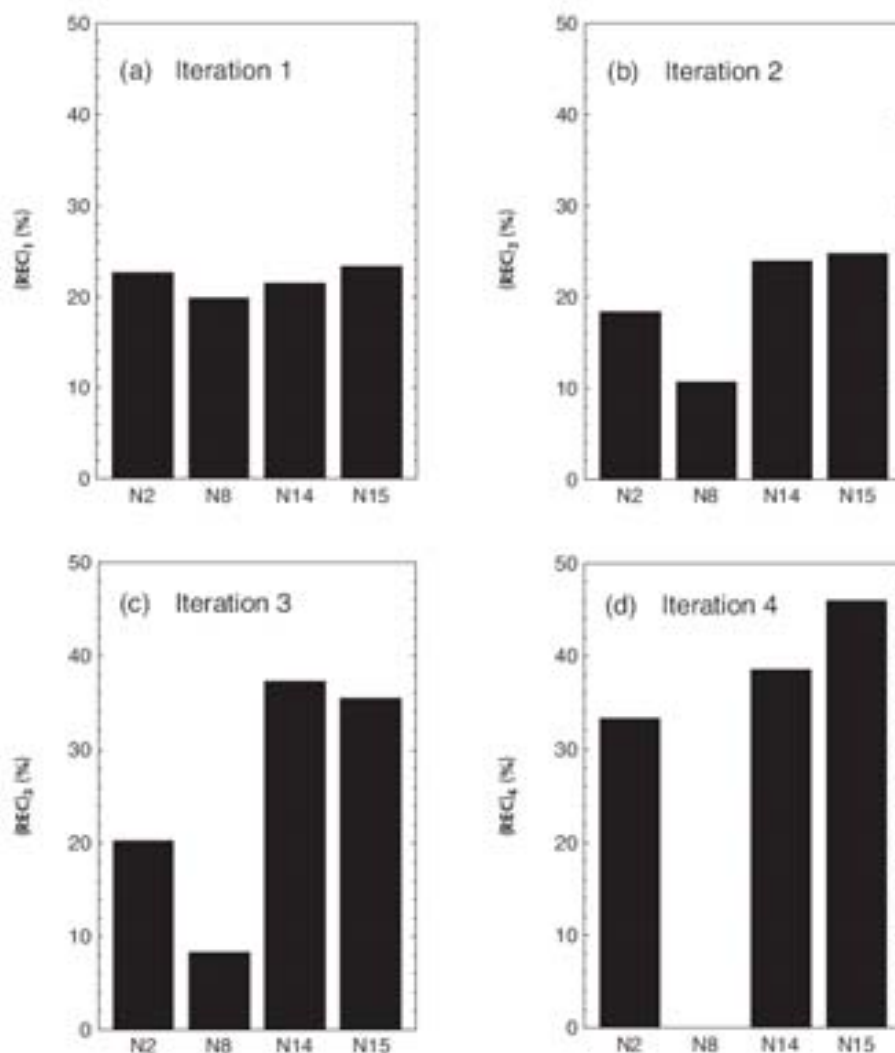


Figure 3. The relative efficiency criterion (REC) of the discordancy tests in Test category 3 (see Table 2 for categorisation). The four tests (two-sided) of the extreme outlier type compared in Test category 3 are: N2 (Grubbs type), N8 (Dixon type), N14 (skewness), and N15 (kurtosis). (a) Iteration 1 (5336 cases for N2, 4490 for N8, and 3889 for N14 and N15); (b) Iteration 2 (1210 cases for N2, 893 for N8, 838 for N14, and 988 for N15); (c) Iteration 3 (222 cases for N2, 96 for N8, 201 for N14, and 245 for N15); and (d) Iteration 4 (45 cases for N2, 1 for N8, 75 for N14, and 87 for N15). Iteration 5 (29 cases for N14 and 40 for N15) is not shown.

were detected as discordant for a few cases up to iteration 4 by N11 and N41).

Also note that for iteration 1, the $(REC)_1$ values (8.2-9.7%) were significantly smaller than those for the upper outlier versions of these tests (15.9-18.3%; see

Test category 1 above; compare Figures 2a and 1a). This is an interesting observation concerning the distribution of discordant outliers in geochemical data on RM4; discordant upper outliers (x_{ij}) are more common than discordant lower outliers (x_{i1}). In fact, the comparison of the performance of upper- and lower-outlier

versions of tests (Figures 1a-d and 2a-c; $(REC)_1$, $(REC)_2$ and $(REC)_3$ values of Test category 1 and Test category 2, respectively) suggests that positively skewed inter-laboratory geochemical datasets are more commonly present than negatively skewed ones. This implies that the robust estimates of central tendency, such as the median, the trimean and the Gastwirth's mean, are likely to be biased, especially for small-sized samples.

Test category 3 (single extreme-outlier tests): The results are presented in Figure 3. These tests are also classified by Barnett and Lewis (1994) as two-sided tests, whereas the tests in the earlier categories (Test category 1 and Test category 2) are known as one-sided. For iteration 1 (Figure 3a), the kurtosis test (N15) showed the greatest $(REC)_1$ value (25.4%), followed by the Grubbs type test (N2; 22.7%) and the skewness test (N14; 21.5%), with the Dixon type test (N8) presenting the smallest value (19.9%). For iterations 2 and 3 (Figure 3b, c), both the skewness and kurtosis tests (N14 and N15) showed very high $(REC)_2$ values (24.0-37.3%) as compared to the Grubbs type test (N2; 18.3 and 20.3%) and the Dixon type test (N8; 8.3 and 10.8%). For iteration 4 (Figure 3d), the Dixon type test (N8) was applicable for only eight cases (being the "successful" cases for this test in iteration 3) whereas all the other tests (N2, N14 and N15) were applied to a significantly large number of cases (45, 75 and 87, respectively) and showed very high $(REC)_4$ values (33.3%, 38.6% and 46.0%, respectively; Figure 3d). Iteration 5 (not plotted in Figure 3) also showed good applicability of both the skewness and kurtosis tests, with $(REC)_5$ values of 44.8% and 42.5%, respectively although for the former, the number of "applicable" cases was 29 (slightly less than "30 cases" set by us as the lower limit for the REC calculations). Outliers were detected as discordant by tests of category 3 for up to 6, 7 and 8 iterations for tests N2, N14 and N15, respectively.

Test category 4 (single upper- or lower-outlier tests): The results are shown in Figure 4. For this evaluation, we summed the "successful" cases in a given iteration, from the two versions (upper and lower outlier types) of a given test (N1, N4, N9 or N10). The same was done for the total number of "applicable" cases. For iteration 1 (Figure 4a), the $(REC)_1$ values for the Grubbs type tests (N1; 13.2% and N4; 13.3%) were similar to the Dixon type tests (N9; 13.0% and N10; 13.6%). However, for iterations 2 and 3 (Figure 4b,c), the Grubbs type tests (N1 and N4) showed much greater $(REC)_2$ and $(REC)_3$ values (13.2-19.7%) than the Dixon type tests (6.3-9.5%). For iteration 4 (Figure 4d),

only the Grubbs type tests (N1 and N4) were applicable and showed $(REC)_4$ values of about 32.0%.

Test category 5 (all single-outlier tests): The results are shown in Figure 5 as a bar-plot. In order to carry out a "fair" comparison of all tests in this final test category (5), we counted all discordant outliers detected in all iterations by a given test (both upper and lower outlier versions wherever applicable), normalised this number to the total number of "initial" cases applicable in our database for a given test, and expressed the result in % for the overall $(REC)_5$ value for that particular test (see equation 4 above). The $(REC)_5$ values showed the following test sequence (from the greatest to the smallest values): the kurtosis two-sided test (N15; 35.9%); the Grubbs type one-sided test (N4; 31.6%); the Grubbs type one-sided test (N1; 31.3%); the skewness two-sided test (N14; 29.9%); the Dixon type one-sided test (N10; 29.4%); the Dixon type one-sided test (N9; 28.0%); the Grubbs type two-sided test (N2; 28.0%); the Dixon type two-sided test (N8; 22.2%); and the Dixon type one-sided test (N7; 17.5%). In other words, based on the empirical $(REC)_5$ values obtained by processing the extensive geochemical database for thirty-five RMs, we suggest the following inequality for the relative efficiency of these single-outlier tests: $N15 > N4 = N1 > N14 > N10 > N9 = N2 > N8 > N7$.

REC as a function of sample sizes: We now briefly analyse whether the REC parameter shows any systematic variation with the sample size (n). For this purpose, we identified the "applicable" cases separately for each sample-size, applied the discordancy tests (iteration 1), and calculated the corresponding $(REC)_1$ values separately for all sample sizes ($n \geq 3$), with the condition that the "applicable" cases ≥ 30 . The results are shown in Figure 6. The $(REC)_1$ values generally increased with n for all discordancy tests (Figure 6a-d).

A somewhat different analysis is presented in Figure 7. Here, we first counted the number of "applicable" cases for different sample-sizes (Figure 7a). It is clear that the cases for $n = 3$ or $n = 4$ represent a relatively small proportion of all "applicable" cases (compare the size of the bars for (i) and (ii) with those for (iv) to (vii); Figure 7a). Therefore, it is of little concern if the discordancy tests are applied for all "applicable" sample-sizes (Table 1), including such small sizes as $n = 3$. The other reason for the justification of the application of the discordancy tests to the statistical samples of small sizes, such as $n = 3$, is that the $(REC)_1$ values are relatively small for them (Figure 6a-d). Furthermore,

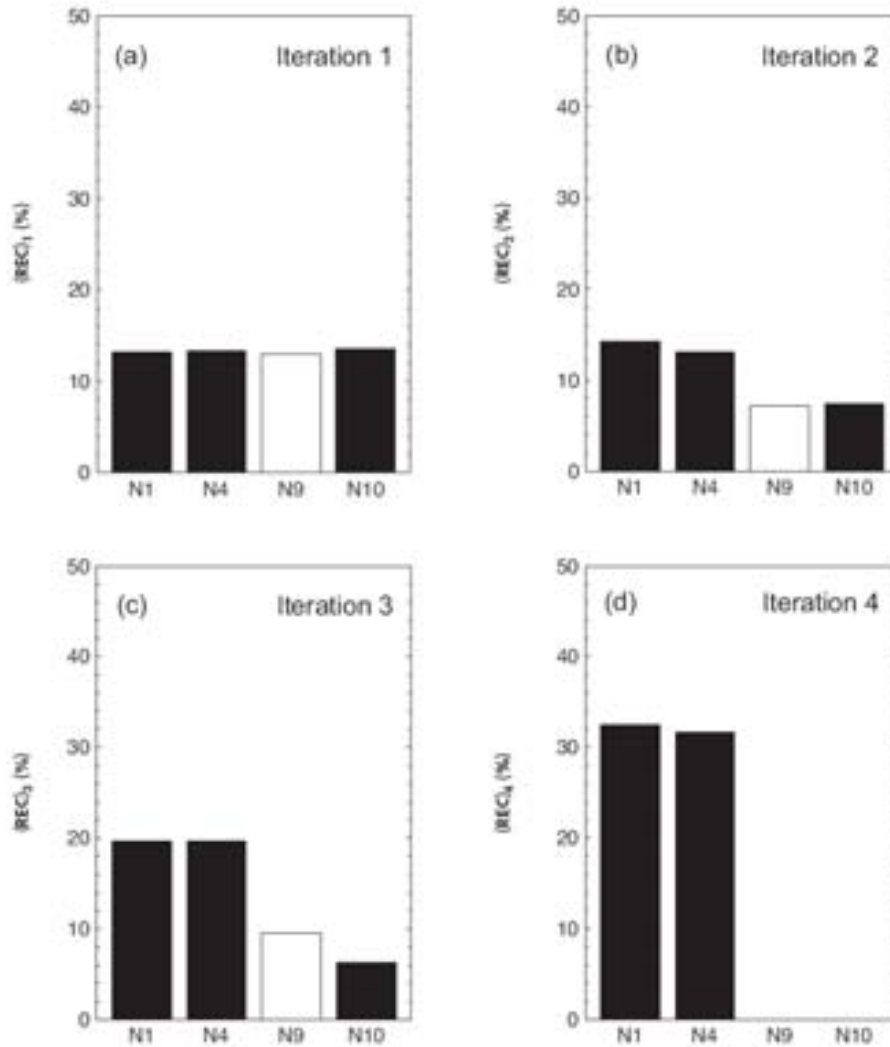


Figure 4. The relative efficiency criterion (REC) of the discordancy tests in Test category 4 (see Table 2 for categorisation). The four tests compared in Test category 4 are: N1 (both upper and lower versions), N2, N9 (both upper and lower versions), and N10 (both upper and lower versions). The number of cases to which a given test was applied is also included. (a) Iteration 1 (10572 cases for N1 and N4, 8980 for N9, and 7778 for N10); (b) Iteration 2 (1412 cases for N1, 1825 for N4, 1167 for N9, and 1058 for N10); (c) Iteration 3 (203 cases for N1, 208 for N4, 84 for N9, and 79 for N10); and (d) Iteration 4 (40 cases for N1, 41 for N4, 0 for N9, and 0 for N10).

for samples of size $n = 3$, $(REC)_2$ does not exist because if for a case an outlier were detected as discordant, the final sample size of the remaining statistical sample after Iteration 1 will be $n = 2$ and no discordancy test would then be applicable (Table 1). Finally, because this discordancy test methodology is applied

to individual method-groups, the final results of the RM concentration mean and standard deviation values for all analytical data by all method-groups are not likely to significantly change if the discordancy tests were applied to a minimum of $n = 6$ instead of $n = 3$ although this point must be confirmed in future.

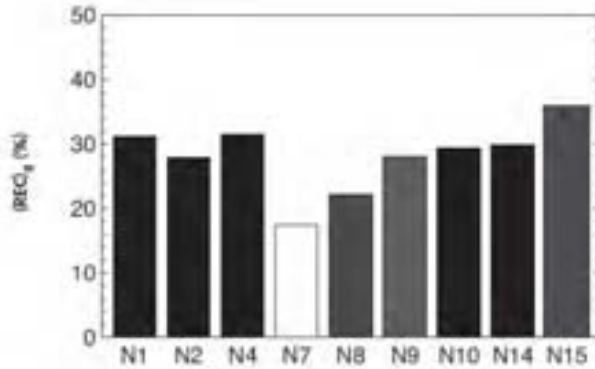


Figure 5. The relative efficiency criterion (REC) of the discordancy tests in the final Test category 5 (see Table 2 for categorisation). All nine tests with thirteen variants compared in Test category 5 are: N1 (both upper and lower versions), N2, N4 (both upper and lower versions), N7, N8, N9 (both upper and lower versions), N10 (both upper and lower versions), N14 and N15. For REC calculations all outliers detected as discordant in all iterations were combined and then referred to the "initial" applicable cases (see text for more details). The database included the following cases for the application of single-outlier tests: 5336 cases for N1, N2, N4 and N7; 4490 cases for N8 and N9; and 3889 cases for N10, N14 and N15.

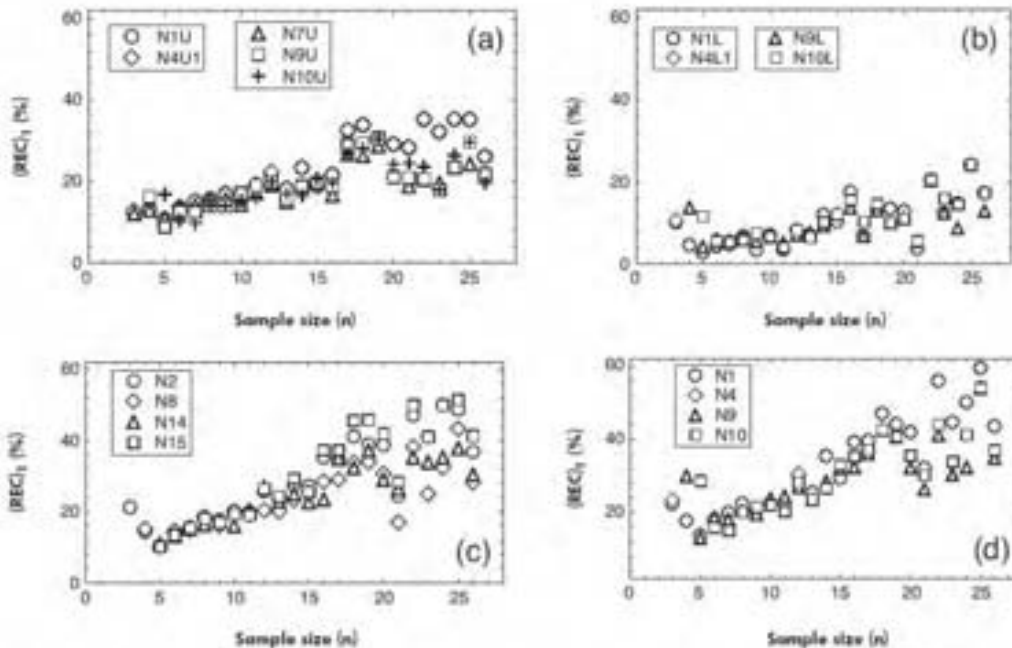


Figure 6. The relative efficiency criterion (REC) of discordancy tests for Iteration 1 (REC₁) as a function of the sample size (n). Note that in order to facilitate a better visual comparison of the (REC₁) values for different test categories, the same y-scale (0-60%) is maintained throughout this Figure. (a) Test Category 1 (one-sided upper outlier tests N1U, N4U1, N7U, N9U and N10U); (b) Test Category 2 (one-sided lower outlier tests N1L, N4L1, N9L and N10L); (c) Test Category 3 (two-sided, extreme outlier tests N2, N8, N14 and N15); and (d) Test Category 4 (combined upper and lower outlier versions of tests N1, N4, N9 and N10).

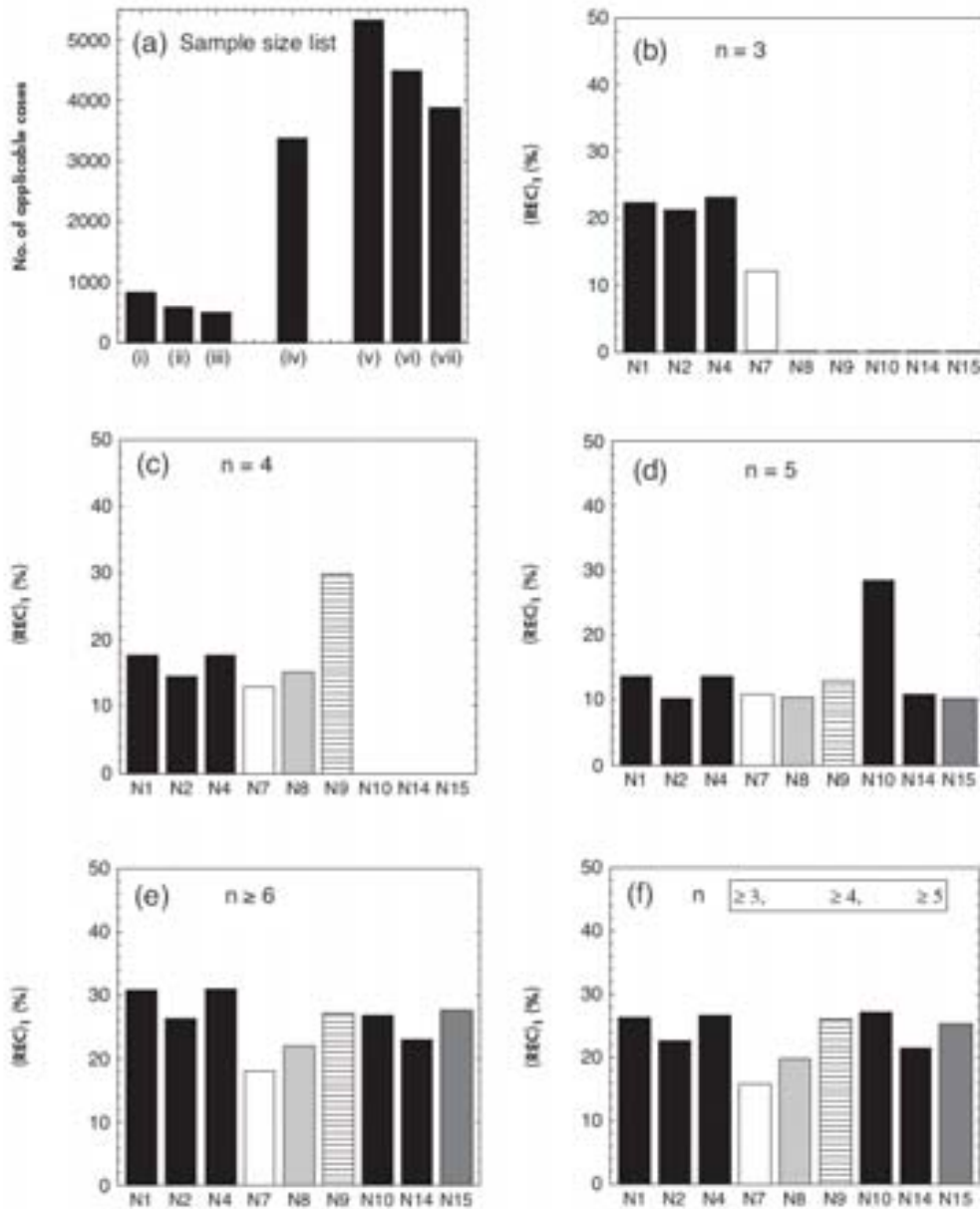


Figure 7. "Applicable" cases and $(REC)_1$ values for all single-outlier tests and selected sample sizes (n). (a) Number of "applicable" cases in our RM database for (i) $n=3$; (ii) $n=4$; (iii) $n=5$; (iv) $n \geq 6$; (v) $n \geq 3$; (vi) $n \geq 4$ and (vii) $n \geq 5$; (b) $(REC)_1$ for tests N1, N2, N4 and N7 and for $n=3$ (note the other single-outlier tests are not applicable for this sample size - Table 1); (c) $(REC)_1$ for tests N1, N2, N4, N7, N8 and N9, and for $n=4$ (note the other single-outlier tests are not applicable for this sample size - Table 1); (d) $(REC)_1$ for all single-outlier tests and for $n=5$ (note all tests are applicable for this sample size - Table 1); (e) $(REC)_1$ for all single-outlier tests and for $n \geq 6$; and (f) $(REC)_1$ for tests N1, N2, N4 and N7 - for $n \geq 3$, tests N8 and N9 - for $n \geq 4$, and tests N10, N14 and N15 - for $n \geq 5$.

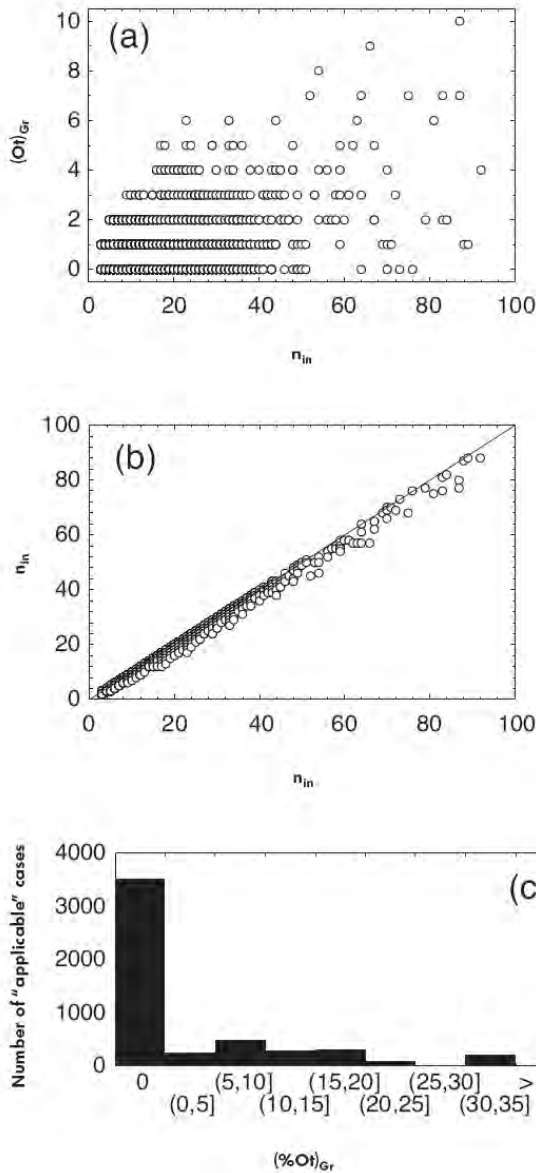


Figure 8. Discordant outliers in the RM database. (a) Number of discordant outliers in a given "applicable" case $(Ot)_{Gr}$, as a function of the initial number of observations n_{in} ; (b) Initial number of observations n_{in} versus final number of observations n_f , note the cases plotting right at the diagonal dashed line showed no discordant outliers; (c) Histogram plot of the %Discordant Outliers $(\%Ot)_{Gr}$ parameter, the width of the histogram window is 5%, except for the first large bar, which is for $(\%Ot)_{Gr} = 0\%$ (i.e., no discordant outliers detected for these numerous cases), the successive bars are, therefore, for : 0 to 5%, > 5% to 10%, and so on.

The $(REC)_1$ values for the discordancy test categories 1 to 5 are presented in Figure 7b-f, respectively. For small sample sizes such as $n = 4$ and $n = 5$ in our RM database, the Dixon-type tests (N9 and N10, respectively), presented very high $(REC)_1$ values (Figures 7c, d). For larger samples of size $n \geq 6$, the Grubbs type tests (N1 and N4) showed somewhat greater $(REC)_1$ values than most other tests (Figure 7e). Finally,

when the tests were applied to all "applicable" cases, i.e., N1 to N7 for $n \geq 3$, N8 and N9 for $n \geq 4$, and N10 to N15 for $n \geq 5$ (Table 1), the Grubbs type tests (N1 and N4), Dixon type tests (N9 and N10) and kurtosis test (N15) showed somewhat greater $(REC)_1$ than the remaining tests (Figure 7f). However, note that this is only a "partial" result - $(REC)_1$ parameter - of the application of discordancy tests; for the "complete"

result, the overall REC value - $(REC)_g$ parameter - should instead be considered (Figure 5).

Distribution of discordant outliers

We now briefly present in Figure 8 the distribution of discordant outliers in the "applicable" cases in our RM database. First, it is important to note that a large number of "applicable" cases showed no discordant outliers. This is true because a large number of cases plot close to the 0 value of the y-axis in Figure 8a, or close to the diagonal "zero discordant outlier" dashed line in Figure 8b, and constitute the largest bar in the histogram plot in Figure 8c. Second, the number of discordant outliers was limited to about ten observations for a maximum sample size of 100 (Figure 8a). This is also shown in Figure 8b, in which most cases plot close to the diagonal dashed line. Finally, the frequency of the occurrence of $(\%O)_{CO}$ values of 5-35% is significantly less than the frequency of $(\%O)_{CO} = 0\%$ (compare the size of the bars with that of the "zero discordant outlier" large bar; Figure 8c); furthermore, in our RM database the $(\%O)_{CO}$ values are limited to a maximum value of about 35%, with only a few exceptions of greater $(\%O)_{CO}$.

Calibration of X-ray fluorescence spectrometry

The present methodology involving the application of the single-outlier tests to our RM database (compiled from Steele *et al.* (1972, 1978), Abbey (1979),

Flanagan (1986), Gladney (1988), Gladney and Roelandts (1988, 1990), Gladney *et al.* (1991, 1992), and the internet address <http://niodb02.lbase.oist.go.jp/earthsci/welcome.html> accessed for the last time in June 2007) is capable of providing reliable estimates of both central tendency (mean) and dispersion (standard deviation) parameters. The actual values of these parameters for the RMs compiled in this work will be estimated and reported in future, after the application of the complete multiple-test method of Verma (1997), i.e., involving both single-outlier as well as multiple-outlier tests, after the RM database is adequately updated by incorporating more recent literature data (i.e., those reported after 1978 for the RMs from South Africa and after 1988-1992 for the RMs from Canada and USA); this work is currently in progress.

The availability of "statistically correct" estimates of both central tendency and dispersion parameters can enable us to achieve instrumental calibrations using a WLR model instead of the conventional OLR model generally applied for this purpose in most laboratories. We present the improvement achieved by a WLR model for the calibration of the major elements by X-ray fluorescence spectrometry (XRF) as compared to the OLR model. In addition to the RMs, the blank and monitor discs (or beads) were also routinely analysed by XRF. All intensities were corrected for matrix effects and instrumental drift; more details on the XRF procedure will be presented elsewhere (M. Guevara and S.P. Verma, manuscript in preparation). The final calibration results were documented as the intercept and slope

Table 5.
Comparison of the ordinary least-squares linear regression (OLR) model with weighted least-squares linear regression (WLR) model for the calibration of the major elements by X-ray fluorescence spectrometry (XRF)

Element or oxide	OLR model			WLR model		
	r (n _i)	Intercept (counts)	Slope (counts/%m/m)	r (n _i)	Intercept (counts)	Slope (counts/%m/m)
SiO ₂	0.9969 (21)	-144,000 ± 103,000	103,700 ± 1,900	0.9998 (21)	6,000 ± 5,800	101,590 ± 460
TiO ₂	0.9999 (22)	-12,400 ± 5,100	1,463,200 ± 5,400	0.9996 (22)	1,170 ± 290	1,448,000 ± 8,900
Al ₂ O ₃	0.9989 (21)	-16,600 ± 20,400	141,730 ± 1,550	0.9998 (21)	1,630 ± 570	140,800 ± 600
Fe ₂ O ₃	0.9982 (23)	30,000 ± 63,000	487,800 ± 6,700	0.9993 (23)	25,300 ± 12,900	482,200 ± 4,100
MnO	0.9995 (22)	8,700 ± 10,400	6,198,000 ± 46,000	0.9998 (22)	6,800 ± 1,100	6,195,700 ± 30,900
MgO	0.9998 (22)	-38,300 ± 8,100	338,000 ± 1,500	0.9988 (22)	520 ± 400	332,400 ± 3,900
CuO	0.9999 (23)	-72,800 ± 24,500	862,200 ± 3,200	0.9998 (23)	4,300 ± 1,700	851,000 ± 3,700
H ₂ O	0.9969 (23)	-79,100 ± 7,900	137,800 ± 2,400	0.9977 (23)	6,000 ± 3,600	129,700 ± 2,000
K ₂ O	0.9987 (22)	59,600 ± 33,300	713,400 ± 8,000	0.9993 (22)	4,000 ± 1,500	766,200 ± 6,300
P ₂ O ₅	0.9979 (22)	-14,900 ± 3,400	1,584,600 ± 12,000	0.9973 (22)	1,780 ± 520	1,521,100 ± 18,000

The parameter r is the linear correlation coefficient, n_i is the number of data points in a given calibration. The intercept and slope refer to a linear regression model (OLR or WLR). Note rounded values are reported according to the criteria put forth by Berrington and Robinson (2003) and Verma (2005).

values, including their uncertainty estimates (95% confidence limits as suggested by ISO 1989). These statistical parameters thus define the regression equations for all major elements or oxides (Table 5).

Both OLR and WLR models showed relatively high, statistically significant (Bevington and Robinson 2003, Verma 2005) linear correlation coefficient (r) values ($r = 0.9969$ for the number of calibration points = 21-23; Table 3). The WLR model always provided a smaller intercept (closer to zero) than the OLR model. The respective uncertainties were also smaller for the WLR model. Although the slopes of the regression lines for the two models were generally similar, the WLR model generally showed smaller uncertainty values for the slopes. Thus, the regression equations of the WLR model were of "better" quality (intercepts close to zero and smaller uncertainties on both intercept and slope coefficients) than the respective equations of the OLR model. The WLR should, therefore, provide more accurate estimates of the major element data in unknown samples as compared to the frequently used OLR model.

Final remarks

These REC results are generally consistent with the earlier empirical evaluations of Velasco and Verma (1998) and Velasco *et al.* (2000), in which the importance of skewness and kurtosis tests was well established for efficiently detecting discordant outliers. The superiority of the kurtosis test (N15) over all other tests is documented for the first time. Nevertheless, because a very extensive database of thirty-five RMAs, involving all major and trace elements, was used in the present work and all "applicable" cases could be processed using new, precise and accurate critical values (Verma and Quiroz-Ruiz 2006a, b, 2008, Verma *et al.* 2008), the statistical inferences drawn here are more firmly valid than in these earlier studies. The kurtosis test (N15), Grubbs type tests (N1 and N4) and skewness test (N14) are to be preferred in comparison with the Dixon type tests (N7, N8, N9 and N10) as well as the Grubbs type test (N2).

We should also point out that our outlier-based procedure should be especially useful for objectively identifying discordant outliers in many fields of science and engineering as has been discussed in detail by Verma and Quiroz-Ruiz (2006a, b). The fields, for which example cases and literature references were provided by these authors, are: agriculture, astronomy, biology, biomedicine, biotechnology, chemistry, electronics,

environmental and pollution research, food science and technology, geochemistry, geochronology, isotope geology, meteorology, nuclear science, palaeontology, petroleum research, quality assurance and assessment programmes, soil science, structural geology, water research and zoology. After the proper identification of discordant outliers, these can be interpreted accordingly and appropriately depending on the expertise of the scientific or engineering field.

Conclusions

An extensive geochemical database of thirty-five reference materials from Canada, Japan, South Africa and the USA was used to evaluate nine single-outlier tests with thirteen variants. The statistical evaluation of these single-outlier tests showed the relative efficiencies of these tests (represented by the $(REC)_0$ parameter) in the following sequence: N15 > N4 > N1 > N14 > N10 > N9 > N2 > N8 > N7. The number of discordant outliers in the database identified by the present single-outlier method was relatively small (from 0 to 10) for sample sizes up to 100. The $(\%O)_0$ parameter most frequently was zero and for a significantly lesser number of cases its value was limited to a maximum of about 35%. The WLR model was shown to provide "better quality" calibration equations with the intercepts closer to zero and the uncertainties on both the intercept and slope generally smaller than the OLR model.

Acknowledgements

We thank the editor Mirella Polve for extending an informal invitation to contribute to Geostandards and Geoanalytical Research and for an efficient and unbiased editorial handling of our manuscript. This work was motivated by an invitation from Phil Potts to the first author (SPV) to review the statistical methods for processing geochemical databases. The present contribution is the first step towards this goal; further work is currently in progress. The second and third authors (LDG and RGR) are grateful to Conacyt for a scholarship to carry out doctoral studies at CIE-UNAM. Alfredo Quiroz-Ruiz is thanked for efficiently maintaining our personal computers and guiding us at the data compilation stage of our work. Our presentation was considerably improved from the critical comments and suggestions from three anonymous reviewers. We are also grateful to Mima Guevara for allowing us to use the unpublished XRF laboratory data, which enabled us to demonstrate the importance of our work for instrumental calibrations.

References

- Abbey S. (1979)**
Reference materials - rock samples SF-2, SF-3, MRG-1.
Energy, Mines and Resources (Canada), Report 79-35,
66pp.
- Azevedo A.G. and González G. (2007)**
Fitting straight lines with replicated observations by linear
regression II: Weighting data. *Critical Reviews in
Analytical Chemistry*, 37, 143-172.
- Barnett V. and Lewis T. (1994)**
Outliers in statistical data. John Wiley and Sons
(Chichester), 584pp.
- Bourman K. (1997)**
Regression and calibration for analytical separation
techniques. Part II: Validation, weighted and robust
regression. *Process Control and Quality*, 10, 75-112.
- Beverington P.R. and Robinson D.K. (2003)**
Data reduction and error analysis for the physical
sciences. McGraw-Hill (New York), 320pp.
- Dixon W.J. (1951)**
Ratios involving extreme values. *Annals of Mathematical
Statistics*, 22, 68-78.
- Dybczynski R., Tuganov A. and Suschey O. (1979)**
Sol-5, a new IAEA certified reference material for trace
element determinations. *Geostandards Newsletter*, 3,
61-67.
- Dybczynski R., Polkownik-Motienko H.,
Samczynski Z. and Szopa Z. (1998)**
Virginia tobacco leaves (CIAVTS-2) - new Polish CRM for
inorganic trace analysis including microanalysis.
Fresenius' Journal of Analytical Chemistry, 360,
384-387.
- Ferre M., Martinez E., Hernandez M.D.,
Fernandez-Alba A., Fritz J., Uenoh E., Mihail O.,
Sakkas V., Morbey A., Albanis T., Brito F.,
Hansen P.D. and Baraco 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, 323-333.
- Flanagan F.J. (1986)**
Rock reference samples, San Marcos Gabbro, GSM-1
and Lakeview Mountain Tonalite, TUM-1. *Geostandards
Newsletter*, 10, 111-119.
- Gobronská D., Rysava J., Filova V., Plicka J.,
Cuhra P., Kubik M. and Baranova S. (2006)**
Gluten determination by gliadin enzyme-linked
immunosorbent assay kit. Interlaboratory study. *Journal of
AOAC International*, 89, 154-160.
- Gladney E.S. (1988)**
1987 compilation of elemental concentration data for
USGS BR-1, DNC-1 and W-2. *Geostandards Newsletter*,
12, 63-118.
- Gladney E.S. and Roelands I. (1988)**
1987 compilation of elemental concentration data for
USGS BHVO-1, MAG-1, QLO-1, RGM-1, SCO-1, SDC-1,
SCR-1 and SRM-1. *Geostandards Newsletter*, 12, 253-262.
- Gladney E.S. and Roelands I. (1990)**
1988 compilation of elemental concentration data for
CORMP reference rock samples SX-2, SX-3 and MRG-1.
Geostandards Newsletter, 14, 373-458.
- Gladney E.S., Jones E.A., Nickell E.J. and
Roelands I. (1991)**
1988 compilation of elemental concentration data for
USGS DS-1, G-1, PCC-1 and W-1. *Geostandards
Newsletter*, 15, 199-396.
- Gladney E.S., Jones E.A., Nickell E.J. and
Roelands I. (1992)**
1988 compilation of elemental concentration data for
USGS AQV-1, GSR-1 and G-2. *Geostandards
Newsletter*, 16, 111-300.
- Govindaraju K. and Roelands I. (1989)**
1988 compilation report on trace elements in six ANRT
rock reference samples: Diorite DR-N, serpentinite UB-N,
bauxite BX-N, diathene DR-N, granite GS-N and potash
feldspar FK-N. *Geostandards Newsletter*, 13, 5-67.
- Govindaraju K., Potts P.J., Webb P.C. and
Watson J.S. (1994)**
1994 Report on Whin Sill dolomite WS-E from England
and Pitscutie microgabbro PM-S from Scotland.
Assessment by one hundred and four international
laboratories. *Geostandards Newsletter*, 18, 211-300.
- Grubbs F.E. (1950)**
Sample criteria for testing outlying observations. *Annals
of Mathematical Statistics*, 21, 27-58.
- Grubbs F.E. (1969)**
Procedures for detecting outlying observations in samples.
Technometrics, 11, 1-21.
- Guevara M., Yermo S.P. and
Velasco-Tapia E. (2001)**
Evaluation of GSI intrusive rocks JG1, JG2, JG3, JG1a
and JG61 by an objective outlier rejection statistical
procedure. *Revista Mexicana de Ciencias Geológicas*,
18, 74-88.
- Guevara M., Yermo S.P., Velasco-Tapia E.,
Lozano-Santa Cruz R. and Girón F. (2005)**
Comparison of linear regression models for quantitative
geochemical analysis. An example using X-ray
fluorescence spectrometry. *Geostandards and
Geoanalytical Research*, 29, 271-284.
- Hayes K., Kinsella A. and Coffey N. (2007)**
A note on the use of outlier criteria in Ontario laboratory
quality control schemes. *Clinical Biochemistry*, 40,
147-152.
- Isoi H., Terashima S., Itoh S. and Ando A. (1995)**
1994 compilation values for GSI reference samples,
"igneous rock series". *Geochemical Journal*, 29, 91-95.
- ISO (1989)**
ISO Guide 35: Certification of reference materials -
general and statistical principles. International
Organization for Standardization (Geneva), 32pp.



references

- Li X.J., Zhang H., Ranish J.A. and Aebersold R. (2003)
Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry *Analytical Chemistry*, 75, 6648-6657.
- Mahwar R.S., Verma N.K., Chakrabarti S.P. and Biswas D.K. (1998)
Development and use of reference materials in India - status and plans. *Fresenius' Journal of Analytical Chemistry*, 360, 291-295.
- Namiesnik J. and Zygmont B. (1999)
Role of reference materials in analysis of environmental pollutants. *The Science of the Total Environment*, 218, 243-257.
- Potts P.J. and Kane J.S. (1992)
Terminology for geological reference material values: A proposal to the International Organization for Standardization (ISO), producers and users. *Geostandards Newsletter*, 16, 333-341.
- Querauville P., Maier E.A., Giepinik B., Fortmann U., Vercootere K. and Muntau H. (1996)
Certified reference materials of soils and sewage sludges for the quality control of trace element environmental control. *Trends in Analytical Chemistry*, 15, 504-513.
- Song H.Q., Wang F., He H.Y., Wang Y.L., Yang L.K. and Zhu R.X. (2006)
Inter-calibration of ZBH-25 biotope reference material utilized for K-Ar and Ar-40/Ar-39 age determination. *Acta Petrologica Sinica*, 22, 3059-3078.
- Sorlayo E. and Verma S.P. (2003)
Determination of lanthanides in synthetic standards by reversed-phase high performance liquid chromatography with the aid of a weighted least-squares regression model. Estimation of method sensitivities and detection limits. *Journal of Chromatography A*, 997, 171-182.
- Sorlayo E., Guevara M. and 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, 73-81.
- Soyogo A., Boccia M. and Asuero A.G. (2004)
Fitting straight lines with replicated observations by linear regression: The least squares postulates. *Critical Reviews in Analytical Chemistry*, 34, 39-50.
- Serbst J.R., Burgess R.M., Kuhn A., Edwards P.A., Cartwell M.G., Pelletier M.C. and Berry W.J. (2003)
Precision of dialysis (peeped sampling) of cadmium in marine sediment interstitial water. *Archives of Environmental Contamination and Toxicology*, 45, 297-305.
- Steele T.W., Russell B.G., Goudris R.G., Daniel G. and Levin J. (1972)
Preliminary report on the analysis of the six NIMROC geochemical standard samples. National Institute for Metallurgy (Randburg, South Africa), Report 1351, 74pp.
- Steele T.W., Wilson A., Goudris R., Ellis P.J. and Radford A.J. (1978)
Analysis of the NIMROC reference samples for minor and trace elements. National Institute for Metallurgy (Randburg, South Africa), Report 1945, 216pp.
- Stelopoulos F. and Stichel E. (2007)
Estimation of performance characteristics of a confirmation method for thyrostatin in plasma by means of a weighted least-squares approach. *Analytica Chimica Acta*, 592, 181-186.
- Tellinghuisen J. (2007)
Weighted least-squares in calibration: What difference does it make? *The Analyst*, 132, 536-543.
- Vendemiato M.A. and Enzweiler J. (2001)
Routine control of accuracy in silicate rock analysis by X-ray fluorescence spectrometry. *Geostandards Newsletter: The Journal of Geostandards and Geoanalysis*, 25, 283-291.
- Velasco F. and 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, 109-128.
- Velasco F., Verma S.P. and Guevara M. (2000)
Comparison of the performance of fourteen statistical tests for detection of outlying values in geochemical reference material databases. *Mathematical Geology*, 32, 439-464.
- Velasco-Tapia F., Guevara M. and Verma S.P. (2001)
Evaluation of concentration data in geochemical reference materials. *Chemie der Erde*, 61, 69-71.
- Verma S.P. (2004)
A revised analytical method for HCO₃⁻ and CO₃²⁻ determinations in geothermal waters. An assessment of WQC and IAEA interlaboratory comparisons. *Geostandards and Geoanalytical Research*, 28, 391-409.
- Verma S.P. (1997)
Sixteen statistical tests for outlier detection and rejection in evaluation of international geochemical reference materials: Example of microgabbaro PM-5. *Geostandards Newsletter: The Journal of Geostandards and Geoanalysis*, 21, 59-75.
- Verma S.P. (1998)
Improved concentration data in two international geochemical reference materials (USGS basalt BH-1 and GSI peridotite JP-1) by outlier rejection. *Geobios Internacional*, 37, 215-250.
- Verma S.P. (2005)
Estadística básica para el manejo de datos experimentales. Aplicación en la geoquímica (geoquimiometría). UNAM (Mexico), D.F., 186pp.
- Verma S.P. and Quiroz-Ruiz A. (2006a)
Critical values for six Dixon tests for outliers in normal samples up to size 100, and applications in science and engineering. *Revista Mexicana de Ciencias Geológicas*, 23, 133-161.

references

- Werna S.P. and 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, 302-319.
- Werna S.P. and Quiroz-Ruiz A. (2008)**
Critical values for 33 discordancy test variants for outliers in normal samples for very large sizes of 1,000 to 30,000. *Revista Mexicana de Ciencias Geológicas*, 25, 369-381.
- Werna S.P., Ordóñez-Galván L.J. and Guevara M. (1998)**
SIPACE: A new computer programme with seventeen statistical tests for outlier detection in evaluation of international geochemical reference materials and its application to Whin Sill diorite WS-E from England and Sol-5 from Peru. *Geostandards Newsletter: The Journal of Geostandards and Geoanalysis*, 22, 209-234.
- Werna S.P., Quiroz-Ruiz A. and Diaz-González L. (2008)**
Critical values for 33 discordancy test variants for outliers in normal samples up to sizes 1000, and applications in quality control in Earth Sciences. *Revista Mexicana de Ciencias Geológicas*, 25, 82-96.
- Vileneuve J.-P., de Mora S. and Cattini C. (2004)**
Determination of organochlorinated compounds and petroleum in fish-homogenate sample WEA-406: Results from a worldwide interlaboratory study. *Trends in Analytical Chemistry*, 23, 501-510.

APÉNDICE 2

González-Ramírez R., Díaz-González, L., Verma S.P. (2009)

Eficiencia Relativa de 15 pruebas de discordancia con 33 variantes aplicadas al procesamiento de datos geoquímicos. Revista Mexicana de Ciencias Geológicas, 26 (2), 501-515

Eficiencia relativa de 15 pruebas de discordancia con 33 variantes aplicadas al procesamiento de datos geoquímicos

Rosalinda González-Ramírez¹, Lorena Díaz-González¹ y Surendra P. Verma^{2*}

¹ Posgrado en Ingeniería (Energía), Universidad Nacional Autónoma de México, Privada Xochicalco s/n, Centro, 62580 Temixco, Morelos, México.

² Centro de Investigación en Energía, Universidad Nacional Autónoma de México, Privada Xochicalco s/n, Centro, 62580 Temixco, Morelos, México.

* spv@cie.unam.mx

RESUMEN

Las pruebas de discordancia son una herramienta estadística útil en los diferentes campos de las ciencias e ingenierías, incluyendo Ciencias de la Tierra. El procedimiento consiste en una metodología rigurosa para la detección de datos desviados en una muestra estadística "contaminada" y posteriormente su depuración, logrando que los datos restantes tengan una distribución normal sin contaminación estadística, con los cuales pueden ser determinadas correctamente las medidas de tendencia central (media) y de dispersión (desviación estándar). En la evaluación empírica de las 15 pruebas de discordancia con 33 variantes, se utilizó una base de datos geoquímicos grande con información sobre 35 materiales de referencia geoquímica (MRG) procedentes de cuatro países (Canadá, E.U.A., Japón y Sudáfrica) que representa 2220 casos con 41,821 datos individuales geoquímicos. Fueron evaluadas nueve pruebas sencillas con 15 variantes y siete pruebas múltiples con 20 variantes (la prueba N4 pertenece a ambas tipos) utilizando valores críticos nuevos de gran precisión y exactitud en la obtención de los resultados. Para la eficiencia de las pruebas de discordancia se emplearon dos términos estadísticos: (1) Criterio de eficiencia relativa ("relative efficiency criterion", REC) previamente conocido; y (2) criterio de valores desviados relativo ("relative outlier criterion" ROC) propuesto en el presente trabajo. Adicionalmente, se utilizó una metodología combinada de regresión lineal y pruebas de significancia de F de Fisher y t de Student. En pruebas de discordancia sencillas, la eficiencia mayor fue para el coeficiente de exceso o curtosis (N15) seguida por las pruebas tipo Grubbs (N1 y N4) y de coeficiente de asimetría (N14), mientras que en pruebas de discordancia múltiples, la prueba N4 en sus tres variantes se caracterizó por eficiencias mayores. Las pruebas tipo Dixon, mucho más populares que las de Grubbs, por lo general presentaron valores menores de la eficiencia. Una implicación importante de estos resultados sería otorgar preferencias a las pruebas N15, N1, N4 y N14 para la aplicación de la metodología de valores desviados en el manejo de datos geoquímicos. Las interpretaciones cuantitativas de regresiones lineales combinadas con pruebas de significancia confirman los resultados de los parámetros REC y ROC. Finalmente, se afirma que independientemente del método analítico usado para determinar la composición geoquímica de materiales de referencia, los valores desviados altos son mucho más comunes que los bajos y las muestras con contaminación estadística simétrica, a ambos lados de la muestra, son relativamente escasas. Los parámetros robustos, como la mediana o la media de Gastwirth, serán muy probablemente seguras para este tipo de datos geoquímicos. Así mismo, la aplicación rigurosa de las pruebas de discordancia antes de estimar los valores de la media y desviación estándar parece ser un requerimiento básico.

Palabras clave: materiales de referencia geoquímica, métodos de valores desviados, pruebas de discordancia, pruebas de Dixon, pruebas de Grubbs, sesgo, coeficiente de asimetría, curtosis, valores críticos, pruebas de significancia.

ABSTRACT

Discordancy tests provide us with a statistical tool that is useful in different areas of science and engineering, including Earth Sciences. Their application represents a rigorous methodology for the detection and elimination of discordant outliers in statistically contaminated normal samples and provides us remaining data without any statistical contamination, which can then be used to estimate the central tendency (mean) and dispersion (standard deviation) parameters. For the empirical evaluation of 15 discordancy tests with 33 variants, an extensive database of 35 reference materials (RM) from four countries (Canada, U.S.A., Japan, and South Africa) having 2220 applicable cases with 41,821 individual geochemical data, was established. Nine single-outlier tests with 13 variants and seven multiple-outlier tests with 20 variants (test N4 belongs to both types) along with the new, most precise and accurate critical values, were employed for this evaluation. Two statistical parameters quantified the efficiency of discordancy tests: (1) Relative efficiency criterion (REC) known from previous work; and (2) relative outlier criterion (ROC) proposed in this work. Additionally, a methodology was used that combines linear regression analysis with Fisher F and Student t significance tests. Among the single-outlier discordancy tests, the greatest efficiency was shown by kurtosis test (N15), followed by Grubbs type tests (N1 and N4) and skewness test (N14), whereas, among multiple-outlier tests, the Grubbs test N4 in its three variants seemed to be characterized by the greatest efficiency values. The Dixon tests, being much more popular than the Grubbs tests, in general presented the smallest efficiencies. One important implication of these results would be to prefer N15, N1, N4, and N14 tests for the application of this outlier-based methodology for geochemical data handling. The quantitative interpretation using the combined methodology of linear regressions and significance tests confirms the results of REC and ROC parameters. Finally, it is inferred that independently of the analytical methods used for the determination of geochemical composition of reference materials, upper discordant outliers are much more common than the lower ones, and samples with a symmetrical statistical contamination on both sides of the sample are relatively scarce. Robust estimates, such as the median or Gastwith mean, are likely to be biased for such geochemical data. The application of discordancy tests before estimating the mean and standard deviation values is a basic requirement.

Key words: reference material, outlier-based methods, discordancy tests, Dixon tests, Grubbs test, skewness, kurtosis, critical values, significance tests.

INTRODUCCIÓN

Existe una amplia literatura relacionada con los métodos aplicados a la estimación de los parámetros de la tendencia central y la dispersión de datos univariados (Barnett y Lewis, 1994). Estos se distribuyen en dos tipos de metodologías generales (Verma, 2005): (a) métodos robustos, que utilizan criterios de estimación que se ven poco afectados por la presencia de valores desviados; y (b) métodos de valores desviados, que consisten en la aplicación de pruebas de discordancia para la detección y eliminación de valores desviados en muestras estadísticas previo a los cálculos de parámetros de la tendencia central y la dispersión.

Los métodos robustos no serán abordados en el presente trabajo, ni los métodos de valores desviados basados en la desviación estándar (σ) de las poblaciones, siendo éstos de tipo "múltiples de σ " como 2σ o 3σ . En la práctica, para la aplicación de estos métodos a datos experimentales no se conoce σ y se le reemplaza, en forma aproximada, por la desviación estándar (s) de las muestras estadísticas (Verma, 2005). Una discusión sobre las limitaciones estadísticas de los métodos robustos y aquéllos de valores desviados basados en el criterio de dos desviaciones estándar ($\pm 2s$) puede ser consultada en Verma (1997, 1998, 2005), Verma et al. (1998) y Verma y Quiroz-Ruiz (2008).

Las pruebas de discordancia, aplicadas en la detección de valores desviados, se distribuyen en cuatro tipos básicos (Barnett y Lewis, 1994; Verma, 2005): (a) estadísticos de desviación o dispersión (tipo Grubbs N1-N3; Grubbs, 1950); (b) estadísticos de suma de cuadrados (tipo Grubbs N4 y N5); (c) estadísticos de intervalo total o de dispersión (N6 y tipo Dixon N7-N13; Dixon, 1951); y (d) estadísticos de momento de alto orden ("skewness", sesgo o coeficiente de asimetría N14 y curtosis N15).

Estas pruebas de discordancia para la detección de valores desviados o discordantes también pueden ser clasificadas en dos tipos o clases: (a) pruebas sencillas, las cuales evalúan un valor a la vez como discordante, tipo $k=1$; y (b) pruebas múltiples, las cuales evalúan dos o más valores a la vez como discordantes, tipo $k=2-4$. Esta distinción será usada en el presente trabajo.

Las pruebas de discordancia han sido aplicadas o apreciadas de forma extensiva en diferentes campos de las ciencias geológicas y de otras áreas del conocimiento, incluyendo el control de calidad mediante materiales de referencia. Aunque en la literatura moderna existen miles de estudios sobre el tema de aplicación de pruebas de discordancia, se pueden citar como ejemplos los siguientes trabajos: Dybczynsky (1980); Linkosalo et al. (1996); Freeman et al. (1997); Verma (1997, 1998); Zarić y Niketic (1997); Velasco y Verma (1998); Verma et al. (1998, 2008);

Taylor (2000); Guevara *et al.* (2001); Velasco-Tapia *et al.* (2001); Schaber y Badeck (2002); Li *et al.* (2003); Serbest *et al.* (2003); Graybeal *et al.* (2004); Farre *et al.* (2006); Gabrovská *et al.* (2006); Sang *et al.* (2006); Colombo *et al.* (2007); Gutiérrez-Ruiz *et al.* (2007); Hayes *et al.* (2007); Kasper-Zubillaga y Zolezzi-Ruiz (2007); Méndez-Ortiz *et al.* (2007); Nagarajan *et al.* (2007, 2008); Ram *et al.* (2007); Salleh *et al.* (2007); Shekhawat *et al.* (2007); Castrellon-Urbe *et al.* (2008); Díaz-González *et al.* (2008); Jafarzadeh y Hosseini-Barzi (2008); Obeidat *et al.* (2008); Palabiyik y Serpen (2008); Vargas-Rodríguez *et al.* (2008); Vattuone *et al.* (2008); Gómez-Arias *et al.* (2009); Madhavaraju y Lee (2009); Marroquín-Guerra *et al.* (2009); Pandarinath (2009); Torres-Alvarado *et al.* (en prensa); y Verma (2009a).

Dada la gran diversidad de pruebas de discordancia, es pertinente conocer sus potencias o eficiencias relativas para detectar valores discordantes. Las eficiencias dependerán seguramente del tipo de estadísticos (las fórmulas o expresiones matemáticas), en los cuales estas pruebas han sido basadas (Barnett y Lewis, 1994; Verma, 2005). Sin embargo, el funcionamiento de los estadísticos y, por consecuencia, la eficiencia de las pruebas pueden estar afectados por el así denominado efecto de enmascaramiento (*masking effect*; Barnett y Lewis, 1994), el cual se define como la inhabilidad del procedimiento para identificar un valor desviado debido a la presencia de un valor sospechoso cercano (Tietjen y Moore, 1972). Este efecto se conoce desde hace mucho tiempo (Pearson y Chandra Sekar, 1936) y constituye una posible limitación, especialmente de las pruebas de discordancia de tipo sencillo, aunque puede estar presente también en las de tipo múltiple. Un efecto contrario llamado *swamping effect* (Barnett y Lewis, 1994) puede alterar –en realidad aumentar– la potencia de pruebas de discordancia de tipo múltiple, ya que uno o más datos verdaderamente discordantes pueden llevar otro(s) dato(s) aparentemente concordante(s), resultando ambos tipos como discordantes. Este efecto no influye a las pruebas sencillas que evalúan un valor a la vez.

Es importante, por lo tanto, establecer la eficiencia (definida en la siguiente sección) de las pruebas, ya que a partir de este conocimiento es posible decidir la aplicación de un esquema basado en un número pequeño de ellas o, si es necesario, aplicar todas las disponibles como inicialmente fue propuesto por Verma (1997). Hasta el momento, no existe una evaluación completa de estas pruebas.

En este contexto, Velasco *et al.* (2000) compararon la eficiencia de 14 pruebas estadísticas, utilizando datos de composición de lantánidos (*rare-earth elements*, REE) en 26 materiales de referencia geoquímica (MRG). Estos autores reportaron: (a) que en la detección individual de valores desviados ($k=1$), las pruebas más eficientes son los estadísticos de momento de alto orden (N14 y N15); y (b) que existe una mayor eficiencia de los procedimientos en bloque ($k=2-4$), en comparación con la aplicación consecutiva de pruebas de discordancia de detección individual ($k=1$). Sin embargo, la evaluación de Velasco *et al.* (2000)

debió de haber sido limitada, ya que no fue posible hacer una comparación completa entre pruebas. Esto fue porque en ese tiempo no existían valores críticos disponibles para determinados tamaños de muestra (n), por ejemplo, para las pruebas Dixon N7-N13 para $n > 30$ (Dixon, 1951), o para la prueba de desviación N2 para $n > 20$ (Barnett y Lewis, 1994; Verma, 2005). Este problema fue solucionado por Verma y Quiroz-Ruiz (2006a, 2006b, 2008) y Verma *et al.* (2008), quienes a partir de simulaciones tipo Monte Carlo, generaron nuevos valores críticos, de mayor precisión y exactitud que los disponibles en la literatura, y con tamaño de muestra de $n=3-30,000$. Con estos nuevos trabajos se hizo evidente que los valores críticos de la literatura anteriores a los nuevos no sólo se caracterizaban por poca precisión sino también por poca exactitud, por lo cual la evaluación empírica de todas las pruebas de discordancia debe ser realizada nuevamente.

Por lo tanto, recientemente Verma *et al.* (2009) llevaron a cabo una evaluación de la eficiencia relativa de las pruebas de discordancia, específicamente las variantes que detectan un valor desviado ($k=1$) y considerando valores críticos nuevos de hasta $n=30,000$. Esta evaluación se basó en la información de composición de elementos mayores y traza en 35 MRG. Los resultados de esta evaluación indican que la prueba N15 (coeficiente de curtosis) presenta la mayor eficiencia entre los estadísticos estudiados, mientras que la prueba N7 (tipo Dixon) fue la menos eficaz en esta detección.

Con el fin de profundizar en el conocimiento sobre la capacidad de detección de valores desviados por las pruebas de discordancia, se presentan los resultados de la evaluación de 15 estadísticos de aplicación sencilla ($k=1$) y múltiple ($k=2-4$), que incluyen 33 variantes. El presente análisis también fue realizado considerando una base de datos geoquímicos (elementos mayores y traza) de los 35 MRG, empleando para ello otros indicadores de eficiencia nuevos y adicionales al usado por Velasco *et al.* (2000) y Verma *et al.* (2009). Los resultados de este análisis de eficiencia han permitido: (a) la comparación entre variantes de tipo sencillo ($k=1$) y múltiple ($k=2-4$), siendo k el número de datos a probar "a la vez"; (b) contar con principios y lineamientos estadísticos para elegir el número y tipo de estadísticos necesarios para la detección de valores desviados en muestras geoquímicas univariadas y decidir si la aplicación de un número menor de pruebas de discordancia es suficiente, o si deben ser aplicadas todas las pruebas disponibles; y (c) demostrar la potencia de las pruebas de discordancia, al evaluar la eficiencia relativa de datos desviados en forma de un parámetro estadístico existente y otro propuesto en el presente trabajo (ambos serán definidas en este trabajo) así como una combinación de regresión lineal y pruebas de significancia.

Cabe aclarar que la importancia del presente trabajo radica en apreciar que los parámetros de la media aritmética y la desviación estándar pertenecen al método estadístico de valores desviados y, por lo tanto, es necesario aplicar

las pruebas de discordancia a los datos univariados, antes de emplear estos parámetros para la estimación de la tendencia central y la dispersión, respectivamente (Verma, 2005). Además, estos parámetros son fundamentales en las calibraciones de instrumentos analíticos que nos proporcionan datos geoquímicos que sirven para inferir procesos geológicos dominantes. Por consecuencia, es pertinente contar con mayor conocimiento sobre estas pruebas de discordancia, por ejemplo, sus potencias o eficiencias relativas globales (Verma et al., 2009), siendo el objetivo primordial del presente trabajo.

METODOLOGÍA

Base de datos

Se llevó a cabo la recopilación exhaustiva de la composición química (11 elementos mayores y 71 elementos traza, incluyendo REE) para 35 MRG provenientes de cuatro países: (a) Canadá (*Energy, Mines and Resources*) – gabro MRG-1 y sienitas SY-2 y SY-3 (Abbey, 1979; Gladney y Roelandts, 1990); (b) E.U.A. (*U. S. Geological Survey; USGS*) – basalto BIR-1 y diabasa W-2 (Gladney, 1988), gabro GSM-1 (Flanagan, 1986), basalto BHVO-1, cuarzo QLO-1 y riolita RGM-1 (Gladney y Roelandts, 1988), dunita DTS-1, diabasa W-1 (Gladney et al., 1991) y granito G-2 (Gladney et al., 1992); (c) Japón (*Geological Survey of Japan, GSJ*) – <http://riodb02.ibase.aist.go.jp/earthsci/welcome.html> – para andesitas JA-1, JA-2 y JA-3, basaltos JB-1, JB-1a, JB-2 y JB-3, feldespatos JF-1 y JF-2, granodioritas JG-1 y JG-1a, granitos JG-2 y JG-3, hornblendita JH-1, peridotita JP-1, y riolitas JR-1 y JR-2; y (d) Sudáfrica (*National Institute of Metallurgy, NIM*) – dunita NIM-D, granito NIM-G, hujavarita NIM-L, norita (gabro) NIM-N, pyroxenita NIM-P y sienita NIM-S (Steele et al., 1972, 1978). Los datos fueron capturados en el software comercial *Statistica*® que es una herramienta estadística para el manejo de gran cantidad de datos y permite exportar o importar datos en otros formatos como *Excel*®.

Aplicación estadística

La detección de datos desviados en muestras univariadas se realizó aplicando el programa DODESYS (*Discordant Outlier Detection and Elimination SYStem*; Verma y Díaz-González, en preparación), el cual incluye 15 pruebas de discordancia y sus 33 variantes (nueve pruebas sencillas y siete pruebas múltiples), considerando que la prueba N4 pertenece a ambas clasificaciones (sencilla una variante $N4_{(p=1)}$ y múltiple tres variantes $N4_{(p=2,3,4)}$). Además, DODESYS utiliza, en su versión actual, los mejores valores críticos, precisos y exactos, para un nivel de confianza estricto de 99%, simulados recientemente por Verma et al.

(2008). Una versión anterior fue programada con valores de Verma y Quiroz-Ruiz (2006a, 2006b) mientras que una posterior (UDASYS; *Univariate Data Analysis SYStem*; Verma y Díaz-González, en preparación), más actualizada, se basaría en ecuaciones para diferentes niveles de confianza en vez de tablas (Verma y Quiroz-Ruiz, 2008), e involucraría adicionalmente pruebas de significancia (Verma, 2009b) y cálculos de parámetros robustos para el manejo de datos experimentales.

Para un arreglo de n datos univariados $x_1, x_2, x_3, \dots, x_{n-1}, x_n$, suponemos que el arreglo ordenado de estos datos es dado por $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n-2)}, x_{(n-1)}, x_{(n)}$. En el presente trabajo, los datos analíticos para un elemento químico generados por un grupo de métodos en diferentes laboratorios para un MRG, representan un arreglo estadístico o una muestra estadística. Todos los métodos analíticos fueron agrupados de acuerdo con los criterios físicos en los cuales se basaron Velasco-Tapia et al. (2001), formándose un total de ocho grupos de métodos: (1) clásicos (químicos, gravimétricos y volumétricos); (2) absorción atómica; (3) fluorescencia de rayos-X; (4) espectrometría de emisión; (5) nucleares; (6) espectrometría de masas; (7) cromatográficos; y (8) misceláneos (que no pertenecen a ninguno de los grupos anteriores).

Para la evaluación de las pruebas de discordancia se emplea el arreglo ordenado $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n-2)}, x_{(n-1)}, x_{(n)}$ el cual nos permite identificar valores extremos (*outliers*) susceptibles a probar como valores discordantes. Siguiendo a Verma y Quiroz-Ruiz (2006a, 2006b, 2008), se utilizó, para pruebas sencillas ($k=1$), la terminología "Upper" (U), cuando se evalúa al valor más alto en una serie de datos, es decir, $x_{(n)}$, p. ej., para N1, se dice N1U. De igual manera, el término "Lower" (L), fue usado cuando se trataba del valor más bajo $x_{(1)}$, p. ej., para N1, sería N1L. Para un valor extremo cuando la prueba está evaluando el valor más alejado de los dos extremos, es decir, $x_{(n)}$ o $x_{(1)}$, p. ej., para N2, sería N2E, pero, por simplicidad, se conservó el nombre original N2. Al contrario de las pruebas tipo "Upper" o "Lower", también llamadas de una cola, las de un valor "extremo" fueron denominadas como de dos colas por Barnett y Lewis (1994). Cabe aclarar también que para más de un dato a probar ($k=2-4$), la terminología fue similar, aunque algo más compleja. Por ejemplo, una prueba múltiple, como la N3, cuando evalúa dos datos a la vez, fue expresada como N3U2 (U para "Upper" y 2 para k) y como N3L2 (L para "Lower" y 2 para k). De igual manera, la prueba múltiple N4 se denominaría como N4U4 cuando se evalúan 4 datos a la vez en el lado más alto de la muestra estadística. No se utilizó la terminación de k cuando se trató de una prueba sencilla; en otras palabras, por simplicidad a N1, p. ej., se le denominó NIU y no NIU1. Finalmente, cuando se necesitó referir al resultado de la combinación de estas dos pruebas (tipo U y L) para un determinado k , se utilizó el signo de +, por ejemplo, para la prueba N3 con $k=2$ se diría, N3U2+N3L2.

Las pruebas de discordancia, para ser aplicadas, re-

quieren el "número mínimo de datos" que varía del 3 (p. ej., N1U) hasta el 9 (p. ej., N3U4 o N3L4). Se decidió que, para evaluar y comparar las pruebas de discordancia de la forma más objetiva, era necesario realizar la aplicación en exactamente los mismos tamaños de muestras. Por lo tanto, de la base de datos de MRG se separaron todos los casos con $n \geq 9$, siendo un total de 2220 casos con estas características. De esta manera, un "caso" (una muestra estadística) se define como un conjunto de datos de un MRG para un elemento químico que ha sido analizado por un determinado grupo de métodos analíticos.

Con el fin de comparar las eficiencias o potencias de las pruebas, se definieron dos parámetros estadísticos (REC y ROC), los cuales se describen en la sección de "Criterios cuantitativos de comparación". Para comparar las eficiencias relativas mediante el primer parámetro REC, las pruebas de discordancia fueron clasificadas en 14 categorías, mostradas en la Tabla 1. Las primeras cinco categorías (1–5) constan de pruebas de discordancia sencillas ($k=1$). De esta manera, la categoría 5 se formó de todas las pruebas sencillas (N1, N2, N4, N7, N8, N9, N10, N14 y N15). Las pruebas múltiples ($k=2-4$), por su parte, fueron agrupadas en ocho categorías (6–13). Por ejemplo, en la categoría 6 existen todas las pruebas múltiples que evalúan los dos

datos ($k=2$) más altos (N3U2, N4U2, N11U2, N12U2 y N13U2). La categoría 13 consta de todas las pruebas múltiples. Finalmente, todas las 15 pruebas de discordancia con sus 33 variantes fueron comparadas en la categoría 14. De igual manera, se establecieron cuatro categorías (A–D) para el parámetro ROC (Tabla 1), el cual nos permitió una comparación más justa que el parámetro REC para las categorías 13 y 14.

Por otra parte, para calcular los parámetros de eficiencia en una iteración se estableció, de forma arbitraria, que el número mínimo de casos aplicables sea, al menos, 30 para ser considerado como representativo para estos cálculos, aunque esto no es tan arbitrario ya que el número 30 es estadísticamente significativo. Cuando los casos "aplicables" para la evaluación de una prueba no cumplían con la condición de ≥ 30 , no se calculó ningún parámetro de eficiencia. De esta manera, llegaron a finalizar las iteraciones para una determinada prueba de discordancia.

Una evaluación estadística adicional consistió en usar regresiones lineales empleando el programa OYNYL (Verma *et al.*, 2006) y posteriormente en aplicar las pruebas F de Fisher y t de Student (Jensen *et al.*, 2000; Verma, 2005, 2009b) para la comparación estadística de las pendientes de las regresiones obtenidas por OYNYL.

Tabla 1. Categorización de las pruebas de discordancia definida para el presente estudio comparativo de la eficiencia relativa de 15 pruebas de discordancia con 33 variantes para muestras estadísticas normales univariadas.

No.	Categorización de las pruebas de discordancia (según REC y ROC) Descripción de la categoría	Número de pruebas / sus variantes	Código de las pruebas de discordancia
Según REC			
1	Pruebas sencillas que evalúan el dato más alto	5 / 5	N1U, N4U, N7U, N9U y N10U
2	Pruebas sencillas que evalúan el dato más bajo	4 / 4	N1L, N4L, N9L y N10L
3	Pruebas sencillas que evalúan un dato extremo	4 / 4	N2, N8, N14 y N15
4	Pruebas sencillas con dos versiones: más alto y más bajo	4 / 8	N1, N4, N9 y N10
5	Todas las pruebas sencillas	9 / 13	N1, N2, N4, N7, N8, N9, N10, N14 y N15
6	Pruebas múltiples que evalúan los dos datos más altos	5 / 5	N3U2, N4U2, N11U2, N12U2 y N13U2
7	Pruebas múltiples que evalúan los dos datos más bajos	5 / 5	N3L2, N4L2, N11L2, N12L2 y N13L2
8	Pruebas múltiples que evalúan tanto el dato más alto como el más bajo	2 / 2	N5U y N6U
9	Pruebas múltiples que evalúan los tres datos más altos	2 / 2	N3U3 y N4U3
10	Pruebas múltiples que evalúan los tres datos más bajos	2 / 2	N3L3 y N4L3
11	Pruebas múltiples que evalúan los cuatro datos más altos	2 / 2	N3U4 y N4U4
12	Pruebas múltiples que evalúan los cuatro datos más bajos	2 / 2	N3L4 y N4L4
13	Todas las pruebas múltiples	7 / 20	N3, N4, N5, N6, N11, N12 y N13
14	Todas las pruebas, tanto sencillas como múltiples	15 / 33	N1, N2, N3, N4, N5, N6, N7, N8, N9, N10, N11, N12, N13, N14 y N15
Según ROC			
A	Todas las pruebas que evalúan de uno hasta cuatro valores más altos y más bajos	9 / 14	N1U, N3U2, N3U3, N3U4, N4U, N4U2, N4U3, N4U4, N7U, N9U, N10U, N11U2, N12U2 y N13U2
B	Todas las pruebas que evalúan de uno hasta cuatro valores más bajos	9 / 14	N1L, N3L2, N3L3, N3L4, N4L2, N4L3, N4L4, N7L, N9L, N10L, N11L2, N12L2 y N13L2
C	Todas las pruebas que evalúan un valor extremo o la combinación de dos valores más alto y más bajo	6 / 6	N2, N5, N6, N8, N14 y N15
D	Todas las pruebas que evalúan de uno hasta cuatro valores más altos	15 / 33	N1, N2, N3, N4, N5, N6, N7, N8, N9, N10, N11, N12, N13, N14 y N15

CRITERIOS CUANTITATIVOS DE COMPARACIÓN

En este trabajo se utilizaron parámetros estadísticos, tanto el existente "criterio de eficiencia relativa" (REC; Velasco *et al.*, 2000; Verma *et al.*, 2009) como el nuevo "criterio de datos desviados relativo" (ROC). Además, se aplicó una manera distinta de evaluar la eficiencia de las pruebas mediante regresiones lineales y pruebas de significancia de F y t. El suplemento electrónico del presente trabajo proporciona detalles sobre estos parámetros y, a continuación, se presenta una síntesis.

El criterio de eficiencia relativa (*relative efficiency criterion*, REC), definido con anterioridad por Velasco *et al.* (2000) y Verma *et al.* (2009), es un parámetro estadístico que expresa la potencia que tiene una prueba de discordancia para detectar y eliminar datos desviados en una muestra estadística. La fórmula para el REC global $(REC)_g$ en una prueba de discordancia se define de la siguiente manera:

$$(REC)_g = \frac{\sum_{i=1}^m \text{número de casos exitosos en la iteración } i}{\text{número total de casos "aplicables"}} \times 100 \quad (1)$$

donde un "caso" (una muestra estadística) es un conjunto de datos de un MRG para un elemento químico que ha sido determinado por un grupo de métodos analíticos; el "éxito" indica el número de casos en los cuales, en una determinada iteración, se detectaron 1 ó 2-4 datos desviados, para pruebas tipo $k=1$ o $k=2-4$, respectivamente; y m es el total de iteraciones. Para mayores detalles sobre la definición de REC para cada iteración $(REC)_1, (REC)_2$, etc., ver el suplemento electrónico. En realidad, la expresión matemática reporta el porcentaje de eficiencia que tiene la prueba para detectar datos desviados. Un ejemplo detallado del cálculo del parámetro REC se presenta en el suplemento electrónico (ver Tabla A1 y la explicación correspondiente).

El criterio de datos desviados relativo (*relative outlier criterion*, ROC) se define, por primera vez, en el presente trabajo e indica el porcentaje de datos desviados que se ha depurado en una muestra, después de haberse aplicado una prueba de discordancia. El ROC global $(ROC)_g$ se puede calcular por:

$$(ROC)_g = \frac{\sum_{i=1}^m \text{número de datos "desviados" en la iteración } i}{\text{número total de datos de todos los casos}} \times 100 \quad (2)$$

La utilidad de esta expresión radica en que se tiene la posibilidad de conocer el número de datos desviados que fueron depurados en la totalidad de las muestras estadísticas iniciales. Además, $(ROC)_g$ proporciona una comparación más justa para algunas categorías en comparación con el parámetro $(REC)_g$ como se verá en la siguiente sección.

RESULTADOS

Organizamos nuestra presentación de acuerdo con los parámetros $(REC)_g$ y $(ROC)_g$ y las distintas categorías

formadas para las pruebas, seguido por la combinación de regresiones lineales y las pruebas de significancia de F de Fisher y t de Student.

Criterio de eficiencia relativa (REC)

Para esta evaluación las pruebas de discordancia fueron clasificadas en 14 categorías (Tabla 1), definidas con la finalidad de comparar los parámetros de eficiencia relativa. El parámetro REC fue calculado para todas las categorías. Se presentan los resultados de estos cálculos en las Figuras A1-A11 (en el suplemento electrónico) y Figuras 1-3 (en este trabajo); en todas ellas (excepto Figuras 1 y 3) se emplea la misma escala en el eje "y" del REC, a fin de facilitar su comparación visual. Cabe aclarar que todas las Figuras y Tablas con numeración iniciada en "A" se ubican en el suplemento electrónico.

Categoría 1: N1U, N4U, N*U, N9U y N10U (Figura A1): Pruebas sencillas que evalúan el dato más alto $x_{(n)}$

En la primera iteración (iteración 1, Figura A1a) aplicada a 2220 casos, las pruebas tipo Grubbs (N1U y N4U) presentaron valores de $(REC)_1$ de 25.59% y 25.63%, respectivamente, mayores que las pruebas tipo Dixon (N7U, N9U y N10U; 19.81%, 21.04% y 20.90%, respectivamente). Adicionalmente, los valores de los parámetros $(REC)_2$ y $(REC)_3$ (iteraciones 2 y 3, Figura A1b-c) para las pruebas tipo Grubbs fueron significativamente mayores (20.21%-27.83%) que los valores obtenidos para las pruebas tipo Dixon (8.62%-12.20%). En la cuarta iteración (iteración 4, Figura A1d), únicamente las dos pruebas tipo Grubbs presentaron un número significativo de casos "aplicables" (≥ 30), por lo cual el parámetro $(REC)_4$ se calculó sólo para estas pruebas de discordancia, resultándose un valor de 34.4% para ambas. Para las pruebas tipo Dixon (N7U, N9U y N10U), los casos "aplicables" para la iteración 4 fueron solamente 4-5, por lo cual no se calculó $(REC)_4$. Estas consideraciones fueron aplicadas de manera similar en la descripción de las categorías subsecuentes (categorías 2-14).

Categoría 2: N1L, N4L, N9L y N10L (Figura A2): Pruebas sencillas que evalúan el dato más bajo $x_{(1)}$

En la primera iteración (iteración 1, Figura A2a) las pruebas tipo Grubbs (N1L y N4L) presentaron valores de $(REC)_1$ de 11.22% para ambas, ligeramente menores que la prueba tipo Dixon N10L (11.31%) y mayores que la prueba tipo Dixon N9L (10.45%). Para fines prácticos, se puede considerar que, de acuerdo con este parámetro, las pruebas N1L, N4L y N10L son parecidas y sólo ligeramente mejores que N9L. No obstante, en la iteración 2 (Figura A2b) las pruebas tipo Grubbs presentaron valores de $(REC)_2$ significativamente mayores (14.9% para ambas N1L y N4L) que las pruebas tipo Dixon N9L y N10L (6.5% y 6.4%). En la tercera iteración (Figura A1c), únicamente las pruebas

tipo Grubbs tuvieron valores de $(REC)_3$ que fueron 16.2% para ambas.

Categoría 3: N2, N8, N14 y N15 (Figura A3):

Pruebas sencillas que evalúan un dato extremo $x_{(1)}$ o $x_{(n)}$

En la Figura A3a (iteración 1) se puede apreciar que la prueba del coeficiente de exceso o curtosis N15 presentó el valor del $(REC)_3$ mayor (34.10%), seguido por la prueba tipo Grubbs N2 (31.94%) y por la prueba del coeficiente de asimetría o sesgo N14 (27.48%). Mientras tanto, la prueba tipo Dixon N8 dió el valor (25.50%) de $(REC)_3$ menor que N14. En las iteraciones 2 y 3 (Figura A3b-c), las pruebas N14 y N15 mostraron valores de $(REC)_2$ y $(REC)_3$ (29.67%-40.33%) mayores que las pruebas tipo Grubbs N2 (22.43% y 27.04%) y tipo Dixon N8 (12.19% y 8.70%). En la iteración 4 (Figura A3d), la prueba tipo Dixon N8 únicamente fue aplicable a seis casos (que corresponden al número de casos "exitosos" en la iteración 3) mientras que las pruebas N2, N14 y N15 fueron aplicadas a un número significativo de casos (43, 73 y 86, respectivamente) y presentaron valores significativamente grandes del $(REC)_4$ (34.9%, 39.7% y 46.5%, respectivamente). Finalmente, en la iteración 5 (Figura A3e) se puede observar que las pruebas N14 y N15 presentaron valores del $(REC)_5$ significativamente grandes (44.8% y 42.5%).

Categoría 4: N1, N4, N9 y N10 (Figura A4):

Todas las pruebas sencillas con dos versiones (más alto y más bajo) $x_{(1)}, x_{(n)}$

En esta evaluación se sumaron los casos "exitosos" obtenidos en una iteración, de ambas versiones (más alto y más bajo) de las pruebas de discordancia que integran esta categoría. En la iteración 1 (Figura A4a), las pruebas tipo Grubbs (N1 y N4) presentaron valores del $(REC)_1$ muy parecidos (18.40% y 18.42%) mientras que las pruebas tipo Dixon (N9 y N10) dieron valores menores (15.74% y 16.10%). En la iteración 2 (Figura A4b), los valores de $(REC)_2$ para las pruebas tipo Grubbs fueron significativamente mayores (18.60% para N1 y 18.58% para N4) que los de las pruebas tipo Dixon (8.15% para N9 y 7.83% para N10). En la iteración 3 (Figura A4c), los valores de $(REC)_3$ para las pruebas tipo Grubbs N1 y N4 fueron aún mayores (ambos 25%) que los de las de Dixon N9 y N10 (12.2% y 10.0%, respectivamente). Finalmente, en la cuarta iteración (Figura A4d) el parámetro $(REC)_4$ para las pruebas tipo Grubbs (N1 y N4) proporcionó valores altos de 34%.

Categoría 5 (Figura 1): Todas las pruebas sencillas (N1, N2, N4, N7, N8, N9, N10, N14 y N15)

Esta categoría incluye todas las pruebas sencillas, es decir, aquellas que evalúan un sólo dato discordante: N1 (con ambas versiones: dato más alto y más bajo), N2 (extremo), N4 (con ambas versiones: más alto y más bajo), N7 (más alto), N8 (extremo), N9 (con ambas versiones: más alto y más bajo), N10 (con ambas versiones: más alto y más bajo), N14 (extremo, coeficiente de asimetría "skewness")

y N15 (extremo, coeficiente de exceso "kurtosis"). Con la finalidad de realizar una comparación "justa" de todas las pruebas de discordancia de esta categoría, se contaron todos los casos "exitosos" en todas las iteraciones de una prueba determinada (incluyendo ambas versiones, más alto y más bajo). Se consideró como casos "aplicables" el número de casos inicial (2220 casos con $n \geq 9$ datos cada uno). Los resultados del parámetro REC global $(REC)_g$ (ver ecuación 1) de las pruebas de discordancia de esta categoría (Figura 1) mostraron la siguiente secuencia (en orden descendente): (i) la prueba de dos colas basada en el coeficiente de exceso o curtosis N15 (51.58%); (ii) la prueba de una cola tipo Grubbs N4 (46.13%); (iii) la prueba de una cola tipo Grubbs N1 (46.08%); (iv) la prueba de dos colas tipo Grubbs N2 (41.85%); (v) la prueba de dos colas basada en el coeficiente de asimetría o sesgo N14 (41.08%); (vi) la prueba de una cola tipo Dixon N10 (34.91%); (vii) la prueba de una cola tipo Dixon N9 (34.28%); (viii) la prueba de dos colas tipo Dixon N8 (28.87%); y (ix) la prueba de una cola tipo Dixon N7 (21.94%). En otras palabras, la eficiencia o la potencia de las pruebas de discordancia sencillas tiene el orden siguiente: $N15 > N4 = N1 > N2 = N14 > N10 = N9 > N8 > N7$.

Categoría 6: N3U2, N4U2, N11U2, N12U2 y N13U2

(Figura A5): Pruebas que evalúan los dos datos más altos $x_{(n-1)}, x_{(n)}$

En la primera iteración (Figura A5a), los valores de $(REC)_1$ presentaron la siguiente secuencia (el valor del $(REC)_1$ del mayor al menor): la prueba tipo Grubbs N4U2 (28.60%), la prueba tipo Dixon N13U2 (26.17%), la prueba tipo Dixon N12U2 (26.13%), la prueba tipo Dixon

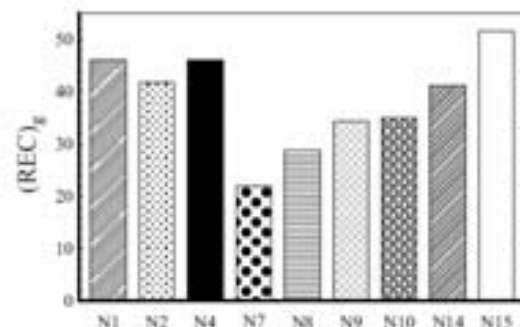


Figura 1. Criterio de eficiencia relativa global $(REC)_g$ para la categoría 5 (ver Tabla 1 para la categorización) que incluye todas las pruebas de discordancia sencillas. Esta categoría se compone de las siguientes pruebas: N1 (con ambas versiones: más alto y más bajo; N1U+N1L), N2 (tipo extremo), N4 (con ambas versiones de tipo $h=1$: más alto y más bajo; N4U+N4L), N7 (más alto; N7U), N8 (tipo extremo), N9 (con ambas versiones: más alto y más bajo; N9U+N9L), N10 (con ambas versiones: más alto y más bajo; N10U+N10L), N14 (tipo extremo) y N15 (tipo extremo). En esta categoría se calculó el REC combinando todos los datos discordantes detectados en todas las iteraciones de cada prueba de discordancia y se consideró como "casos aplicables" al número de casos inicial (2220 casos con $n \geq 9$ datos cada uno).

N11U2 (24.82%) y finalmente, la prueba tipo Grubbs N3U2 (18.47%). Mientras tanto, en la segunda iteración (Figura A5b) los valores de $(REC)_2$ para las pruebas tipo Grubbs (N3U2; 16.10% y N4U2; 11.65%) fueron mayores que los valores obtenidos para las pruebas tipo Dixon N11U2, N12U2 y N13U2 (6.17%, 6.21% y 4.99%, respectivamente). En la iteración 3 (Figura A5c), nuevamente las pruebas tipo Grubbs (N3U2 y N4U2) presentaron valores significativamente mayores (19.7% y 17.6%) que la prueba tipo Dixon N12 (2.8%). El resto de las pruebas tipo Dixon (N11U2 y N13U2) no presentaron un número de casos “aplicables” significativo para esta iteración.

Categoría 7: N3L2, N4L2, N11L2, N12L2 y N13L2 (Figura A6): Pruebas que evalúan los dos datos más bajos $x_{(1)}, x_{(2)}$

En la primera iteración (Figura A6a) las pruebas tipo Dixon N12L2 y N13L2 presentaron los valores de $(REC)_1$ ligeramente mayores (12.16% y 13.24%) que el resto de las pruebas N3L2, N4L2 y N11L2 (con 8.51%, 12.07% y 10.68%, respectivamente). Sin embargo, en la segunda iteración (Figura A6b) los valores de $(REC)_2$ fueron mayores para las pruebas tipo Grubbs N3L2 (7.94%) y N4L2 (5.22%) que las de tipo Dixon N11L2 (2.11%), N12L2 (1.48%) y N13L2 (2.38%). El número de casos “aplicables” en la tercera iteración no fueron significativos.

Categoría 8: N5UL y N6UL (Figura A7): Pruebas que evalúan el dato más alto y el más bajo $x_{(1)}, x_{(n)}$

Las pruebas de discordancia que integran esta categoría son solamente las dos pruebas tipo Grubbs. En la primera iteración (Figura A7a) la prueba N5UL presentó un valor mayor (29.86%) que la prueba N6UL (16.67%). En la iteración 2 (Figura A7b) el valor de $(REC)_2$ de ambas pruebas fueron similares (N5UL: 15.84% y N6UL: 15.14%). Mientras tanto, en la iteración 3 (Figura A8c) se puede apreciar que la prueba N6UL (21.4%) presentó un valor ligeramente mayor de $(REC)_3$ que la prueba N5UL (20.0%).

Categoría 9: N3U3 y N4U3 (Figura A8): Pruebas que evalúan los tres datos más altos $x_{(n)}, x_{(n-1)}, x_{(n-2)}$

Estas pruebas múltiples tipo Grubbs pudieron lograr una sola iteración. Por lo tanto, únicamente se calculó el $(REC)_1$. La prueba N3U3 presentó un valor de $(REC)_1$ significativamente menor (10.95%) que la prueba N4U3 (27.03%).

Categoría 10: N3L3 y N4L3 (Figura A9): Pruebas que evalúan los tres datos más bajos $x_{(1)}, x_{(2)}, x_{(3)}$

Al igual que la categoría 9, las pruebas lograron sólo una iteración. La prueba N3L3 presentó un valor de $(REC)_1$ significativamente menor (4.77%) que la prueba N4L3 (11.49%). El resultado en esta categoría coincidió con el de la categoría 9 ya que la prueba N4 presentó mayor eficiencia que la N3.

Categoría 11: N3U4 y N4U4 (Figura A10): Pruebas que evalúan los cuatro datos más altos $x_{(n)}, x_{(n-1)}, x_{(n-2)}, x_{(n-3)}$

La eficiencia de la primera iteración $(REC)_1$ para la prueba N3U4 presentó un valor significativamente menor (6.76%) que para la prueba N4U4 (25.01%), consistente al resultado de las variantes de $k=2$ y 3 (categorías 9 y 10).

Categoría 12: N3L4 y N4L4 (Figura A11): Pruebas que evalúan los cuatro datos más bajos $x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}$

El valor del $(REC)_1$ para la prueba N3L4 (2.61%) fue significativamente menor al de la prueba N4L4 (10%). Esto nuevamente sugiere que la prueba N4 es más eficiente que la prueba N3.

Categoría 13 (Figura 2): Todas las pruebas de discordancia múltiples (N3, N4, N5, N6, N11, N12 y N13) con sus variantes

La presente categoría se encuentra integrada por todas las pruebas de discordancia múltiples (estas pruebas evalúan de dos hasta cuatro datos a la vez, $k=2-4$). Con el fin de realizar una justa comparación de las pruebas en esta categoría, se sumaron los casos “exitosos” de todas las iteraciones realizadas por una determinada prueba (incluyendo ambas versiones que evalúan tanto los datos más altos como los más bajos). La comparación de las pruebas de esta categoría requirió el cálculo del parámetro de eficiencia relativa global $(REC)_g$ (ecuación 1). Los valores de $(REC)_g$ de las pruebas múltiples (Figura 2) mostraron la siguiente características: (i) la prueba tipo Grubbs de una cola N3 (con tres combinaciones de seis variantes N3U2+N3L2, N3U3+N3L3 y N3U4+N3L4; 31.58, 15.72 y 9.37%, respectivamente); (ii) la prueba tipo Grubbs de una cola N4 (con tres combinaciones de

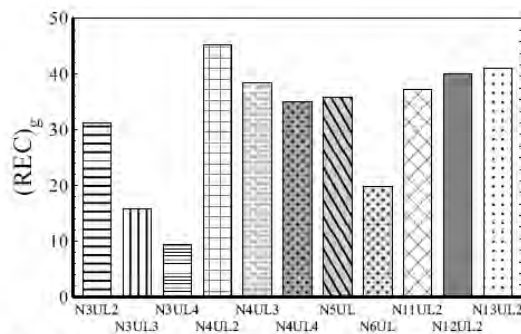


Figura 2. Criterio de eficiencia relativa global $(REC)_g$ de la categoría 13 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas de discordancia múltiples: N3UL2 (o sea, combinación de tipo “Upper” y “Lower” N3U2+N3L2), N3UL3 (N3U3+N3L3), N3UL4 (N3U4+N3L4), N4UL2 (N4U2+N4L2), N4UL3 (N4U3+N4L3), N4UL4 (N4U4+N4L4), N5UL (prueba tipo extremo con $k=2$, o sea los dos datos extremos a la vez), N6UL (prueba tipo extremo con $k=2$), N11UL2 (N11U2+N11L2), N12UL2 (N12U2+N12L2) y N13UL2 (N13U2+N13L2). Se calculó el REC considerando todos los casos “exitosos” en todas las iteraciones de cada prueba de discordancia.

seis variantes N4U2+N4L2, N4U3+N4L3 y N4U4+N4L4; 45.36, 38.51 y 35.09%, respectivamente); (iii) la prueba tipo Grubbs de dos colas N5 (N5UL: 35.86%); (iv) la prueba tipo Grubbs de dos colas N6 (N6UL: 19.82%); (v) la prueba tipo Dixon de una cola N11U2+N11L2 (37.25%); (vi) la prueba tipo Dixon de una cola N12U2+N12L2 (40.14%); y (vii) la prueba tipo Dixon de una cola N13U2+N13L2 (41.04%).

Si comparamos las pruebas de discordancia múltiples para un determinado valor de k , se puede inferir el siguiente orden en la eficiencia relativa $(REC)_g$ de las pruebas: (i) pruebas que evalúan dos datos a la vez ($k=2$): N4U2+N4L2 > N13U2+N13L2 > N12U2+N12L2 > N11U2+N11L2 > N5UL > N3U2+N3L2 > N6UL; (ii) pruebas que evalúan tres valores ($k=3$): N4U3+N4L3 > N3U3+N3L3; y (iii) pruebas que evalúan cuatro valores ($k=4$): N4U4+N4L4 > N3U4+N3L4.

Categoría 14 (Figura 3): Todas las pruebas de discordancia sencillas y múltiples

Los resultados finales de la presente evaluación empírica de 15 pruebas de discordancia con 33 variantes mediante el parámetro $(REC)_g$ fueron calculados de manera similar que en las categorías 5 y 13. Se obtuvo el número total de casos “exitosos” (numerador de la ecuación 1) sumando todos los casos “exitosos” en todas las iteraciones de una prueba determinada que incluyó ambas versiones de datos desviados (tipo más alto y más bajo). Por ejemplo, para la prueba N1 se sumaron los casos “exitosos” de N1U y N1L, por lo cual aquí N1 (Tabla 1) representa N1U+N1L; N3UL2 equivale a N3U2+N3L2; etc. De esta manera, se obtuvieron 20 pruebas con variantes de prueba combinadas (Figura 3). La secuencia de las eficiencias $(REC)_g$ para

estas variantes combinadas fueron: N15 > N4U+N4L ≈ N1U+N1L ≈ N4U2+N4L2 > N2 ≈ N14 ≈ N13U2+N13L2 ≈ N12U2+N12L2 > N4U3+N4L3 ≈ N11U2+N11L2 > N5UL ≈ N4U4+N4L4 ≈ N10U+N10L ≈ N9U+N9L > N3U2+N3L2 > N8 > N7 > N6UL > N3U3+N3L3 > N3U4+N3L4.

Necesidad de un nuevo criterio estadístico: Criterio relativo de datos desviados (relative outlier criterion, ROC)

Es importante señalar que el parámetro REC es útil para comparar justamente las pruebas en la mayoría de las 14 categorías, excepto en las categorías 13 y 14. En estas dos categorías, debido a la comparación de pruebas con distintos valores de k (1–4), el número de datos probados a la vez no se toma en cuenta correctamente. El parámetro REC es simplemente la relación de casos “exitosos” con casos “aplicables” (ecuación 1). Es deseable, por lo tanto, que el parámetro estadístico tome en cuenta tanto el número de datos desviados en casos “exitosos” como el total de los datos en los casos “aplicables”. Se formuló un nuevo criterio llamado criterio relativo de datos desviados $(ROC)_g$ (el subíndice g representa “global”) para la evaluación de las 15 pruebas de discordancia con 33 variantes. La evaluación basada en este nuevo parámetro estadístico $(ROC)_g$ se presenta a continuación.

La determinación de este parámetro $(ROC)_g$ (ecuación 2) para una prueba se obtiene al dividir el número de datos desviados detectados por la prueba entre el número total de datos que contienen los 2220 casos “aplicables”, siendo un total de 41,821 datos individuales geoquímicos. Debido a la definición en términos de la relación de datos desviados entre el total de datos, este parámetro es totalmente libre de la unidad, por lo cual cuando en una categoría se comparan pruebas con distintos k (1–4), $(ROC)_g$ pudiera ser más objetiva y justa para realizar una comparación que el parámetro $(REC)_g$, que aunque se basa también en la relación de número de casos y es libre de unidad, no toma en cuenta el valor de k (1–4). Por todo esto, se puede afirmar que el valor de $(ROC)_g$ permite conocer el número de datos totales que una prueba de discordancia detecta y elimina en las muestras estadísticas en función de todos los datos presentes. Para la aplicación de $(ROC)_g$, se formaron cuatro categorías de pruebas (Tabla 1).

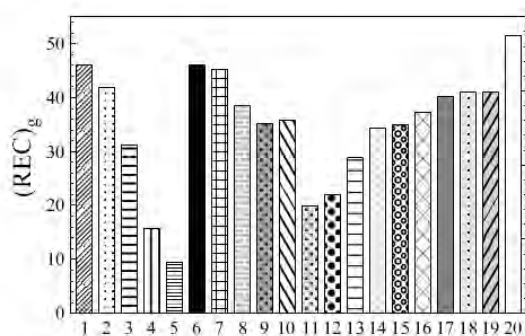


Figura 3. Criterio de eficiencia relativa global $(REC)_g$ de la categoría 14 (ver Tabla 1 para categorización) integrada por todas las pruebas de discordancia. Esta categoría es una combinación de la categoría 5 (la cual incluye 9 pruebas con 9 combinaciones de 13 variantes) y la categoría 13 (que incluye 7 pruebas con 11 combinaciones de 20 variantes): 1–N1 (N1U+N1L), 2–N2, 3–N3UL2 (N3U2+N3L2), 4–N3UL3 (N3U3+N3L3), 5–N3UL4 (N3U4+N3L4), 6–N4UL (N4U+N4L), 7–N4UL2 (N4U2+N4L2), 8–N4UL3 (N4U3+N4L3), 9–N4UL4 (N4U4+N4L4), 10–N5UL, 11–N6UL, 12–N7 (N7U), 13–N8, 14–N9 (N9U+N9L), 15–N10 (N10U+N10L), 16–N11UL2 (N11U2+N11L2), 17–N12UL2 (N12U2+N12L2), 18–N13UL2 (N13U2+N13L2), 19–N14 y 20–N15.

Categoría A (Figura A12a): $(ROC)_g$ para pruebas que detectan desde uno hasta cuatro valores discordantes altos $x_{(n)}$, $x_{(n-1)}$, $x_{(n-2)}$, $x_{(n-3)}$

En la categoría A fueron incluidas todas las pruebas de discordancia que evaluaron a los valores más altos. Esta categoría cuenta con cinco variantes de cuatro pruebas sencillas (N1U, N4U, N7U, N9U y N10U) y nueve variantes de cinco pruebas múltiples (de tipo $k=2$: N3U2, N4U2, N11U2, N12U2 y N13U2; de $k=3$: N3U3 y N4U3; y de $k=4$: N3U4 y N4U4). En la Figura A12a se observa que la prueba múl-

tiple N4, en sus tres versiones ($k=2-4$), presenta valores de $(ROC)_g$ de 3.46%-5.33%, significativamente mayores que todas las otras pruebas en esta categoría (1.16%-2.95%). El valor mayor del $(ROC)_g$ corresponde a la prueba tipo Grubbs N4U4. En su versión de $k=4$, debido a que esta prueba detectó un total de 2228 datos discordantes en un total de 41,821 datos individuales geoquímicos. Por otra parte, la prueba N7U tipo Dixon, con el menor valor del $(ROC)_g$, detectó un total de solamente 487 datos discordantes en el mismo total de 41,821 datos individuales geoquímicos.

Esto aclara también que, aunque el parámetro $(REC)_g$ referente a casos o muestras estadísticas, presentó valores de entre 21.94% y 32.84% para pruebas sencillas tipo "Upper" y entre 6.76% y 27.79% para pruebas múltiples tipo "Upper", el $(ROC)_g$ concierne a los datos desviados se encuentra entre valores significativamente menores (1.16%-5.33%). Esto significa que, en los datos geoquímicos interlaboratorios de MRG, relativamente pocos datos se identificaron como valores discordantes y que la gran mayoría de los datos restantes (94.68%-98.84%) representan fielmente muestras estadísticas normales sin "contaminación" estadística.

Finalmente, se infiere que el orden de $(ROC)_g$ para pruebas de discordancia que evaluaron a los valores más altos fue el siguiente: N4U4 > N4U3 > N4U2 > N12U2 ≈ N13U2 > N11U2 > N3U2 > N3U3 = N1U ≈ N4U > N3U4 > N9U ≈ N10U > N7U.

Categoría B (Figura A12b): $(ROC)_g$ para pruebas que detectan desde uno hasta cuatro valores discordantes bajos $x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}$

Al igual que la categoría A, aquí se consideraron todas las pruebas de discordancia de tipo sencillo y múltiple que evalúan al valor más bajo. La prueba N4 en su versión de $k=4$ representó el valor del $(ROC)_g$ más alto aunque solamente de 2.12%, equivalente a sólo 886 datos discordantes en un total de 41,821 datos individuales geoquímicos. Por su parte, en esta categoría la prueba N3L4 presentó el menor valor de $(ROC)_g$, de 0.55%, equivalente a 232 datos en toda la base de 41,821 datos individuales geoquímicos analizada en este trabajo. En comparación, el parámetro anterior $(REC)_g$ referente a casos o muestras estadísticas, por su parte, presentó valores de entre 11.17% y 13.24% para pruebas sencillas tipo "Lower" y entre 2.61% y 13.56% para pruebas múltiples tipo "Lower".

El orden de $(ROC)_g$ para pruebas de discordancia con valores bajos fue el siguiente: N4L4 > N4L3 > N13L2 > N4L2 > N12L2 > N11L2 > N3L2 > N3L3 > N4L ≈ N1L > N10L2 > N9L > N3L4.

Categoría C (Figura A12c): $(ROC)_g$ para pruebas que detectan desde uno o dos valores discordantes extremos $x_{(1)}$ o $x_{(n)}$ o ambos

En esta categoría se evalúa las eficiencias de las pruebas de tipo "valor(es) extremo(s)", viz., N2, N5, N6, N8, N14 y N15. El mayor valor de $(ROC)_g$ fue de 3.81% para

la prueba múltiple tipo Grubbs N5, seguido por la prueba sencilla N15 (curtosis) con 2.74% y la prueba sencilla tipo Dixon N8 fue la menos eficiente (1.53%). En términos de datos desviados, el mayor número de discordantes fue 1592 (prueba tipo Grubbs N5) y el menor solamente 631 (prueba tipo Dixon N8). En comparación, el parámetro anterior $(REC)_g$ referente a casos o muestras estadísticas presentó valores significativamente mayores de entre 19.82% y 51.58% para pruebas tipo "extremo".

El orden del criterio de eficiencia relativa de valores desviados $(ROC)_g$ en pruebas que evaluaron datos extremos, fue el siguiente: N5 > N15 > N2 ≈ N14 > N6 > N8.

Categoría D (Figura 4): $(ROC)_g$ para todas las pruebas sencillas y múltiples

Finalmente, se presenta una comparación de todas las 15 pruebas de discordancia con las 33 variantes presentados en 20 combinaciones (Figura 4). En el histograma se observa que los valores mayores de $(ROC)_g$ corresponden a las variantes de la prueba múltiple tipo Grubbs N4. Para pruebas sencillas, el mayor valor (2.74%) de $(ROC)_g$ correspondió a la prueba de curtosis N15 mientras que el menor (1.17%) a la prueba tipo Dixon N7. Para pruebas múltiples, por su parte, el mayor valor (7.45%) de $(ROC)_g$ fue para la combinación de la prueba tipo Grubbs N4 (N4U4+N4L4) y el menor (1.99%) también para la prueba tipo Grubbs N3 (N3U4+N3L4). Estos porcentajes son significativamente menores a los representados por el parámetro anterior $(REC)_g$ referente a casos o muestras estadísticas de 21.95%-51.58% para combinaciones iguales de pruebas sencillas y de 9.37%-45.36% para pruebas múltiples.

El orden de $(ROC)_g$ para pruebas de discordancia que evaluaron a los valores más altos y más bajos fue el siguiente: N4U4+N4L4 > N4U3+N4L3 > N4U2+N4L2 > N13U2+N13L2 > N11U2+N11L2 > N5UL > N3U2+N3L2

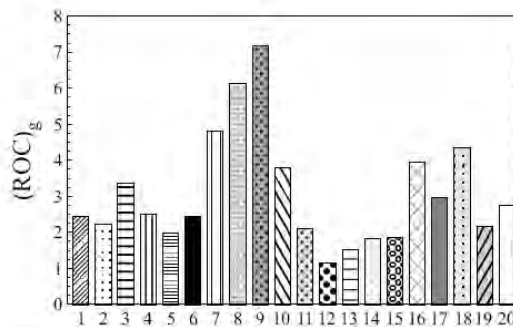


Figura 4. Criterio relativo de datos discordantes $(ROC)_g$ de todas las pruebas de discordancia sencillas y múltiples: 1-N1 (N1U+N1L), 2-N2, 3-N3UL2 (N3U2+N3L2), 4-N3UL3 (N3U3+N3L3), 5-N3UL4 (N3U4+N3L4), 6-N4UL (N4U+N4L), 7-N4UL2 (N4U2+N4L2), 8-N4UL3 (N4U3+N4L3), 9-N4UL4 (N4U4+N4L4), 10-N5UL, 11-N6UL, 12-N7 (N7U), 13-N8, 14-N9 (N9U+N9L), 15-N10 (N10U+N10L), 16-N11UL2 (N11U2+N11L2), 17-N12UL2 (N12U2+N12L2), 18-N13UL2 (N13U2+N13L2), 19-N14 y 20-N15

$> N12U2+N12L2 > N15 > N3U3+N3L3 = N4U+N4L$
 $= N1U+N1L > N2 = N14 > N6UL > N3U4+N3L4 >$
 $N10U2+N10L2 > N9 > N8 > N7.$

DISCUSIÓN

Regresión lineal y pruebas de significancia

Con el fin de evaluar la eficiencia relativa de las pruebas de discordancia mediante este nuevo procedimiento estadístico –la combinación de regresión lineal y pruebas

de significancia–, se investigó la posible dependencia del criterio de eficiencia relativa de la primera iteración (REC)₁ sobre el tamaño de la muestra estadística (n , el número de datos en casos “aplicables” o el tamaño del grupo). Para ello, los casos “aplicables” fueron separados por sus tamaños y el parámetro (REC)₂ fue calculado para cada una de las pruebas de discordancia. Los resultados de las correlaciones lineales entre el tamaño n y (REC)₂ se presentan gráficamente en las Figuras A13 y A14, mientras que una síntesis de los parámetros estadísticos de la regresión (intercepto y pendiente así como sus errores) se muestra en la Tabla 2. Debido a los valores altos del coeficiente de correlación lineal (r) con la

Tabla 2. Ajuste de regresión lineal aplicado al criterio de eficiencia relativa (REC)₂ y el tamaño (n) de los casos de la base de datos para cada prueba de discordancia agrupadas en categorías, usando el programa de computación OYNYL.

Categoría de pruebas	Pruebas de discordancia	N	a	s_a	b	s_b	r	$P_r(r; N)$
1	N1U	18	7.33	3.43	1.083	0.188	0.8219	0.000029
1	N4U	18	7.48	3.44	1.076	0.188	0.8193	0.000032
1	N7U	18	12.68	2.92	0.451	0.160	0.5769	0.012188*
1	N9U	18	13.29	3.64	0.469	0.199	0.5072	0.031690*
1	N10U	18	10.96	3.43	0.611	0.188	0.6307	0.005011
2	N1L	18	-2.46	3.78	0.815	0.207	0.7015	0.001176
2	N4L	18	-2.46	3.78	0.815	0.207	0.7015	0.001176
2	N9L	18	-0.02	3.67	0.627	0.201	0.6151	0.006584
2	N10L	18	-1.57	3.30	0.787	0.181	0.7363	0.000494
3	N2	18	4.0	4.9	1.679	0.266	0.8448	0.000010
3	N8	18	9.7	4.8	0.994	0.264	0.6850	0.001708
3	N14	18	9.97	3.40	1.042	0.186	0.8136	0.000040
3	N15	18	2.0	5.1	1.898	0.281	0.8607	0.000005
4	N1 (N1U+N1L)	18	4.9	5.1	1.898	0.280	0.8608	0.000005
4	N4 (N4U+N4L)	18	5.0	5.1	1.891	0.281	0.8595	0.000005
4	N9 (N9U+N9L)	18	13.3	5.4	1.096	0.296	0.6796	0.001921
4	N10 (N10U+N10L)	18	9.4	4.7	1.398	0.259	0.8040	0.000058
6	N3U2	18	-11.48	3.07	1.687	0.168	0.9290	0.000000
6	N4U2	18	11.94	3.08	0.958	0.169	0.8175	0.000034
6	N11U2	18	15.54	4.53	0.583	0.248	0.5065	0.031950*
6	N12U2	18	13.68	4.45	0.756	0.244	0.6126	0.006878
6	N13U2	18	11.31	3.93	0.897	0.215	0.7219	0.000719
7	N3L2	18	-9.84	3.07	1.075	0.168	0.8477	0.000009
7	N4L2	18	-4.00	4.02	0.973	0.220	0.7413	0.000431
7	N11L2	18	-1.39	3.18	0.761	0.174	0.7377	0.000476
7	N12L2	18	-2.97	3.29	0.944	0.180	0.7950	0.000081
7	N13L2	18	-5.79	2.91	1.184	0.160	0.8802	0.000001
8	N5UL	18	0.10	4.05	1.777	0.222	0.8945	0.000001
8	N6UL	18	-20.00	4.35	2.076	0.238	0.9087	0.000000
9	N3U3	18	-10.87	3.25	1.128	0.178	0.8453	0.000010
9	N4U3	18	6.93	3.44	1.171	0.189	0.8405	0.000013
10	N3L3	18	-3.74	2.37	0.439	0.130	0.6454	0.003820
10	N4L3	18	-4.4	4.6	0.954	0.253	0.6858	0.001678
11	N3U4	18	-7.28	2.06	0.671	0.113	0.8301	0.000020
11	N4U4	18	1.18	3.24	1.400	0.178	0.8916	0.000001
12	N3L4	18	-2.44	1.68	0.241	0.092	0.5472	0.018772*
12	N4L4	18	-5.03	3.82	0.905	0.209	0.7337	0.000528

N : número de pares de datos en la regresión lineal; a : valor de la ordenada en el origen o intercepto; s_a : error del intercepto; b : pendiente; s_b : error de la pendiente; r : coeficiente de regresión lineal; $P_r(r; N)$: probabilidad de no-correlación, por lo tanto, la probabilidad de una correlación lineal sería $(1 - P_r(r; N))$; * correlación lineal no-válida a nivel de confianza de 99% ($P_r(r; N) < 0.01$), pero válida a nivel de confianza de 95% ($P_r(r; N) < 0.05$). Cabe aclarar que los valores de parámetros de regresión (a , s_a , b y s_b) se reportan redondeados de acuerdo con las reglas de redondeo e indicaciones correspondientes (Bevington y Robinson, 2003; Verma, 2003).

probabilidad de no-correlación baja (Miller y Miller, 2000; Bevington y Robinson, 2003; Verma, 2005), se infiere la validez de una correlación lineal entre estos dos parámetros $-n$ y $(REC)_1$ a nivel de confianza de 99%. Esto fue válido para 33 relaciones de las pruebas de discordancia agrupadas en 11 categorías (de las 14 en la Tabla 1), con la excepción de cuatro relaciones que también fueron válidas a nivel de confianza de 95% (ver las relaciones identificadas por un asterisco en la Tabla 2).

En el presente análisis no se tomará en cuenta el intercepto de las regresiones debido a que no tiene significado para las pruebas de discordancia, las cuales son inaplicables para $n < 3$ (Verma, 2005). El intercepto de una regresión lineal se define por el valor en el eje-y para el eje-x=0, en nuestro caso para $n=0$; este valor cero de tamaño de muestra (n) carece de sentido alguno, por lo cual es difícil interpretar el significado del intercepto en el presente estudio. De cualquier manera, los valores negativos grandes de un intercepto en una categoría de pruebas donde otros valores de interceptos son positivos, como es el caso de la categoría 6 (ver el intercepto de -11.48 para N3U2 en comparación con +11.31 a +15.54 para las otras pruebas; Tabla 2), pueden causar problemas en la interpretación de las pendientes que si son importantes para evaluar las eficiencias relativas de las pruebas.

Todas las pendientes son positivas (Tabla 2), lo cual confirma que a mayor tamaño de muestra (n) corresponde una mayor eficiencia de las pruebas de discordancia $(REC)_1$ y viceversa. Esto también permite interpretar que las muestras con un tamaño menor tendrían menor número de valores desviados que las muestras grandes. Aunque una correlación más compleja (tipo polinomial) pudiera ser mejor (Figuras A13 y A14), es útil comparar estadísticamente las pendientes de las regresiones lineales para evaluar las eficiencias de las pruebas. Esto se llevó a cabo mediante las pruebas de significancia de F de Fisher y t de Student (Jensen et al., 2000; Verma, 2005), aplicadas a nivel de confianza de 99%. Para la mayoría de los pares de pruebas no se encontró diferencias significativas entre las dos varianzas.

Para cada par de pruebas en una determinada categoría (Tabla 2) se probó la hipótesis nula si una de las dos pruebas demostraba una mayor eficiencia $(REC)_1$, utilizando para ello valores críticos de t de Student para una cola a nivel de confianza de 99% (Verma, 2005, 2009b).

En la categoría 1 de las pruebas sencillas con su versión "Upper" del dato más alto (Tabla 2), las pendientes (b) para N1U y N4U fueron significativamente mayores que para N7U, N9U y N10U. Se infiere que las pruebas tipo Grubbs N1 y N4 en su versión "Upper" presentaron eficiencias significativamente mayores que las pruebas tipo Dixon N7, N9 y N10. Para las pruebas sencillas tipo "Lower" (categoría 2 en la Tabla 2), N1L y N4L tuvieron pendientes significativamente mayores que N9L, pero no con respecto a N10L. Se concluye que las pruebas tipo Grubbs N1 y N4 ("Lower") demostraron eficiencias ma-

yores que la prueba N9 tipo Dixon. De igual manera, la prueba curtosis N15 presentó eficiencia significativamente mayor que la prueba N8 tipo Dixon y la prueba de sesgo N14 (categoría 3 en la Tabla 2). Las pruebas combinadas ("Upper+Lower"; categoría 4 en la Tabla 2) tipo Grubbs N1 y N4 demostraron significativamente mayor eficiencia que las pruebas N9 y N10, ambas tipo Dixon.

Para las pruebas múltiples tipo $k=2$ para dos datos más altos ("Upper") y sin tomar en cuenta la prueba N3U2 con un valor negativo grande en su intercepto (-11.48; categoría 6 en la Tabla 2), se encontró que la prueba tipo Grubbs N4U2 fue significativamente más eficiente que las pruebas N11U2 y N12U2 tipo Dixon. Para las versiones tipo "Lower", la prueba N4L2 presentó mayor eficiencia que la prueba N11L2 (categoría 7 en la Tabla 2). Solamente dos pruebas tipo Grubbs N3 y N4 tienen versiones con $k=3$ y $k=4$. La prueba N4 presentó mayor eficiencia en relación con la prueba N3, consistente con la relación de las pendientes de regresiones lineales (ver categorías 9-12 en la Tabla 2).

Estas consideraciones son abordadas para $(REC)_2$ con el fin de confirmar las inferencias de $(REC)_1$. Las gráficas (Figuras A15 y A16) en el suplemento electrónico demuestran que los valores de $(REC)_2$ para las pruebas de curtosis N15, coeficiente de asimetría N14 y tipo Grubbs N1 y N4, son más altos que para las pruebas tipo Dixon N7-N13. Esta conclusión es totalmente consistente con el análisis de los parámetros REC y ROC (Figuras 1-4). De igual manera, se confirma que los valores de $(REC)_2$ para las pruebas tipo "Upper" son significativamente mayores a los de tipo "Lower".

Consideraciones finales

El parámetro REC, usado anteriormente por Velasco-Tapia et al. (2001) y Verma et al. (2009), así como el parámetro ROC demostraron su utilidad en la evaluación de las pruebas de discordancia. La combinación de regresiones lineales y aplicación de pruebas de significancia de F y t también fueron útiles para documentar las eficiencias relativas.

De mucha relevancia para datos geoquímicos interlaboratorios es que los resultados de $(ROC)_2$ en MRG demuestran claramente que la mayor parte de los datos desviados se localizan en el extremo de los valores altos, en comparación con los del extremo de los valores bajos. Las distribuciones, por lo tanto, de estos datos geoquímicos, debido a la "contaminación" estadística, estarían sesgadas positivamente en mayor número de los casos y negativamente en menor número de los casos. En la categoría D con todas las pruebas, las pruebas N5 y N6 que evalúan el par de datos formado por un dato a cada lado de las muestras, ocuparon lugares desde intermedios hasta los últimos en la secuencia de eficiencias y no los primeros, lo cual equivale a que la contaminación simétrica (a ambos lados

de las muestras) no es particularmente frecuente. Además, dentro de la categoría 13 de las pruebas tipo $k=2$, estas pruebas N5 y N6 demostraron relativamente eficiencias $(REC)_y$ bajas.

Todo esto tiene implicaciones importantes para los métodos robustos, como la mediana o media de Gastwirth, los cuales, a pesar de su supuesta robustez, presentarían la estimación de tendencia central afectada por la contaminación dominada por valores altos. Trabajos de simulación para evaluar el comportamiento de métodos robustos y de valores desviados se encuentran actualmente en proceso por Verma y colaboradores, mismos que proporcionarían mayores criterios acerca de los métodos más apropiados para el procesamiento de datos experimentales.

Un mayor éxito de las pruebas múltiples puede deberse en parte al "swamping effect" que afecta favorablemente a este tipo de pruebas, estando ausente en las pruebas sencillas. Las pruebas sencillas, por su parte, pueden verse disminuidas en su eficiencia por el "masking effect" que influye mayormente a esta clase de pruebas y no tanto a las múltiples. Trabajos de simulaciones computacionales para abundar más sobre este tipo de comparaciones son sumamente necesarios y se encuentran actualmente en preparación por Verma y colaboradores.

En resumen, los parámetros robustos, como la mediana o la media de Gastwirth, a pesar de su robustez, serían muy probablemente sesgados para este tipo de datos. Así mismo, la aplicación rigurosa de pruebas de discordancia antes de estimar los valores de la media y desviación estándar parece ser un requerimiento básico. Adicionalmente, se recomienda considerar el tamaño de la muestra para la selección de las pruebas de discordancia. En muestras estadísticas pequeñas, lo recomendable es aplicar pruebas de discordancia sencillas, mientras que para las muestras grandes, todas las pruebas con mayores eficiencias, tanto sencillas como múltiples, pueden ser aplicadas.

CONCLUSIONES

Se encontraron los patrones siguientes para las eficiencias de las pruebas de discordancia.

(1) La eficiencia o potencia $(REC)_y$ de las pruebas de discordancia sencillas (categoría 5) tiene el orden: $N15 > N4 = N1 > N2 = N14 > N10 = N9 > N8 > N7$. En pruebas sencillas, la que reportó el mayor valor en la eficiencia fue la de curtos N15 y la de menor eficiencia fue la N7 de tipo Dixon.

(2) El siguiente orden en la eficiencia relativa $(REC)_y$ de las pruebas múltiples (categoría 13) fue: (i) pruebas que evalúan dos datos a la vez ($k=2$): $N4U2+N4L2 > N13U2+N13L2 > N12U2+N12L2 > N11U2+N11L2 > N5UL > N3U2+N3L2 > N6UL$; (ii) pruebas que evalúan tres valores ($k=3$): $N4U3+N4L3 > N3U3+N3L3$; y (iii) pruebas que evalúan cuatro valores ($k=4$): $N4U4+N4L4 > N3U4+N3L4$. En pruebas múltiples, la prueba tipo Grubbs

N4 es la más recomendable.

(3) El orden de $(ROC)_y$ para pruebas de discordancia que evaluaron a los valores más altos (categoría A) fue el siguiente: $N4U4 > N4U3 > N4U2 > N12U2 = N13U2 > N11U2 > N3U2 > N3U3 = N1U = N4U > N3U4 > N9U = N10U > N7U$. Se afirma la importancia de la prueba N4 tipo Grubbs para su uso con datos geoquímicos.

(4) El orden de $(ROC)_y$ para pruebas de discordancia con valores bajos (categoría B) fue: $N4L4 > N4L3 > N13L2 > N4L2 > N12L2 > N11L2 > N3L2 > N3L3 > N4L = N1L > N10L2 > N9L > N3L4$.

(5) El orden del criterio de eficiencia relativa de valores desviados $(ROC)_y$ en pruebas que evaluaron datos extremos (categoría C) fue: $N5 > N15 > N2 = N14 > N6 > N8$.

(6) El orden de $(ROC)_y$ para pruebas de discordancia que evaluaron a los valores más altos y más bajos (categoría D) fue: $N4U4+N4L4 > N4U3+N4L3 > N4U2+N4L2 > N13U2+N13L2 > N11U2+N11L2 > N5UL > N3U2+N3L2 > N12U2+N12L2 > N15 > N3U3+N3L3 = N4U+N4L = N1U+N1L > N2 = N14 > N6UL > N3U4+N3L4 > N10U2+N10L2 > N9$. En esta categoría más completa, se afirma nuevamente la importancia de la prueba N4 tipo Grubbs como la de mayor eficiencia en todas sus versiones de $k=2-4$.

Finalmente, si el objetivo es contar con muestras estadísticas con distribución normal sin contaminación estadística y determinar parámetros estadísticos confiables, como debe ser para cualquier experimento en ciencias o ingenierías, es indispensable aplicar pruebas de discordancia previamente a los cálculos de la media y la desviación estándar.

AGRADECIMIENTOS

La primera autora (RGR) agradece a la Secretaría de Educación Pública por el permiso otorgado para llevar a cabo los estudios de doctorado en la Universidad Nacional Autónoma de México mientras que la segunda (LDG) lo hace al CONACYT por la beca otorgada para realizar y concluir sus estudios doctorales. Expresamos nuestro agradecimiento a Alfredo Quiroz-Ruiz por la constante ayuda computacional durante el desarrollo de esta investigación. De manera similar, agradecemos a los tres árbitros—Alfredo Aparicio, John S. Armstrong-Altrin y uno anónimo—por habernos proporcionado valiosos comentarios para mejorar nuestra presentación.

APÉNDICE. SUPLEMENTO ELECTRÓNICO

Información adicional a este artículo se encuentra disponible en el sitio web de la revista <<http://rmcg.imam.mx/>>, en la tabla de contenido de este número (suplemento electrónico 26-2-02).

REFERENCIAS

- Abbey, S., 1979. Reference materials - rock samples SY-2, SY-3, MRG-1. Energy, Mines and Resources Canada, Report 79-35, 66 pp.
- Barnett, V., Lewis, T., 1994. Outliers in Statistical Data. New York, John Wiley & Sons, Third edition, 584 pp.
- Bevington, P.R., Robinson, D.K., 2003. Data Reduction and Error Analysis for the Physical Sciences. Boston, MA, USA, McGraw-Hill, 320 pp.
- Castellón-Urbe, J., Cuevas-Arteaga, C., Trajillo-Estrada, A., 2008. Corrosion monitoring of stainless steel 304L in lithium bromide aqueous solution using transmittance optical detection technique: Optics and Lasers in Engineering, 46(6), 469-476.
- Colombo, F., Pissinatti-Mayer, E.V., Gay, H.D., Lara, R., Dorais, M.J., 2007. Barbozalita y lipscambita 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.
- Díaz-González, L., Santoyo, E., Reyes-Reyes, J., 2008. Tres nuevos geotermómetros mejorados de Na/K usando herramientas computacionales y geoprogramáticas: aplicación a la predicción de temperaturas de sistemas geotérmicos. Revista Mexicana de Ciencias Geológicas, 25(3), 465-482.
- Dixon, W.J., 1951. Ratios involving extreme values. Annals of Mathematical Statistics, 22(1), 68-78.
- Dybczynsky, R., 1980. Comparison of the effectiveness of various procedures for the rejection of outlying results and assigning consensus values in interlaboratory programs involving determination of trace elements or radioisotopes. Analytica Chimica Acta, 117(1), 53-70.
- Farré, M., Martínez, E., Hernando, M.D., Fernández-Alba, A., Fritz, J., Urrutia, E., Mihail, O., Sakka, V., Morbey, A., Albanis, T., Boto, F., Hansen, P.D., Barceló, 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.
- Flanagan, F.J., 1986. Rock reference samples, San Marcos Gabbeo, GSM-1 and Lakeview Mountain Tonalite, TLM-1. Geostandards Newsletter, 10(1), 111-119.
- Freeman, B.D., Quezada, Z., Zeni, F., Natanson, C., Dunnet, R.L., Banks, S., Qarazdo, M., Fitz, Y., Bucher, J., Eschacker, P.Q., 1997. rG-CSF reduces endotoxaemia and improves survival during *E. coli* pneumonia. Journal of Applied Physiology, 83(5), 1467-1475.
- Gabrovičá, D., Rysová, J., Filová, V., Placka, J., Čuhra, P., Kubík, M., Barová, S., 2006. Glutathione determination by glutathione-S-transferase-linked immunosorbent assay kit. Interlaboratory study. Journal of AOAC International, 89(1), 154-160.
- Gladney, E.S., 1988. 1987 compilation of elemental concentration data for USGS BIR-1, DNC-1 and W-2. Geostandards Newsletter, 12(1), 63-118.
- Gladney, E.S., Roelands, I., 1988. 1987 compilation of elemental concentration data for USGS BHVO-1, MAG-1, QLO-1, RGM-1, SCo-1, SDC-1, SGR-1, and STM-1. Geostandards Newsletter, 12(2), 253-262.
- Gladney, E.S., Roelands, I., 1990. 1988 compilation of elemental concentration data for CCRMP reference rock samples SY-2, SY-3 and MRG-1. Geostandards Newsletter, 14(3), 373-458.
- Gladney, E.S., Jones, E.A., Nickell, E.J., Roelands, 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.
- Gómez-Arias, E., Andavéde, J., Santoyo, E., Urquiza, G., 2009. Determinación de la viscosidad y su incertidumbre en fluidos de perforación usados en la construcción de pozos geotérmicos y su aplicación en el campo de Los Hornos, Puebla, México. Revista Mexicana de Ciencias Geológicas, 26(2), 516-529.
- Graybeal, D.Y., DeGortano, 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), 1725-1735.
- Grubbs, F.E., 1950. Sample criteria for testing outlying observations. Annals of Mathematical Statistics, 21, 27-58.
- Guevara, M., Verma, S.P., Velasco-Tapia, F., 2001. Evaluation of GSD intrusive rocks JG1, JG2, JG3, JG1a, and JG1b. Revista Mexicana de Ciencias Geológicas, 18(1), 74-88.
- 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.
- Hayes, K., Kinsella, A., Coffey, N., 2007. A note on the use of outlier criteria in Outsize laboratory quality control schemes. Clinical Biochemistry, 40 (3-4), 147-152.
- Jensen, J.L., Lake, L.W., Corbett, P.W.N., Goggins, D.J., 2000. Statistics for Petroleum Engineers and Geoscientists. Amsterdam, The Netherlands, Elsevier, 338 pp.
- Kasper-Zobellaga, J.J., Zolezzi-Ruiz, H., 2007. Grain size, mineralogical and geochemical studies of coastal and inland dune sands from El Vizcaíno Desert, Baja California peninsula, México. Revista Mexicana de Ciencias Geológicas, 24(3), 423-438.
- Jafarzadeh, M., Hossaini-Barzi, M., 2008. Petrography and geochemistry of Ahwaz sandstone member of Asmari Formation, Zagros, Iran: implications on provenance and tectonic setting. Revista Mexicana de Ciencias Geológicas, 25(2), 247-260.
- Li, X.J., Zhang, H., Ranish, J.A., Abersold, R., 2003. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. Analytical Chemistry, 75, 6648-6657.
- Linkosalo, T., Hakkinen, R., Hasi, P., 1996. Improving the reliability of a combined phenological time series by analyzing observation quality. Tree Physiology, 16(7), 661-664.
- Madhavaraja, J., Lee, Y.I., 2009. Geochemistry of the Dalmiapuram Formation of the Umaru Group (Early Cretaceous), Cauvery Basin, Southeastern India: Implications on Provenance and Paleoredox conditions. Revista Mexicana de Ciencias Geológicas, 26(2), 380-394.
- Marroquín-Guerra, S.G., Velasco-Tapia, F., Díaz-González, L., 2009. Evaluación estadística de Materiales de Referencia Geoprogramada del Centre de Recherches Petrographiques et Géochimiques (Francia) aplicando un esquema de detección y eliminación de valores desviados y su posible aplicación en el control de calidad de datos geoprogramados. Revista Mexicana de Ciencias Geológicas, 26(2), 530-542.
- Méndez-Ortiz, B.A., Carrillo-Chávez, A., Moutouy-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, 288 pp.
- Nagarajan, R., Madhavaraja, J., Nagendra, R., Armstrong-Altrin, J.S., Moutouy, 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., Madhavaraja, J., Nagendra, R., 2008. Carbon and oxygen isotope geochemistry of Neoproterozoic limestones of the Shahabad Formation, Bhima Basin, Karnataka, southern India. Revista Mexicana de Ciencias Geológicas, 25(2), 225-235.
- Obeidat, M.M., Almasri, F.Y., Hamawi, N.A., Massadeh, A.M., Athamneh, F.S., 2008. Assessment of nitrate contamination of karst springs, Bani Kanana, northern Jordan. Revista Mexicana de Ciencias Geológicas, 25(3), 426-437.
- Palabiyik, Y., Serpen, U., 2008. Geochemical assessment of Sivas geothermal field, Turkey. Revista Mexicana de Ciencias Geológicas, 25(3), 408-425.
- Pandarinath, K., 2009. Clay minerals in SW Indian continental shelf

- sediments cores as indicators of provenance and paleoenvironmental conditions: a statistical approach. *International Geology Review* 51(2), 145-165.
- Pearson, E.S., Chandra Sekar, C., 1936, The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, 28, 308-320.
- Ram, J.L., Thompson, B., Turner, C., Nechaustal, J.M., Sheehan, H., Bobrin, J., 2007, Identification of pets and raccoons as sources of bacterial contamination of urban storm sewers using a sequence-based bacterial source tracking method. *Water Research*, 41(16), 3605-3614.
- 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 ⁴⁰Ar-³⁹Ar age determination. *Acta Petrologica Sinica*, 22(12), 3059-3078.
- Schaber, J., Badeck, F.W., 2002, Evaluation of methods for the combination of phenological time series and outlier detection. *Tree Physiology*, 22(14), 973-982.
- Serbest, J.R., Burgess, R.M., Kuhn, A., Edwards, P.A., Cantwell, M.G., Pelletier, M.C., Berry, W.J., 2003, Precision of dialysis (perpet) sampling of cadmium in marine sediment interstitial water. *Archives of Environmental Contamination and Toxicology*, 45(3), 297-305.
- Shkhwat, L.S., Pandit, M.K., Joshi D.W., 2007, Geology and geochemistry of low grade metabasic volcanic rocks from Salunber area in the Palaeoproterozoic Aarvalls Supergroup, NW India. *Journal of Earth System Science*, 116(6), 511-524.
- Steele, T.W., Russell, B.G., Goodwin, R.G., Dornel, G., Levin, J., 1972, Preliminary report on the analysis of the six NIMROG geochemical standard samples, Randburg, South Africa. National Institute for Metallurgy Report, 1351, 74.
- Steele, T.W., Wilson, A., Goodwin, R., Ellis, P.J., Radford, A.J., 1978, Analyses of the NIMROG reference samples for minor and trace elements, Randburg, South Africa. National Institute for Metallurgy Report, 1945, 218.
- Taylor, B.J., 2000, A statistical analysis of the metallicities of nine old superclusters and moving groups. *Astronomy and Astrophysics*, 362, 563-579.
- Tietjen, G.L., Moore, R.H., 1972, Some Grubbs-type statistics for the detection of several outliers. *Technometrics*, 14(3), 583-597.
- Tores-Alvarado, I.S., Smith, A.D., Castillo-Román, J., en prensa, Sr, Nd, and Pb isotopic and geochemical constraints for the origin of magmas in Popocatepetl volcano (Central Mexico) and their relationship with adjacent volcanic fields. *International Geology Review*.
- Vargas-Rodríguez, Y.M., Gómez Vidales, V., Vázquez-Labastida, E., García-Bórquez, A., Aguilar-Sahagún, G., Marrieta-Sánchez, H., Salasón, M., 2008, Caracterización espectroscópica, química y morfológica y propiedades superficiales de una montmorillonita mexicana. *Revista Mexicana de Ciencias Geológicas*, 25(1), 135-144.
- Vattuone, M.E., Leal, P.R., Crosta, S., Berbeglia, Y., Gallegos, E., Martínez-Dopico, C., 2008, Paragénesis de zeolitas alcalinas en un afloramiento de basaltos olivínicos amigdaloides de Junín de Los Andes, Neuquén, Patagonia, Argentina. *Revista Mexicana de Ciencias Geológicas*, 25(3), 483-493.
- 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, S.P., 1997, Sixteen statistical tests for outlier detection and rejection in evaluation of international geochemical reference materials: example of microgablos PM-5. *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 GSI 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 (Geoquímica): México, D. F., Universidad Nacional Autónoma de México, 186 pp.
- Verma, S.P., 2009a, Continental rift setting for the central part of the Mexican Volcanic Belt: A statistical approach. *The Open Geology Journal*, 3, 8-29.
- Verma, S.P., 2009b, Evaluation of polynomial regression models for the Student t and Fisher F critical values, the best interpolation equations from double and triple natural logarithm transformation of degrees of freedom up to 1000, and their applications to quality control in science and engineering. *Revista Mexicana de Ciencias Geológicas*, 26(1), 79-92.
- Verma, S.P., Quiróz-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., Quiróz-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.
- Verma S.P., Quiróz-Ruiz, A., 2008, Critical values for 33 discordancy test variants for outliers in normal samples for very large sizes of 1000 to 30,000. *Revista Mexicana de Ciencias Geológicas*, 25(3), 369-381.
- Verma, S.P., Ondul-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 Sol-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, O'NYL: A new Computer Program for Ordinary, York, and New York least-squares linear regression. *WSEAS Transactions on Environment and Development*, 2(8), 997-1002.
- Verma, S.P., Quiróz-Ruiz, A., Díaz-González, L., 2008, Critical values for 33 discordancy test variants for outliers in normal samples up to sizes 1000, and applications in quality control in Earth Sciences. *Revista Mexicana de Ciencias Geológicas*, 25(1), 82-96.
- Verma, S.P., Díaz-González, L., González-Ramírez, R., 2009, Relative efficiency of single-outlier discordancy tests for processing geochemical data on reference materials and application to instrumental calibrations by a weighted least-squares linear regression model. *Geostandards and Geoanalytical Research*, 33(1), 29-49.
- Zaric, S., Niketic, S.R., 1997, The anisotropic π -effect of the nitro group in ammine-nitro cobalt(III) complexes. *Polyhedron*, 16(20), 3565-3569.

Manuscrito recibido: Enero 27, 2009

Manuscrito corregido recibido: Marzo 31, 2009

Manuscrito aceptado: Marzo 31, 2009

Suplemento electrónico 26-2-02

al artículo

Eficiencia relativa de 15 pruebas de discordancia con 33 variantes aplicadas al procesamiento de datos geoquímicos

por

Rosalinda González-Ramírez, Lorena Díaz-González y Surendra P. Verma

publicado en

Revista Mexicana de Ciencias Geológicas, 2009, 26(2), 501-515

Contenido

Definición del parámetro de evaluación de pruebas de discordancia: criterio de eficiencia relativa (<i>relative efficiency criterion</i> , REC)		Pág. 2
Resultados del criterio de eficiencia relativa REC calculado en las 14 categorías de pruebas de discordancia		Pág. 4
Figuras A1-A11	Parámetro REC de las pruebas	
Ejemplo del cálculo del criterio de eficiencia relativa (REC)		Pág. 10
Tabla A1	Casos aplicables ($n=16$) seleccionados de la base de datos para ilustrar el cálculo del parámetro REC. Esta tabla también presenta los casos que fueron exitosos en cada una de las pruebas.	
Parámetro de evaluación del porcentaje de datos desviados		Pág. 17
Ejemplos del cálculo del porcentaje de datos desviados		Pág. 18
Tabla A2	Resumen de la aplicación de todas las pruebas de discordancia a 14 casos seleccionados como ejemplos para ilustrar el cálculo del porcentaje de datos desviados.	
Resultados del criterio relativo de datos discordantes (<i>relative outlier criterion</i> , ROC)		Pág. 19
Figuras A12	Valores del parámetro ROC calculado para las pruebas que evalúan el(los) dato(s): más alto(s), más bajo(s), extremo(s) y todas las pruebas.	
Resultados de la correlación lineal del criterio de eficiencia relativa REC de la primera iteración y el tamaño de la muestra estadística		Pág. 20
Figuras A13-A16	Resultados de la correlación lineal del $(REC)_1$ o $(REC)_2$ y el tamaño de la muestra	

Parámetro de evaluación de pruebas de discordancia: criterio de eficiencia relativa (*relative efficiency criterion*, REC)

El criterio de eficiencia relativa de una prueba de discordancia para la primera iteración $(REC)_1$ se define de la siguiente manera:

$$(REC)_1 = \frac{\text{numero de casos "exitosos" en la primera iteración}}{\text{numero de casos "aplicables" en la primera iteración}} \times 100 \quad (A1)$$

En las iteraciones subsecuentes (p.ej., iteración 2, 3, 4 y etc.), el parámetro REC fue definido mediante la ecuación general:

$$(REC)_i = \frac{\text{numero de casos "exitosos" en la iteración } i}{\text{numero de casos "exitosos" en la iteración } (i-1)} \times 100 \quad (A2)$$

Por ejemplo, el REC para la segunda iteración se expresa de la siguiente forma:

$$(REC)_2 = \frac{\text{numero de casos "exitosos" en la segunda iteración}}{\text{numero de casos "exitosos" en la primera iteración}} \times 100 \quad (A3)$$

Los parámetros REC subsecuentes (a partir de la iteración 2) fueron definidos con la Ec. general (A2) porque los casos aplicables para la segunda iteración son los mismos que resultaron exitosos en la primera iteración, como se indica en la ecuación A3. En el caso de las iteraciones subsecuentes (p. ej., tercera o cuarta iteración), el denominador del parámetro REC_i es el número de casos "exitosos" en la iteración anterior $(i-1)$. Finalmente, el criterio de eficiencia relativa global $(REC)_g$ para una prueba de discordancia se define mediante la siguiente ecuación:

$$(REC)_g = \frac{\sum_{i=1}^n \text{numero de casos "exitosos" en una iteración}}{\text{numero inicial de casos "aplicables"}} \times 100 \quad (A4)$$

donde el número de iteración i varía de 1 a n (siendo n la iteración final considerada para el cálculo del parámetro REC).

El parámetro global $(REC)_g$ considera todos los casos "exitosos" (en las pruebas sencillas, este número de casos "exitosos" es igual al número de datos discordantes identificados) de todas las iteraciones realizadas por una prueba determinada (incluyendo ambas variantes: más alto "Upper" y más bajo "Lower"). En otras palabras, para calcular el parámetro $(REC)_g$ de una prueba de discordancia se requiere conocer: (i) el número total de casos "exitosos" de la prueba, considerando todas las iteraciones que ésta realizó (en sus dos versiones); y, (ii) el número de casos (o muestras

González-Ramírez et al., 2009, Revista Mexicana de Ciencias Geológicas, 26(2), 501-515

estadísticas univariadas), a los cuales fue aplicada la prueba de discordancia en la primera iteración (en este trabajo, todas las pruebas fueron aplicadas a 2220 casos).

Por tanto, si una prueba de discordancia realiza dos iteraciones y sólo en la primera detecta datos desviados, el valor de $(REC)_g$ es el mismo que $(REC)_1$: $(REC)_g = (REC)_1$. Sin embargo, si una prueba de discordancia detecta datos desviados en la primera y en la segunda iteración, se tiene que: $(REC)_g > (REC)_1$.

Por otro lado, cabe señalar que con la finalidad de considerar el parámetro REC estadísticamente válido, establecimos arbitrariamente el número 30 como el límite inferior de casos "aplicables". Esto implica que, por debajo de este límite, no se calculó ningún parámetro REC. Teóricamente, se considera que el valor esperado del parámetro REC debe aproximarse a cero, bajo la hipótesis nula (H_0) de que las muestras estadísticas univariadas fueron extraídas de una población normal, sin ningún tipo de contaminación "estadística". La hipótesis alterna (H_1), puesta a prueba está relacionada con un modelo de contaminación, según el cual un dato extremo proviene de una distribución diferente (media desplazada, mayor desviación estándar, o ambas) que el resto de los datos en una muestra estadística determinada. Barnett y Lewis (1994) es una excelente referencia para obtener más detalles sobre estas hipótesis y modelos. Por último, enfatizamos que Verma (1998) aplicó las pruebas de discordancia a muestras estadísticas univariadas procedentes de un sólo grupo de métodos analíticos a un estricto nivel de confianza del 99%. La misma metodología fue usada en la presente evaluación.

Resultados del criterio de eficiencia relativa REC calculado para las 14 categorías de pruebas de discordancia

Se presentan en las Figuras A1-A11 (además de las Figuras 1-3 en el texto principal) los resultados del cálculo del parámetro REC de todas las categorías de pruebas de discordancia. En todas las Figuras A1-A11 se emplea la misma escala en el eje "y" del parámetro REC, a fin de facilitar la comparación visual de cada una de las gráficas. Con el fin de evitar la repetición con el texto del artículo principal, no se presentarán mayores detalles sobre estas figuras.

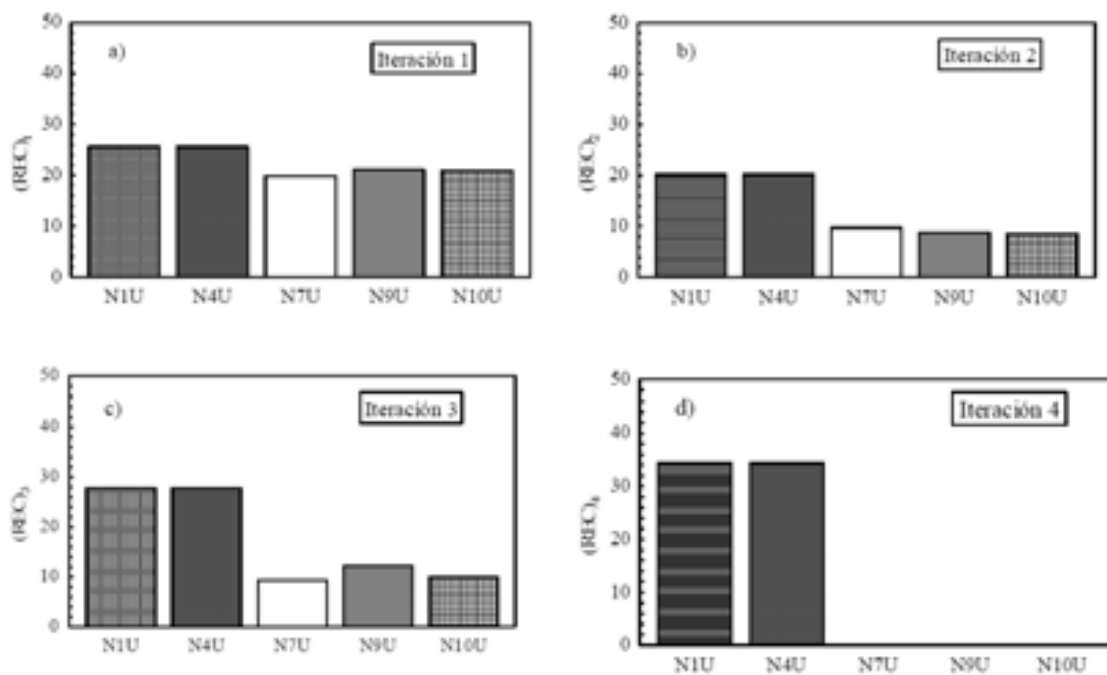


Figura A1. Criterio de eficiencia relativa (REC) de las pruebas de discordancia de la categoría 1 (ver Tabla 1 para categorización). Las cinco pruebas de discordancia sencillas que integran la categoría 1 evalúan el dato más alto ("Upper"): N1U, N4U, N7U, N9U y N10U. Cada prueba de discordancia fue aplicada en cada iteración a un determinado número de casos: (a) Iteración 1 (2220 casos con $n \geq 9$ datos para todas las pruebas); (b) Iteración 2 (568 casos para N1U; 569 casos para N4U; 440 casos para N7U; 467 casos para N9U; y, 464 casos para N10U); (c) Iteración 3 (115 casos tanto para N1U como para N4U; 43 casos para N7U; 49 casos para N9U; y, 40 casos para N10U); y (d) iteración 4 (32 casos tanto para N1U como para N4U; en esta iteración no se calculó el REC para las pruebas de discordancia N7U, N9U y N10U debido a que el número de casos aplicables fue menor a 30; restricción establecida para el cálculo del REC a fin de considerar sólo los números de casos aplicables estadísticamente representativos).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

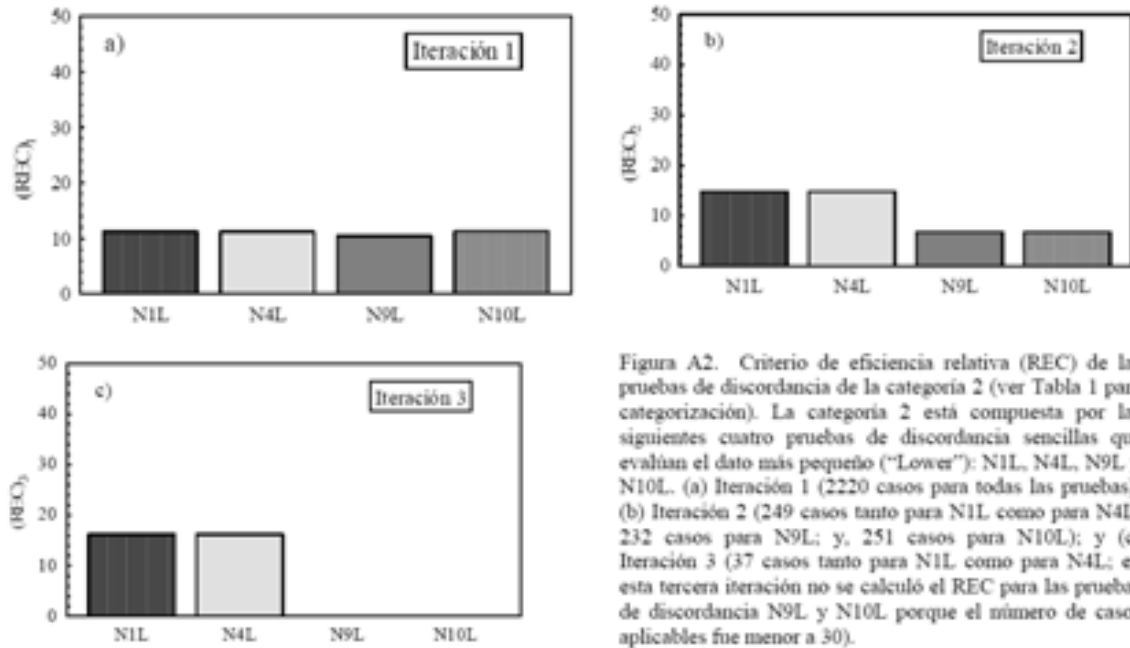


Figura A2. Criterio de eficiencia relativa (REC) de las pruebas de discordancia de la categoría 2 (ver Tabla 1 para categorización). La categoría 2 está compuesta por las siguientes cuatro pruebas de discordancia sencillas que evalúan el dato más pequeño ("Lower"): NIL, N4L, N9L y N10L. (a) Iteración 1 (2220 casos para todas las pruebas); (b) Iteración 2 (249 casos tanto para NIL como para N4L; 232 casos para N9L; y 251 casos para N10L); y (c) Iteración 3 (37 casos tanto para NIL como para N4L; en esta tercera iteración no se calculó el REC para las pruebas de discordancia N9L y N10L porque el número de casos aplicables fue menor a 30).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

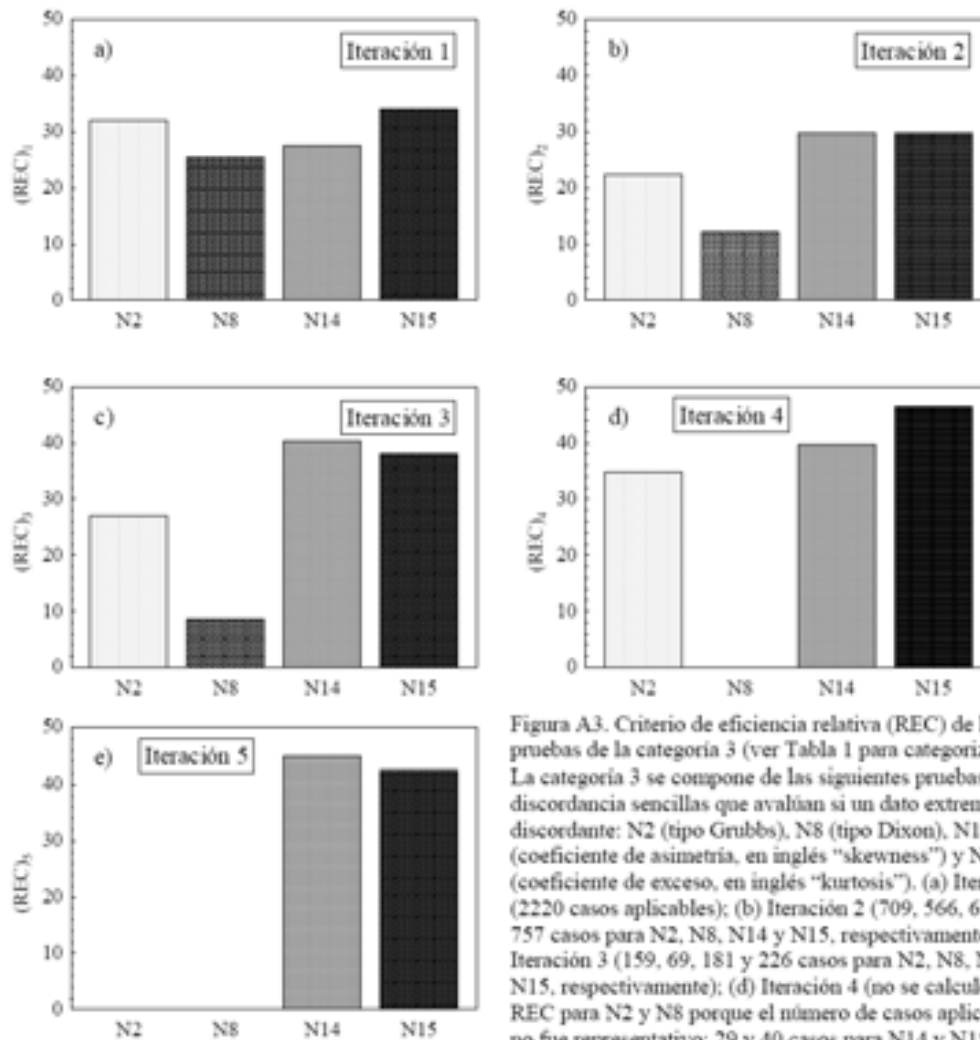


Figura A3. Criterio de eficiencia relativa (REC) de las pruebas de la categoría 3 (ver Tabla 1 para categorización). La categoría 3 se compone de las siguientes pruebas de discordancia sencillas que avalúan si un dato extremo es discordante: N2 (tipo Grubbs), N8 (tipo Dixon), N14 (coeficiente de asimetría, en inglés "skewness") y N15 (coeficiente de exceso, en inglés "kurtosis"). (a) Iteración 1 (2220 casos aplicables); (b) Iteración 2 (709, 566, 610 y 757 casos para N2, N8, N14 y N15, respectivamente); (c) Iteración 3 (159, 69, 181 y 226 casos para N2, N8, N14 y N15, respectivamente); (d) Iteración 4 (no se calculó el REC para N2 y N8 porque el número de casos aplicables no fue representativo; 29 y 40 casos para N14 y N15, respectivamente).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

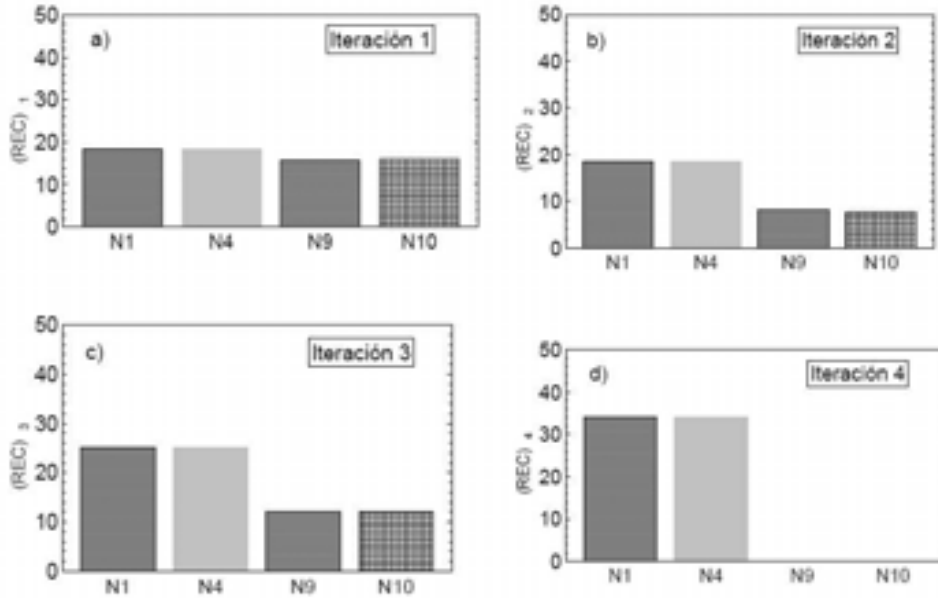


Figura A4. Criterio de eficiencia relativa (REC) de la categoría 4 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas sencillas N1, N4, N9 y N10, incluyendo sus dos versiones: más alto y más bajo. (a) Iteración 1 (2220 casos aplicables para todas las pruebas). (b) Iteración 2 (817 casos para N1; 818 casos para N4; 699 casos para N9; y, 715 casos para N10). (c) Iteración 3 (152 casos para las pruebas N1 y N4; en esta tercera iteración no se calculó el REC para N9 y N10 porque el número de casos aplicables no fue representativo).

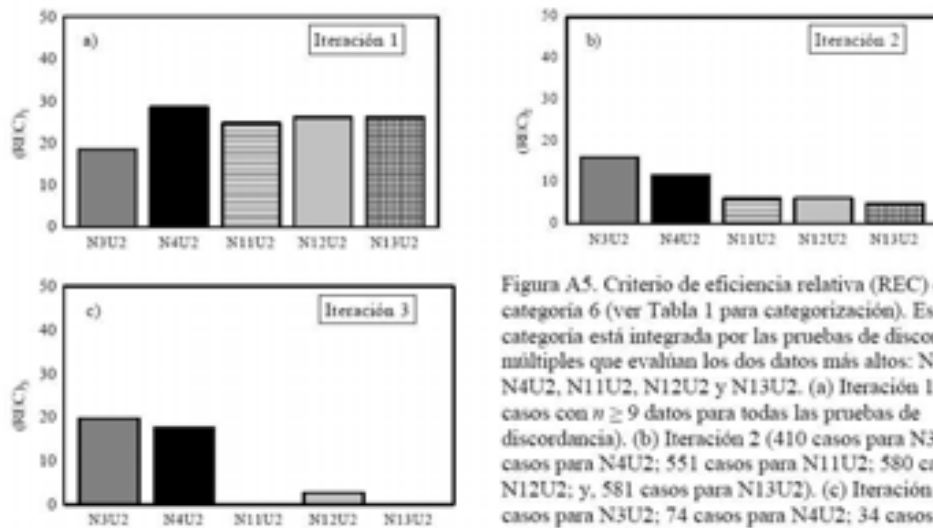


Figura A5. Criterio de eficiencia relativa (REC) de la categoría 6 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas de discordancia múltiples que evalúan los dos datos más altos: N3U2, N4U2, N11U2, N12U2 y N13U2. (a) Iteración 1 (2220 casos con $n \geq 9$ datos para todas las pruebas de discordancia). (b) Iteración 2 (410 casos para N3U2; 635 casos para N4U2; 551 casos para N11U2; 580 casos para N12U2; y, 581 casos para N13U2). (c) Iteración 3 (66 casos para N3U2; 74 casos para N4U2; 34 casos para N11U2; 36 casos para N12U2; y, 29 casos para N13U2).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

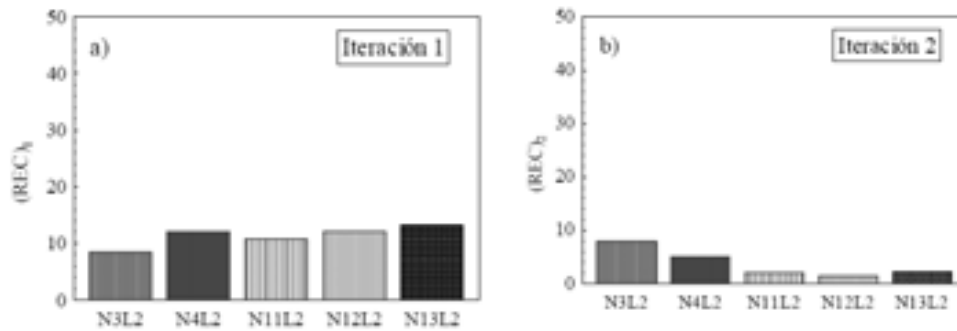


Figura A6. Criterio de eficiencia relativa (REC) de la categoría 7 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas de discordancia múltiples que evalúan los dos datos más bajos: N3L2, N4L2, N11L2, N12L2 y N13L2. (a) Iteración 1 (2220 casos de aplicación con $n \geq 9$). (b) Iteración 2 (189, 268, 237, 270 y 294 casos para N3L2, N4L2, N11L2, N12L2 y N13L2).

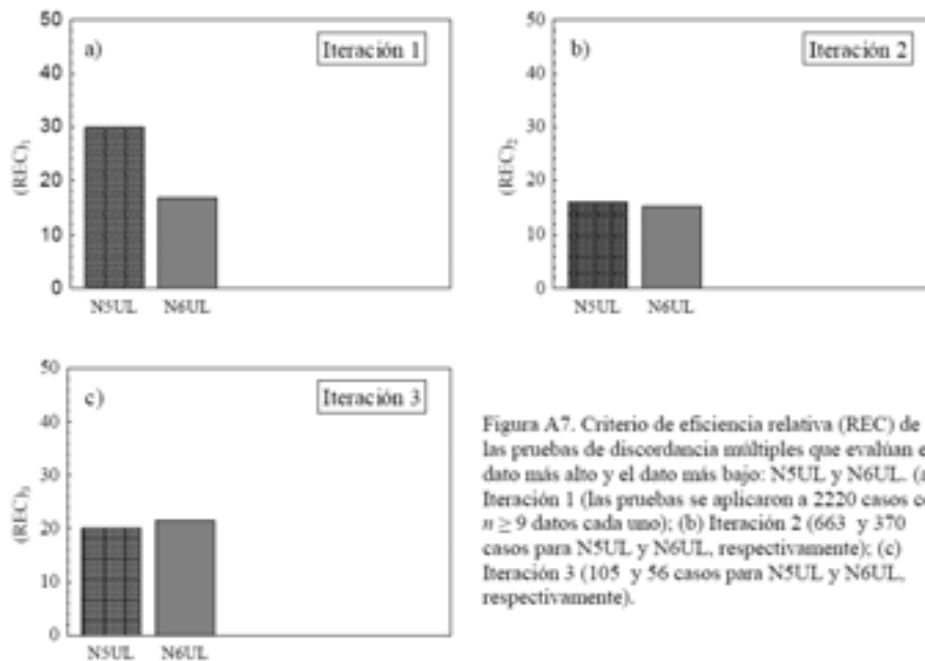


Figura A7. Criterio de eficiencia relativa (REC) de las pruebas de discordancia múltiples que evalúan el dato más alto y el dato más bajo: N5UL y N6UL. (a) Iteración 1 (las pruebas se aplicaron a 2220 casos con $n \geq 9$ datos cada uno); (b) Iteración 2 (663 y 370 casos para N5UL y N6UL, respectivamente); (c) Iteración 3 (105 y 56 casos para N5UL y N6UL, respectivamente).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

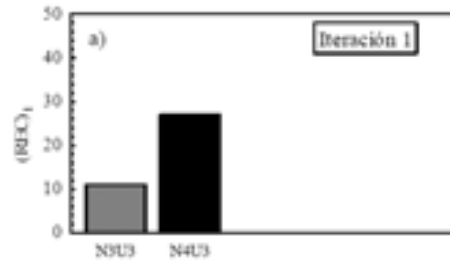


Figura A8. Criterio de eficiencia relativa (REC) de la categoría 9 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas de discordancia múltiples que evalúan los tres datos más altos: N3U3 y N4U3. (a) Iteración 1 (las pruebas N3U3 y N4U3 fueron aplicadas a 2220 casos, fueron exitosas en la detección de datos desviados en 245 y 600 casos, respectivamente).

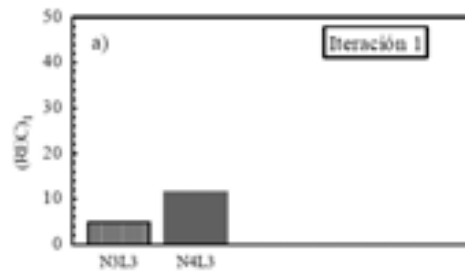


Figura A9. Criterio de eficiencia relativa (REC) de la categoría 10 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas múltiples que evalúan los tres datos más bajos: N3L3 y N4L3. Iteración 1 (las pruebas N3L3 y N4L3 fueron aplicadas a 2220 casos y detectaron datos desviados en 106 y 255 casos, respectivamente).

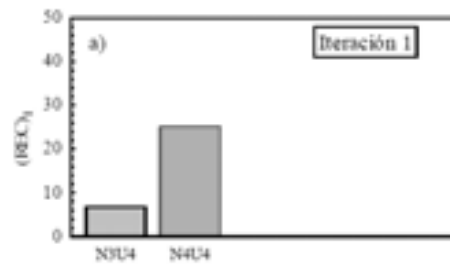


Figura A10. Criterio de eficiencia relativa (REC) de la categoría 11 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas de discordancia múltiples que evalúan los cuatro datos más altos: N3U4 y N4U4. Iteración 1 (las pruebas fueron aplicadas a 2220 casos; de los cuales, la prueba N3U4 fue exitosa en 557 casos y la prueba N4U4 en un número de casos significativamente mayor: 150).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

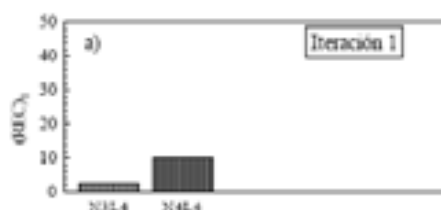


Figura A11. Criterio de eficiencia relativa (REC) de la categoría 12 (ver Tabla 1 para categorización). Esta categoría está integrada por las pruebas de discordancia múltiples que evalúan los cuatro datos más bajos: N3L4 y N4L4. (a) Iteración 1 (las pruebas N3L4 y N4L4 obtuvieron 58 y 222 casos “exitosos”, respectivamente, en 2220 casos de aplicación).

Ejemplo del cálculo del criterio de eficiencia relativa (REC)

En este apartado, se presenta un ejemplo del cálculo del parámetro REC usando un conjunto de casos, en particular, con un tamaño de muestra $n=16$ de la base de datos. Inicialmente, se identificaron en la base de datos todos los casos que tienen este tamaño de muestra, en total se encontraron 102 casos (se presentan en la Tabla A1 en las primeras tres subcolumnas identificadas con el encabezado “Casos aplicables”). Por ejemplo, algunos casos “aplicables” son: BIR-1_Co_Gr5 – métodos nucleares, DTS-1_SiO2_Gr1 – métodos clásicos, DTS-1_Sc_Gr5 – métodos nucleares, DTS-1_U_Gr5 – métodos nucleares y G-2_Ce_Gr3 – métodos de fluorescencia de rayos-X.

Posteriormente, identificamos los casos “exitosos” de cada una de las 33 variantes de pruebas de discordancia en todas sus iteraciones. La Tabla A1 contiene una columna para cada variante de prueba (esto con el propósito de ilustrar mejor la aplicación de las pruebas en estos 102 casos aplicables). En dicha columna se muestran los casos “exitosos” obtenidos en la primera iteración de una prueba determinada (p.ej., la prueba tipo Grubbs N1U fue exitosa en la detección del dato discordante más alto sólo en 22 casos de 102 casos aplicables; algunos de estos casos “exitosos” son: DTS-1_U_Gr5, G-2_Rb_Gr4, W-1_Sn_Gr4 y W-1_Gd_Gr6; por lo tanto, para calcular el (REC); de esta prueba se dividen los 22 casos “exitosos” entre los 102 casos “aplicables” y para obtener el porcentaje se multiplica el resultado por 100; 21.57%).

Asimismo, en la Tabla A1 se encuentran identificados aquellos casos en los cuales la prueba también fue exitosa en la segunda (identificados por el símbolo †) y tercera (identificados por el símbolo ‡). Por ejemplo, en la segunda iteración la prueba tipo Grubbs N1U fue exitosa únicamente en los casos W-1_Sn_Gr4 y W-1_Gd_Gr6; sin presentarse ningún caso exitoso en la tercera iteración.

Mientras tanto, la prueba del coeficiente de exceso o curtosis N15 fue exitosa no sólo en la primera y segunda iteración sino también en la tercera iteración (en los casos JB-1_V_Gr3 y JB-1_La_Gr6; identificados por el símbolo ‡).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

Tabla A1. Casos aplicables con tamaño de muestra $n = 16$ para ilustrar el cálculo del criterio de eficiencia relativa (REC).

#	Casos aplicables			Casos exitosos (nivel de confianza 99%)		
				N1U	N1L	N2
1	BIR-1_Ce_Ge5	JB-3_Ga_Ge3	JR-2_Hf_Ge5	DTS-1_U_Ge5	BIR-1_Ce_Ge5	BIR-1_Ce_Ge5
2	DTS-1_SiO2_Ge1	JB-3_Ba_Ge5	JR-2_La_Ge5	G-2_Rb_Ge4	DTS-1_Sc_Ge5	DTS-1_Sc_Ge5
3	DTS-1_Sc_Ge5	JB-3_FE2O3T_Ge5	JR-2_Na2O_Ge5	JA-1_La_Ge6	G-2_Ce_Ge3 †	DTS-1_U_Ge5
4	DTS-1_U_Ge5	JB-3_Hf_Ge5	JR-2_Nd_Ge5	JA-2_Th_Ge3	JA-1_Rb_Ge6	G-2_Ce_Ge3
5	G-2_Ce_Ge3	JB-3_Na2O_Ge5	JR-2_Rb_Ge5	JA-2_Sc_Ge5	JA-2_U_Ge5	G-2_Rb_Ge4
6	G-2_Rb_Ge4	JB-3_Yb_Ge5	MRG-1_Cu_Ge8	JB-1_La_Ge6	JB-1a_Sm_Ge5	JA-1_La_Ge6
7	G-2_Y_Ge4	IF-1_Pb_Ge3	NIM-D_MNO_Ge2	JB-1_Na2O_Ge2	JG-1a_LOI_Ge1	JA-2_Th_Ge3
8	G-2_Ho_Ge5	IF-1_Th_Ge5	NIM-D_V_Ge2	JB-3_FE2O3T_Ge5	JB-1a_K2O_Ge2	JA-2_Sc_Ge5
9	G-2_Rb_Ge6	IF-2_FE2O3T_Ge3	NIM-G_MNO_Ge3	IF-1_Pb_Ge3	JG-1a_Ce_Ge5	JA-2_U_Ge5
10	JA-1_H2OM_Ge1	JG-1a_LOI_Ge1	NIM-G_Rb_Ge3	IF-2_FE2O3T_Ge3	JG-1a_Ce_Ge6	JB-1a_Sm_Ge5
11	JA-1_Zn_Ge2	JG-1a_K2O_Ge2	NIM-L_Zn_Ge2	JG-1_V_Ge3	JG-1a_Gd_Ge6	JB-1_La_Ge6
12	JA-1_Ce_Ge5	JG-1a_Ce_Ge5	NIM-L_Sc_Ge3	JG-2_La_Ge5	JG-2_K2O_Ge2	JB-3_Na2O_Ge2
13	JA-1_Yb_Ge5	JG-1a_Ce_Ge6	NIM-N_Sr_Ge3	JP-1_Cu_Ge3	JP-1_LOI_Ge1	JB-3_FE2O3T_Ge5
14	JA-1_Dy_Ge6	JG-1a_Gd_Ge6	NIM-N_MNO_Ge8	JR-1_FE2O3T_Ge5	JR-2_Nd_Ge5	IF-1_Pb_Ge3
15	JA-1_La_Ge6	JG-1a_Yb_Ge6	NIM-P_Zn_Ge2	NIM-N_Sc_Ge3	JR-2_Rb_Ge5	IF-2_FE2O3T_Ge3
16	JA-1_La_Ge6	JG-1_K2O_Ge1	RGM-1_Sm_Ge5	NIM-P_Zn_Ge2	NIM-D_MNO_Ge2	JG-1a_LOI_Ge1
17	JA-1_Rb_Ge6	JG-1_Cr_Ge2	SY-2_AL2O3_Ge1	SY-2_TiO2_Ge3	W-1_Yb_Ge6	JG-1a_K2O_Ge2
18	JA-2_LOI_Ge1	JG-1_V_Ge3	SY-2_Rb_Ge3	SY-3_TiO2_Ge3	W-2_TiO2_Ge3	JG-1a_Ce_Ge6
19	JA-2_Th_Ge3	JG-2_K2O_Ge2	SY-2_TiO2_Ge3	W-1_Sn_Ge4 †		JG-1a_Gd_Ge6
20	JA-2_Ce_Ge5	JG-2_Ce_Ge5	SY-3_FE2O3T_Ge2	W-1_Dy_Ge6		JG-1_V_Ge3
21	JA-2_Ce_Ge5	JG-2_La_Ge5	SY-3_Sr_Ge2	W-1_Gd_Ge6 †		JG-2_K2O_Ge2
22	JA-2_Sc_Ge5	JG-2_Sc_Ge5	SY-3_TiO2_Ge3	W-2_Yb_Ge5		JG-2_La_Ge5
23	JA-2_U_Ge5	JP-1_LOI_Ge1	W-1_Na2O_Ge3			JP-1_Cu_Ge3
24	JA-3_Cu_Ge3	JP-1_Cu_Ge3	W-1_Sn_Ge4			JR-1_LOI_Ge1
25	JB-1a_Hf_Ge5	JP-1_P2O5_Ge3	W-1_Ce_Ge6			JR-1_FE2O3T_Ge5
26	JB-1a_Sm_Ge5	JR-1_LOI_Ge1	W-1_Dy_Ge6			JR-2_Nd_Ge5
27	JB-1_P2O5_Ge1	JR-1_Cr_Ge5	W-1_Gd_Ge6			JR-2_Rb_Ge5
28	JB-1_La_Ge2	JR-1_FE2O3T_Ge5	W-1_Yb_Ge6			NIM-D_MNO_Ge2
29	JB-1_V_Ge3	JR-1_La_Ge5	W-2_TiO2_Ge3			NIM-N_Sr_Ge3
30	JB-1_La_Ge6	JR-1_La_Ge5	W-2_Yb_Ge5			NIM-P_Zn_Ge2
31	JB-1_P2O5_Ge8	JR-1_Ta_Ge5				SY-2_TiO2_Ge3
32	JB-2_Cu_Ge2	JR-1_Nd_Ge6				SY-3_TiO2_Ge3
33	JB-2_MGO_Ge2	JR-1_Sm_Ge6				W-1_Sn_Ge4 †
34	JB-2_Pb_Ge6	JR-2_Cr_Ge3				W-1_Dy_Ge6
35	JB-3_H2OM_Ge1	JR-2_Th_Ge3				W-2_TiO2_Ge3
36	JB-3_Na2O_Ge2	JR-2_Cr_Ge5				W-2_Yb_Ge5

(Continúa)

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

Tabla A1. (Continuación)

#	Casos exitosos (nivel de confianza 99%)			Casos exitosos (nivel de confianza 99%)		
	N3U2	N3U3	N3U4	N3L2	N3L3	N3L4
1	G-2_Rb_Ge4	SY-3_Sr_Ge2	JA-1_Zn_Ge2	G-2_Ce_Ge3 †	JB-1_Li_Ge2	
2	JA-1_Dy_Ge6			G-2_Rb_Ge6	JB-1_P2O5_Ge8	
3	JA-1_Lu_Ge6			JB-1a_Sm_Ge5	NIM-N_MNO_Ge8	
4	JA-2_Sc_Ge5			JG-1a_Gd_Ge6		
5	JB-1_La_Ge6			JG-1a_Yb_Ge6		
6	JF-2_FE2O3T_Ge3			NIM-N_MNO_Ge8		
7	JR-1_FE2O3T_Ge5			W-2_TiO2_Ge3		
8	JR-1_Ta_Ge5					
9	JR-2_Th_Ge3					
10	NIM-N_Sr_Ge3					
11	SY-3_Sr_Ge2					
12	W-1_Sn_Ge4					
13	W-1_Ce_Ge6					
14	W-1_Gd_Ge6					
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						

(Continúa)

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

Tabla A1. (Continuación)

#	Casos exitosos (nivel de confianza 99%)			Casos exitosos (nivel de confianza 99%)		
	N4U	N4U2	N4U3	N4L4	N4L	N4L2
1	DTS-1_U_Gr5	DTS-1_U_Gr5	DTS-1_U_Gr5	DTS-1_U_Gr5	BIR-1_Co_Gr5	DTS-1_Sc_Gr5
2	G-2_Rb_Gr4	G-2_Rb_Gr4	G-2_Rb_Gr4	G-2_Rb_Gr4	DTS-1_Sc_Gr5	G-2_Ce_Gr3
3	JA-1_La_Gr6	JA-1_Dy_Gr6	JA-1_La_Gr6	JA-1_La_Gr6	G-2_Ce_Gr3	G-2_Rb_Gr6
4	JA-2_Th_Gr3	JA-1_La_Gr6	JA-2_Th_Gr3	JA-2_Th_Gr3	JA-1_Rb_Gr6	JA-2_U_Gr5
5	JA-2_Sc_Gr5	JA-2_Th_Gr3	JA-2_Sc_Gr5	JA-2_Sc_Gr5	JA-2_U_Gr5	JB-1a_Sm_Gr5
6	JB-1_La_Gr6	JA-2_Sc_Gr5	JB-1_La_Gr6	JB-3_NA2O_Gr2	JB-1a_Sm_Gr5	JG-1a_LOI_Gr1
7	JB-3_NA2O_Gr2	JB-1_La_Gr6	JB-3_NA2O_Gr2	JB-3_FE2O3T_Gr5	JG-1a_LOI_Gr1	JG-1a_K2O_Gr2
8	JB-3_FE2O3T_Gr5	JB-3_NA2O_Gr2	JB-3_FE2O3T_Gr5	JP-1_Pb_Gr3	JG-1a_K2O_Gr2	JG-1a_Ce_Gr6
9	JP-1_Pb_Gr3	JB-3_FE2O3T_Gr5	JP-1_Pb_Gr3	JP-2_FE2O3T_Gr3	JG-1a_Ce_Gr5	JG-1a_Gd_Gr6
10	JP-2_FE2O3T_Gr3	JP-1_Pb_Gr3	JP-2_FE2O3T_Gr3	JG-2_La_Gr5	JG-1a_Ce_Gr6	JG-1a_Yb_Gr6
11	JG-1_V_Gr3	JP-2_FE2O3T_Gr3	JG-2_La_Gr5	JP-1_Cu_Gr3	JG-1a_Gd_Gr6	JG-2_K2O_Gr2
12	JG-2_La_Gr5	JG-1_V_Gr3	JP-1_Cu_Gr3	JP-1_FE2O3T_Gr5	JG-2_K2O_Gr2	JR-1_LOI_Gr1
13	JP-1_Cu_Gr3	JG-2_La_Gr5	JR-1_FE2O3T_Gr5	JR-1_Ta_Gr5	JR-1_LOI_Gr1	JR-2_Nd_Gr5
14	JR-1_FE2O3T_Gr5	JP-1_Cu_Gr3	JR-1_Ta_Gr5	NIM-P_Zn_Gr2	JR-2_Nd_Gr5	JR-2_Rb_Gr5
15	NIM-N_Sr_Gr3	JR-1_FE2O3T_Gr5	NIM-N_Sr_Gr3	SY-3_Sr_Gr2	JR-2_Rb_Gr5	NIM-D_MNO_Gr2
16	NIM-P_Zn_Gr2	JR-1_Ta_Gr5	NIM-P_Zn_Gr2	W-1_Sn_Gr4	NIM-D_MNO_Gr2	NIM-N_MNO_Gr8
17	SY-2_TiO2_Gr3	NIM-N_Sr_Gr3	SY-2_TiO2_Gr3	W-1_Gd_Gr6	W-1_Yb_Gr6	W-2_TiO2_Gr3
18	SY-3_TiO2_Gr3	NIM-P_Zn_Gr2	SY-3_Sr_Gr2		W-2_TiO2_Gr3	
19	W-1_Sn_Gr4	SY-2_TiO2_Gr3	W-1_Sn_Gr4			
20	W-1_Dy_Gr6	SY-3_Sr_Gr2	W-1_Dy_Gr6			
21	W-1_Gd_Gr6	SY-3_TiO2_Gr3	W-1_Gd_Gr6			
22	W-2_Yb_Gr5	W-1_Sn_Gr4				
23		W-1_Ce_Gr6				
24		W-1_Dy_Gr6				
25		W-1_Gd_Gr6				
26		W-2_Yb_Gr5				
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						

(Continúa)

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

Tabla A1. (Continuación)

#	Casos exitosos (nivel de confianza 99%)			Casos exitosos (nivel de confianza 99%)		
	NaU3	NaU4	NSUL2	NaUL2	N7U	N8
1	DTS-1_Sc_Gr5	DTS-1_Sc_Gr5	BIR-1_Co_Gr5	BIR-1_Co_Gr5	DTS-1_U_Gr5	BIR-1_Co_Gr5
2	G-2_Ce_Gr3	G-2_Ce_Gr3	DTS-1_Sc_Gr5	JA-1_Dy_Gr6	G-2_Rb_Gr4	DTS-1_Sc_Gr5
3	JA-2_U_Gr5	JA-2_U_Gr5	DTS-1_U_Gr5	JA-1_Rb_Gr6	JA-1_Lu_Gr6	DTS-1_U_Gr5
4	JB-1a_Sm_Gr5	JB-1_P2O5_Gr8	G-2_Ce_Gr3	JB-1_V_Gr3	JA-2_Th_Gr3	G-2_Rb_Gr4
5	JB-1_P2O5_Gr8	JG-1a_LOI_Gr1	G-2_Rb_Gr4 †	JB-1_La_Gr6	JA-2_Sc_Gr5	JA-1_Lu_Gr6
6						
7	JG-1a_LOI_Gr1	JG-1a_K2O_Gr2	JA-1_Dy_Gr6	JG-1_K2O_Gr1	JB-3_NA2O_Gr2	JA-2_Th_Gr3
8	JG-1a_K2O_Gr2	JG-1a_Ce_Gr5	JA-1_Lu_Gr6	JG-1_V_Gr3	JB-3_FE2O3T_Gr5	JA-2_Sc_Gr5
9	JG-1a_Ce_Gr5	JG-1a_Ce_Gr6	JA-1_Rb_Gr6	JR-1_LOI_Gr1	IF-1_Pb_Gr3	JA-2_U_Gr5
10	JG-1a_Ce_Gr6	JG-1a_Gd_Gr6	JA-2_Th_Gr3	NIM-N_Sr_Gr3	IF-2_FE2O3T_Gr3	JB-3_NA2O_Gr2
11	JG-1a_Gd_Gr6	JG-1a_Yb_Gr6	JA-2_Sc_Gr5	W-1_Dy_Gr6	JG-2_La_Gr5	JB-3_FE2O3T_Gr5
12	JG-1a_Yb_Gr6	JR-2_Nd_Gr5	JA-2_U_Gr5		JP-1_Cu_Gr3	IF-1_Pb_Gr3
13	JG-2_K2O_Gr2	JR-2_Rb_Gr5	JB-1_V_Gr3 †		JR-1_FE2O3T_Gr5	IF-2_FE2O3T_Gr3
14	JR-1_LOI_Gr1	NIM-D_MNO_Gr2	JB-1_La_Gr6 †		NIM-P_Zn_Gr2	JG-1a_LOI_Gr1
15	JR-2_Nd_Gr5	NIM-N_MNO_Gr8	JB-3_NA2O_Gr2		SY-2_TiO2_Gr3	JG-1a_Ce_Gr6
16	JR-2_Rb_Gr5	W-1_Yb_Gr6	JB-3_FE2O3T_Gr5		SY-3_TiO2_Gr3	JG-1a_Gd_Gr6
17	NIM-D_MNO_Gr2	W-2_TiO2_Gr3	IF-1_Pb_Gr3		W-1_Dy_Gr6	JG-2_K2O_Gr2
18	NIM-N_MNO_Gr8		IF-2_FE2O3T_Gr3		W-2_Yb_Gr5	JG-2_La_Gr5
19	W-1_Yb_Gr6		JG-1a_LOI_Gr1			JP-1_Cu_Gr3
20	W-2_TiO2_Gr3		JG-1a_Ce_Gr6			JR-1_LOI_Gr1
21			JG-1a_Gd_Gr6			JR-1_FE2O3T_Gr5
22			JG-1_V_Gr3			JR-2_Nd_Gr5
23			JG-2_La_Gr5			JR-2_Rb_Gr5
24			JP-1_Cu_Gr3			NIM-D_MNO_Gr2
25			JR-1_LOI_Gr1			NIM-P_Zn_Gr2
26			JR-1_FE2O3T_Gr5			SY-2_TiO2_Gr3
27			JR-2_Nd_Gr5			SY-3_TiO2_Gr3
28			JR-2_Rb_Gr5			W-1_Dy_Gr6
29			NIM-D_MNO_Gr2			W-2_TiO2_Gr3
30			NIM-N_Sr_Gr3			W-2_Yb_Gr5
31			NIM-P_Zn_Gr2			
32			SY-2_TiO2_Gr3			
33			SY-3_TiO2_Gr3			
34			W-1_Dy_Gr6			
35			W-2_TiO2_Gr3			
36			W-2_Yb_Gr5			

(Continúa)

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

Tabla A1. (Continuación)

#	Casos exitosos (nivel de confianza 99%)			Casos exitosos (nivel de confianza 99%)		
	N9U	N9L	N10U	N10L	N11U2	N11L2
1	DTS-1_U_Gr5	BIR-1_Co_Gr5	DTS-1_U_Gr5	BIR-1_Co_Gr5	DTS-1_U_Gr5	BIR-1_Co_Gr5
2	G-2_Rb_Gr4	DTS-1_Sc_Gr5	G-2_Rb_Gr4 †	DTS-1_Sc_Gr5	G-2_Rb_Gr4	DTS-1_Sc_Gr5
3	JA-1_La_Gr6	G-2_Ce_Gr3 †	JA-1_La_Gr6	JA-1_Dy_Gr6	JA-1_La_Gr6	G-2_Ce_Gr3
4	JA-2_Th_Gr3	JA-1_Rb_Gr6	JA-2_Th_Gr3	JA-1_Rb_Gr6	JA-2_Th_Gr3	G-2_Rb_Gr6
5	JA-2_Sc_Gr5	JA-2_U_Gr5	JA-2_Sc_Gr5	JA-2_U_Gr5	JA-2_Sc_Gr5	JA-2_U_Gr5
6	JB-1_La_Gr6 †	JG-1a_LOI_Gr1	JB-1_La_Gr6 †	JB-1_V_Gr3	JB-1_La_Gr6	JB-1a_Sm_Gr5
7	JB-3_NA2O_Gr2	JG-1a_Ce_Gr6	JB-3_NA2O_Gr2	JB-1_La_Gr6	JB-3_NA2O_Gr2	JG-1a_LOI_Gr1
8	JB-3_FE2O3T_Gr5	JG-1a_Gd_Gr6	JB-3_FE2O3T_Gr5	JG-1a_LOI_Gr1	JB-3_FE2O3T_Gr5	JG-1a_K2O_Gr2
9	JF-1_Pb_Gr3	JG-2_K2O_Gr2	JF-1_Pb_Gr3	JG-1a_Ce_Gr6	JF-1_Pb_Gr3	JG-1a_Ce_Gr6
10	JF-2_FE2O3T_Gr3	JR-1_LOI_Gr1	JF-2_FE2O3T_Gr3	JG-1a_Gd_Gr6	JF-2_FE2O3T_Gr3	JG-1a_Gd_Gr6
11	JG-1_V_Gr3	JR-2_Nd_Gr5	JG-1_V_Gr3	JG-2_K2O_Gr2	JG-1_V_Gr3	JG-2_K2O_Gr2
12	JG-2_La_Gr5	JR-2_Rb_Gr5	JG-2_La_Gr5	JR-1_LOI_Gr1	JG-2_La_Gr5	JR-1_LOI_Gr1
13	JP-1_Cu_Gr3	NIM-D_MNO_Gr2	JP-1_Cu_Gr3	JR-2_Nd_Gr5	P-1_Cu_Gr3	JR-2_Nd_Gr5
14	JR-1_FE2O3T_Gr5	W-2_TiO2_Gr3 †	JR-1_FE2O3T_Gr5	JR-2_Rb_Gr5	JR-1_FE2O3T_Gr5	JR-2_Rb_Gr5
15	NIM-P_Zn_Gr2		NIM-N_Sr_Gr3	NIM-D_MNO_Gr2	JR-1_Ta_Gr5	NIM-D_MNO_Gr2
16	SY-2_TiO2_Gr3		NIM-P_Zn_Gr2	W-2_TiO2_Gr3	NIM-N_Sr_Gr3	W-2_TiO2_Gr3
17	SY-3_TiO2_Gr3		SY-2_TiO2_Gr3		NIM-P_Zn_Gr2	
18	W-1_Dy_Gr6		SY-3_TiO2_Gr3		SY-2_TiO2_Gr3	
19	W-2_Yb_Gr5		W-1_Dy_Gr6		SY-3_TiO2_Gr3	
20			W-2_Yb_Gr5		W-1_Sn_Gr4	
21					W-1_Ce_Gr6	
22					W-1_Dy_Gr6	
23					W-1_Gd_Gr6	
24					W-2_Yb_Gr5	
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						

(Continúa)

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

Tabla A1. (Continuación)

#	Casos exitosos (nivel de confianza 99%)			Casos exitosos (nivel de confianza 99%)		
	N12U2	N12L2	N12U2	N13L3	N14	N15
1	DTS-1_U_Ge5	BIR-1_Co_Ge5	DTS-1_U_Ge5	BIR-1_Co_Ge5	DTS-1_U_Ge5	BIR-1_Co_Ge5
2	G-2_Rb_Ge4	DTS-1_Sc_Ge5	G-2_Rb_Ge4	DTS-1_Sc_Ge5	G-2_Rb_Ge4	DTS-1_Sc_Ge5
3	JA-1_Dy_Ge6	G-2_Ce_Ge3	G-2_Rb_Ge6	G-2_Ce_Ge3	JA-1_La_Ge6	DTS-1_U_Ge5
4	JA-1_La_Ge6	JA-1_Rb_Ge6	JA-1_Dy_Ge6	G-2_Rb_Ge6	JA-2_Th_Ge3	G-2_Ce_Ge3
5	JA-2_Th_Ge3	JA-2_U_Ge5	JA-1_La_Ge6	JA-1_Rb_Ge6	JA-2_Sc_Ge5	G-2_Rb_Ge4
6	JA-2_Sc_Ge5	JB-1a_Sm_Ge5	JA-2_Th_Ge3	JA-2_U_Ge5	JB-1_La_Ge6	JA-1_La_Ge6
7	JB-1_V_Ge3	JG-1a_LOI_Ge1	JA-2_Sc_Ge5	JB-1a_Sm_Ge5	JB-3_NA2O_Ge2	JA-1_Rb_Ge6
8	JB-1_La_Ge6	JG-1a_Ce_Ge6	JB-1_V_Ge3	JB-1_V_Ge3	JB-3_FE2O3T_Ge5	JA-2_Th_Ge3
9	JB-3_NA2O_Ge2	JG-1a_Gd_Ge6	JB-1_La_Ge6	JB-1_La_Ge6	JP-1_Pb_Ge3	JA-2_Sc_Ge5
10	JB-3_FE2O3T_Ge5	JG-2_K2O_Ge2	JB-3_NA2O_Ge2	JG-1a_LOI_Ge1	JP-2_FE2O3T_Ge3	JA-2_U_Ge5
11	JP-1_Pb_Ge3	JR-1_LOI_Ge1	JB-3_FE2O3T_Ge5	JG-1a_Ce_Ge6	JG-1_V_Ge3	JB-1a_Sm_Ge5
12	JP-2_FE2O3T_Ge3	JR-2_Nd_Ge5	JP-1_Pb_Ge3	JG-1a_Gd_Ge6	JG-2_La_Ge5	JB-1_V_Ge3
13	JG-1_V_Ge3	JR-2_Rb_Ge5	JP-2_FE2O3T_Ge3	JR-1_LOI_Ge1	JP-1_Cu_Ge3	JB-1_La_Ge6
14	JG-2_La_Ge5	NIM-D_MNO_Ge2	JG-1_V_Ge3	JR-2_Nd_Ge5	JR-1_FE2O3T_Ge5	JB-3_NA2O_Ge2
15	JP-1_Cu_Ge3	W-2_TiO2_Ge3	JG-2_La_Ge5	JR-2_Rb_Ge5	JR-1_Ta_Ge5 †	JB-3_FE2O3T_Ge5
16	JR-1_FE2O3T_Ge5		JP-1_Cu_Ge3	NIM-D_MNO_Ge2	NIM-N_Sr_Ge3	JP-1_Pb_Ge3
17	JR-1_Ta_Ge5		JR-1_FE2O3T_Ge5	NIM-P_Zn_Ge2	NIM-P_Zn_Ge2	JP-2_FE2O3T_Ge3
18	NIM-N_Sr_Ge3		JR-1_Ta_Ge5	W-2_TiO2_Ge3	SY-2_TiO2_Ge3	JG-1a_LOI_Ge1
19	NIM-P_Zn_Ge2		NIM-N_Sr_Ge3 †		SY-3_Sr_Ge2	JG-1a_Ce_Ge6
20	SY-2_TiO2_Ge3		NIM-P_Zn_Ge2		SY-3_TiO2_Ge3	JG-1a_Gd_Ge6
21	SY-3_TiO2_Ge3		SY-2_TiO2_Ge3		W-1_Sn_Ge4 †	JG-1_V_Ge3
22	W-1_Sn_Ge4		SY-3_TiO2_Ge3		W-1_Dy_Ge6	JG-2_K2O_Ge2
23	W-1_Ce_Ge6		W-1_Sn_Ge4		W-1_Gd_Ge6 †	JG-2_La_Ge5
24	W-1_Dy_Ge6		W-1_Ce_Ge6		W-2_Yb_Ge5	JP-1_Cu_Ge3
25	W-1_Gd_Ge6		W-1_Dy_Ge6			JR-1_LOI_Ge1
26	W-2_Yb_Ge5		W-1_Gd_Ge6			JR-1_FE2O3T_Ge5
27			W-2_Yb_Ge5			JR-2_Nd_Ge5
28						JR-2_Rb_Ge5
29						NIM-D_MNO_Ge2
30						NIM-N_Sr_Ge3
31						NIM-P_Zn_Ge2
32						SY-2_TiO2_Ge3
33						SY-3_TiO2_Ge3
34						W-1_Sn_Ge4 †
35						W-1_Dy_Ge6
36						W-1_Gd_Ge6 †
37						W-2_TiO2_Ge3
38						W-2_Yb_Ge5

Abreviaturas para grupos de métodos analíticos (Ge1 a Ge8): Ge1 – métodos clásicos; Ge2 – métodos de absorción atómica; Ge3 – métodos de fluorescencia de rayos-X; Ge4 – método de espectrometría de emisión; Ge5 – métodos nucleares; Ge6 – métodos de espectrometría de masas; Ge7 – métodos de cromatografía; Ge8 – miscelánea de métodos. Mayores detalles ser consultados en Velasco-Tapia et al. (2001). Las pruebas de discordancia para algunos casos "exitosos" también detectaron datos desviados en una segunda (identificadas por el símbolo †) y tercera (identificadas por el símbolo ‡).

González-Ramírez *et al.*, 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), XXX-XXX

Parámetro de evaluación del porcentaje de datos desviados o discordantes (%Ot)

Es conveniente destacar que el parámetro REC no es sinónimo de %Ot (%otd en Verma 1997 o %O_t en Velasco-Tapia *et al.* 2001). El parámetro (%Ot)_{Gr} para un determinado grupo de métodos analíticos de un material de referencia se expresa de la siguiente manera:

$$(\%Ot)_{Gr} = \frac{\text{número total de datos discordantes en un caso "aplicable"}}{\text{número total de datos iniciales en este caso "aplicable"}} \times 100 \quad (\text{A5})$$

El parámetro %Ot para un elemento de un material de referencia en una base de datos completa puede ser similarmente definido como:

$$\%Ot = \frac{\text{número total de datos discordantes en un caso "combinado"}}{\text{número total de datos iniciales en este caso "combinado"}} \times 100 \quad (\text{A6})$$

Un caso "combinado" se refiere a todos los datos analíticos de un elemento en un material de referencia, obtenidos de todos los métodos analíticos.

La mayor distinción entre el REC y el parámetro %Ot es que el primero es referido a todos los casos "aplicables" mientras que en el segundo a un sólo caso "aplicable" en particular, tanto para una sólo grupo de métodos (aquí referido como (%Ot)_{Gr}) o para un caso "combinado" de todos los grupos de métodos (simplemente llamado %Ot). El parámetro REC puede ser calculado únicamente si se tiene un número representativo de casos aplicables (Ecs. A1-A4) mientras que el parámetro %Ot para cualquier caso "aplicable" o "combinado" (Ecs. A5-A6) generalmente se considera significativo.

Ejemplos del cálculo del porcentaje de datos desviados o discordantes $(\%Ot)_{Gr}$

A continuación, se presentan catorce muestras con tamaño $n=16$ seleccionadas de la Tabla A1 para ilustrar el cálculo del porcentaje de datos desviados $(\%Ot)_{Gr}$. En la Tabla A2 se muestra un resumen de la aplicación de todas las pruebas de discordancia a estos casos de ejemplo.

Tabla A2. Resumen de la aplicación de todas las pruebas de discordancia a los catorce casos seleccionados como ejemplos para ilustrar el cálculo del porcentaje de datos desviados.

#	Case	$(n_0)_{Gr}$	$(x_0)_{Gr}$	$(y_0)_{Gr}$	$(Ot)_{Gr}$	$(\%Ot)_{Gr}$	(n_0)	(x_0)	(y_0)
1	BIR-1_Co_Gr5	16	51.906	5.819	3	18.8	13	52.731	2.830
2	DTS-1_Sc_Gr5	16	3.361	0.688	5	31.3	11	3.571	0.208
3	DTS-1_U_Gr5	16	0.004	0.002	5	31.3	11	0.003	0.000
4	G-2_Ce_Gr3	16	164.313	20.630	5	31.3	11	171.364	4.744
5	G-2_Rb_Gr4	16	200.500	91.813	6	37.5	10	176.000	11.662
6	JA-1_Lu_Gr6	16	0.466	0.064	5	31.3	11	0.442	0.019
7	MRG-1_Cu_Gr8	16	141.813	23.224	0	0	16	141.81	23.22
8	NIM-L_Sr_Gr3	16	4374.813	586.271	0	0	16	4374.8	586.3
9	NIM-L_Zn_Gr2	16	384.500	139.348	0	0	16	384.5	139.3
10	JA-2_Co_Gr5	16	29.819	1.365	0	0	16	29.819	1.365
11	JA-2_Sc_Gr5	16	19.745	2.497	5	31.3	11	18.875	0.843
12	SY-2_AL2O3_Gr1	16	11.951	0.221	0	0	16	11.951	0.221
13	W-1_Ce_Gr6	16	22.556	4.240	2	12.5	14	21.279	2.517
14	W-2_Yb_Gr5	16	2.103	0.149	3	19	13	2.0754	0.077

Como se puede observar, en la séptima columna de la Tabla A2 se presenta el porcentaje de datos desviados $(\%Ot)_{Gr}$ calculado a partir del número de datos discordantes identificados $(Ot)_{Gr}$ por las pruebas de discordancia.

A continuación se presenta información más detallada sobre la detección de los datos discordantes en cada caso: BIR-1_Co_Gr5 $((\%Ot)_{Gr} = 18.8$; los dos valores más pequeños y el valor más grande fueron identificados como datos discordantes por las pruebas: N1L, N2, N4L, N5UL, N6UL, N8, N9L, N10L, N11L2, N12L2, N13L2 y N15); DTS-1_Sc_Gr5 $((\%Ot)_{Gr} = 31.3$; los cuatro valores más pequeños y el valor más grande se identificaron como datos desviados por las pruebas: N1L, N2, N4L, N4L2, N4L3, N4L4, N5UL, N8, N9L, N10L, N11L2, N12L2, N13L2 y N15); DTS-1_U_Gr5 $((\%Ot)_{Gr} = 31.3$; cinco datos desviados: el dato más pequeño y los cuatro valores más grandes; N1U, N2, N4U, N4U2, N4U3, N4U4, N5UL, N7U, N8, N9U, N10U, N11U2, N12U2, N13U2, N14 y N15); G-2_Ce_Gr3 $((\%Ot)_{Gr} = 31.3$; cinco datos desviados: los cuatro valores más pequeños y el valor más grande; N1L, N2, N3L2, N4L, N4L2, N4L3, N4L4, N5UL, N9L, N11L2, N12L2, N13L2 y N15); G-2_Rb_Gr4 $((\%Ot)_{Gr} = 37.5$; seis datos desviados: los cuatro valores más pequeños y los dos valores más grande; N1U, N2, N3U2, N4U, N4U2, N4U3, N4U4, N5UL, N7U, N8, N9U, N10U, N11U2, N12U2, N13U2, N14 y N15) JA-1_Lu_Gr6 $((\%Ot)_{Gr} = 31.3$; cinco datos desviados: el valor más pequeño y los cuatro valores más grandes; N1U, N2, N3U2, N4U, N4U2, N4U3, N4U4, N5UL, N7U, N8, N9U, N10U, N11U2, N12U2, N13U2, N14 y N15); JA-2_Sc_Gr5 $((\%Ot)_{Gr} = 31.3$; cinco datos desviados: el valor más pequeño y los cuatro valores más grandes; N1U, N2, N3U2, N4U, N4U2, N4U3, N4U4, N5UL, N7U, N8, N9U, N10U, N11U2, N12U2, N13U2, N14 y N15); W-1_Ce_Gr6 $((\%Ot)_{Gr} = 12.5$; los dos datos más grandes fueron identificados como datos

González-Ramírez, et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 301-315

discordantes por las pruebas: N3U2, N4U2, N11U2, N12U2 y N13U2); W-2_Yb_Gr5 ((%Ot)_{Gr} = 18.8; tres datos desviados: el dato más pequeño y los dos valores más grandes: N1U, N2, N4U, N4U2, N5UL, N7U, N8, N9U, N10U, N11U2, N12U2, N13U2, N14 y N15); y finalmente, en los casos: JA-2_Co_Gr5, MRG-1_Cu_Gr8, NIM-L_Sr_Gr3 y NIM-L_Zn_Gr2 no se detectaron datos desviados y por tanto el porcentaje de datos desviados calculado es igual a cero ((%Ot)_{Gr} = 0).

Resultados del criterio relativo de datos discordantes (ROC)

El parámetro ROC permite estimar el número de datos que una prueba de discordancia detecta y rechaza en una muestra estadística, en función de datos totales. En la Figura A15 se presentan los valores del parámetro ROC calculados para las pruebas que evalúan los datos: (a) más altos; (b) más bajos; y, (c) extremos. El parámetro ROC se obtiene de dividir el número de datos discordantes identificados por una determinada prueba entre el número total de datos (41,821 datos) que contienen los 2220 casos a los cuales fue aplicada dicha prueba.

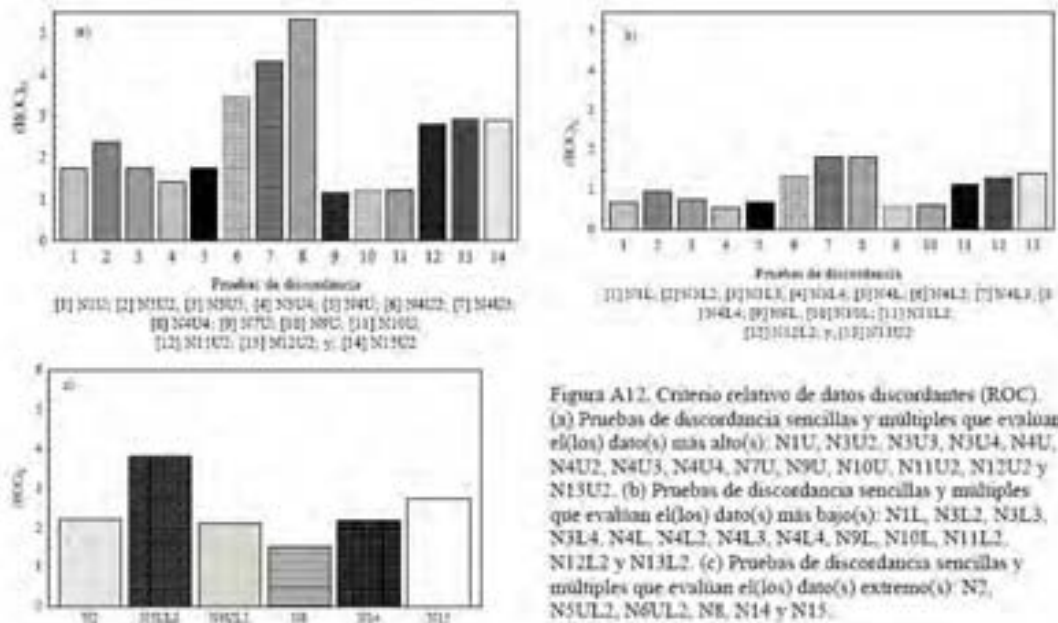


Figura A12. Criterio relativo de datos discordantes (ROC). (a) Pruebas de discordancia sencillas y múltiples que evalúan el(los) dato(s) más alto(s): N1U, N3U2, N3U3, N3U4, N4U, N4U2, N4U3, N4U4, N7U, N9U, N10U, N11U2, N12U2 y N13U2. (b) Pruebas de discordancia sencillas y múltiples que evalúan el(los) dato(s) más bajo(s): N1L, N3L2, N3L3, N3L4, N4L, N4L2, N4L3, N4L4, N9L, N10L, N11L2, N12L2 y N13L2. (c) Pruebas de discordancia sencillas y múltiples que evalúan el(los) dato(s) extremo(s): N2, N5UL, N6UL, N8, N14 y N15.

Resultados de la correlación lineal entre el criterio de eficiencia relativa de la primera iteración $(REC)_1$ o de todas las iteraciones $(REC)_2$ y el tamaño de la muestra estadística

En las Figuras A13 y A14, se presentan las correlaciones lineales entre el $(REC)_1$ y el tamaño de las muestras estadísticas formados a partir de la base de datos. De manera similar, los resultados para $(REC)_2$ se dan a conocer en las Figuras A15 y A16.

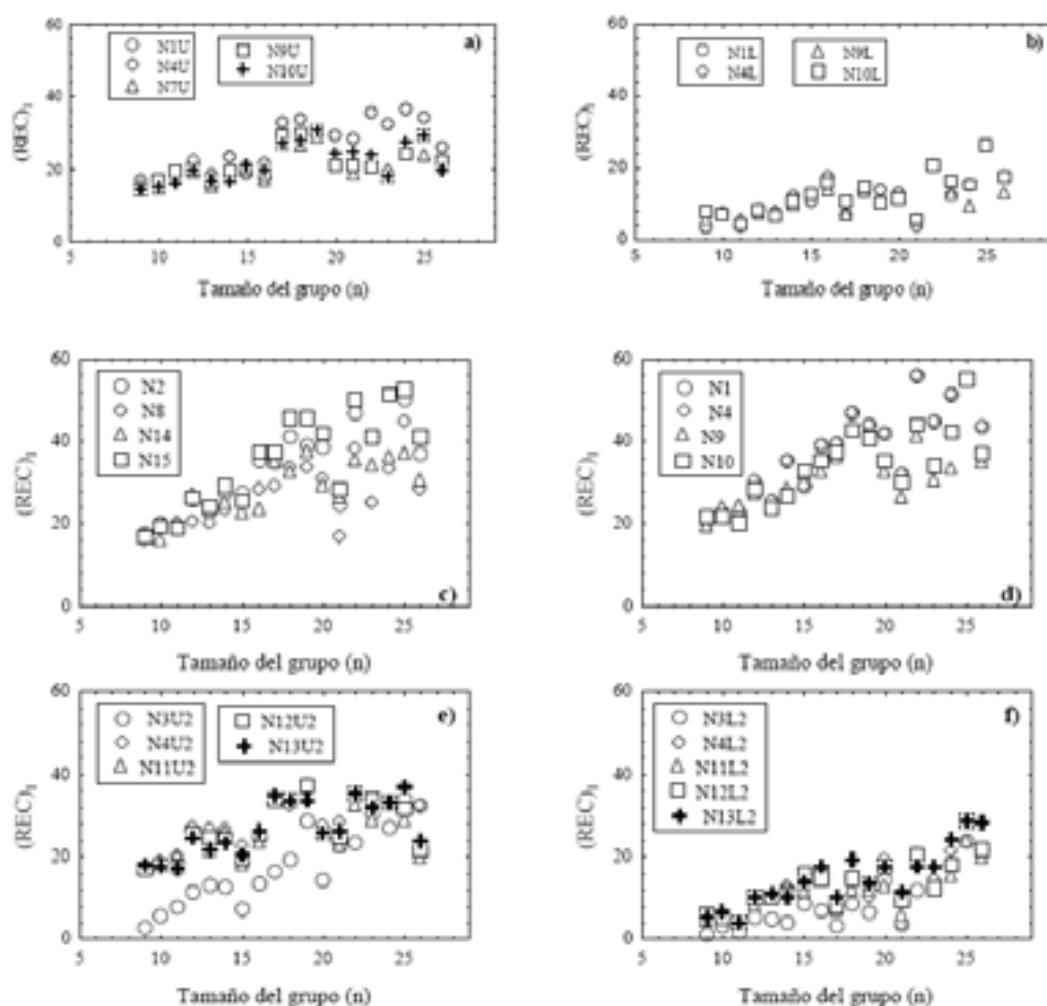


Figura A13. Criterio de eficiencia relativa de la primera iteración $(REC)_1$ versus el tamaño (n) del grupo o caso. Se presentan las pruebas de discordancia agrupadas por categoría con el propósito de facilitar la comparación visual del parámetro $(REC)_1$: (a) categoría 1 (N1U, N4U, N7U, N9U y N10U); (b) categoría 2 (N1L, N4L, N9L, N10L); (c) categoría 3 (N2, N8, N14 y N15); (d) categoría 4 (N1, N4, N9 y N10; con ambas versiones: más alto y más bajo); (e) categoría 6 (N3U2, N4U2, N11U2, N12U2 y N13U2); y, (f) categoría 7 (N3L2, N4L2, N11L2, N12L2 y N13L2).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

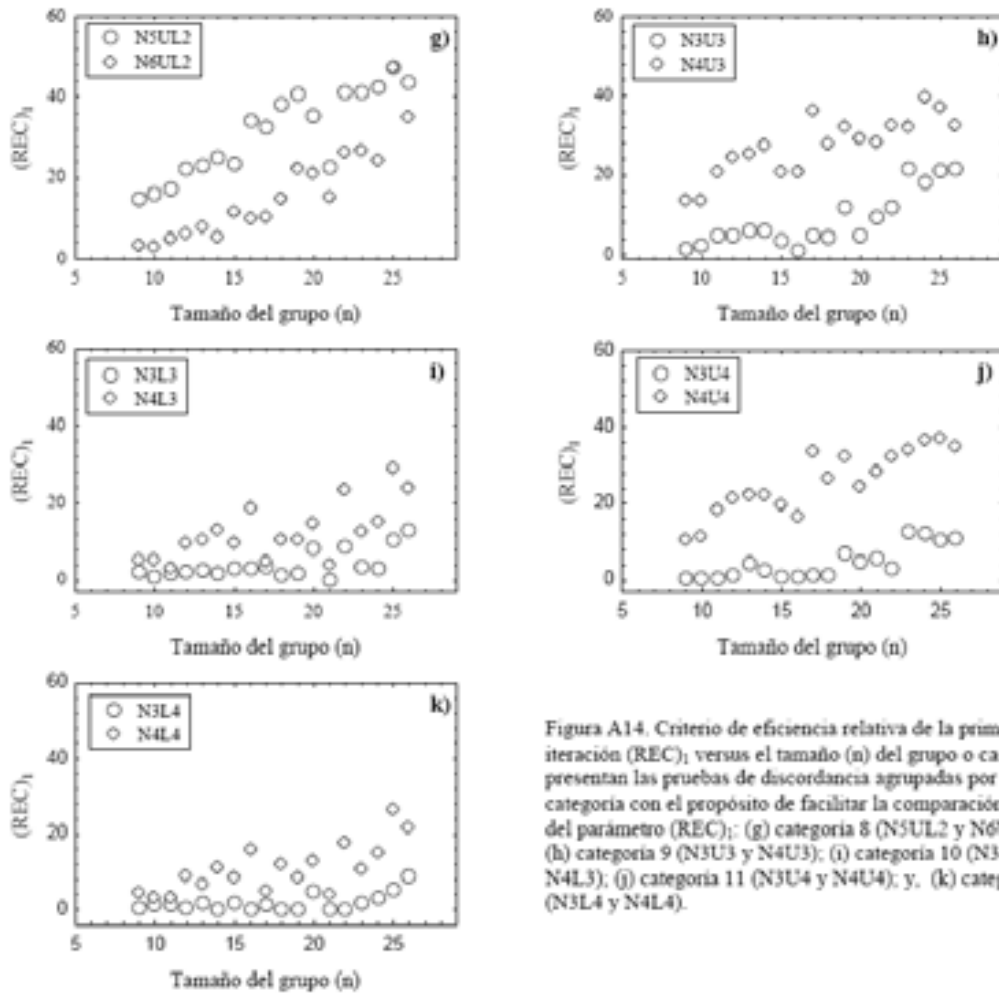


Figura A14. Criterio de eficiencia relativa de la primera iteración $(REC)_1$ versus el tamaño (n) del grupo o caso. Se presentan las pruebas de discordancia agrupadas por categoría con el propósito de facilitar la comparación visual del parámetro $(REC)_1$: (g) categoría 8 (N5UL2 y N6UL2); (h) categoría 9 (N3U3 y N4U3); (i) categoría 10 (N3L3 y N4L3); (j) categoría 11 (N3U4 y N4U4); y. (k) categoría 12 (N3L4 y N4L4).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

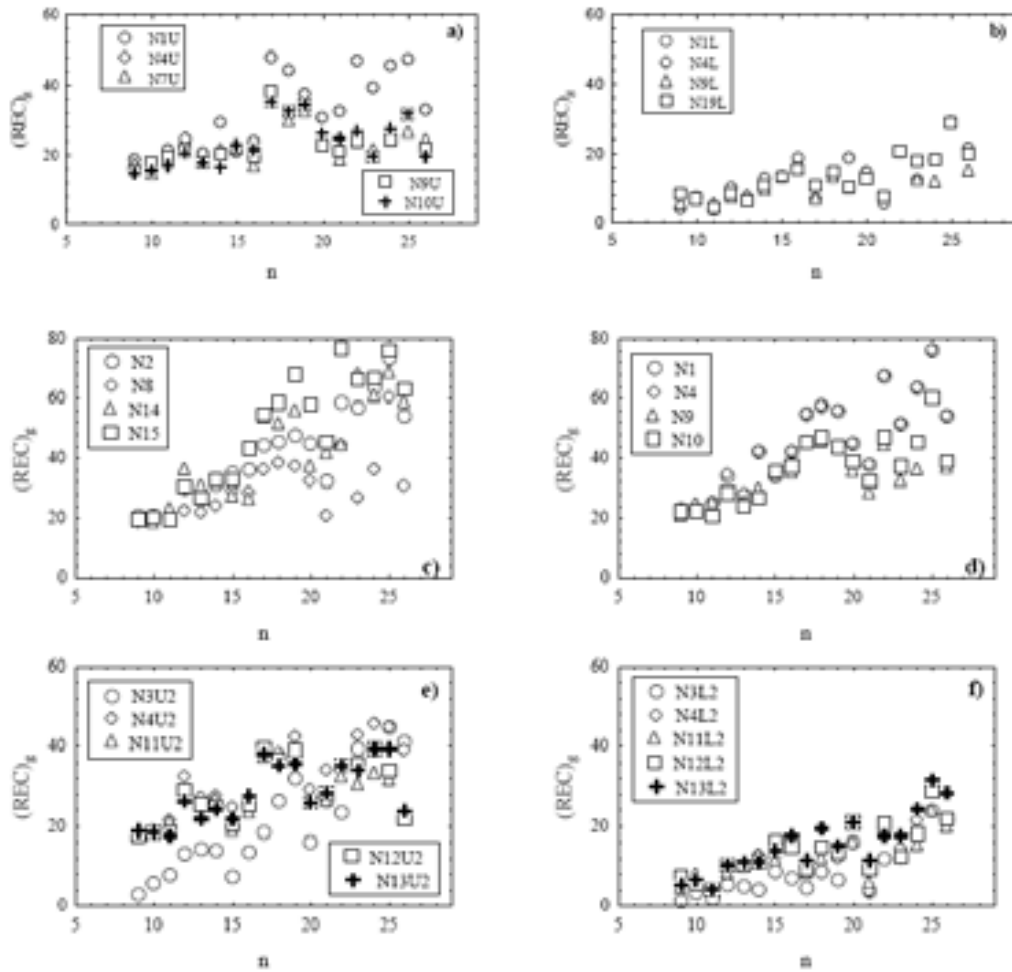


Figura A15. Criterio de eficiencia relativa global $(REC)_g$ versus el tamaño (n) del grupo o caso. Se presentan las pruebas de discordancia agrupadas por categoría con el propósito de facilitar la comparación visual del parámetro $(REC)_g$: (a) categoría 1 (N1U, N4U, N7U, N9U y N10U); (b) categoría 2 (N1L, N4L, N9L, N10L); (c) categoría 3 (N2, N8, N14 y N15); (d) categoría 4 (N1, N4, N9 y N10; con ambas versiones: más alto y más bajo); (e) categoría 6 (N3U2, N4U2, N11U2, N12U2 y N13U2); y, (f) categoría 7 (N3L2, N4L2, N11L2, N12L2 y N13L2).

González-Ramírez et al., 2009, *Revista Mexicana de Ciencias Geológicas*, 26(2), 501-515

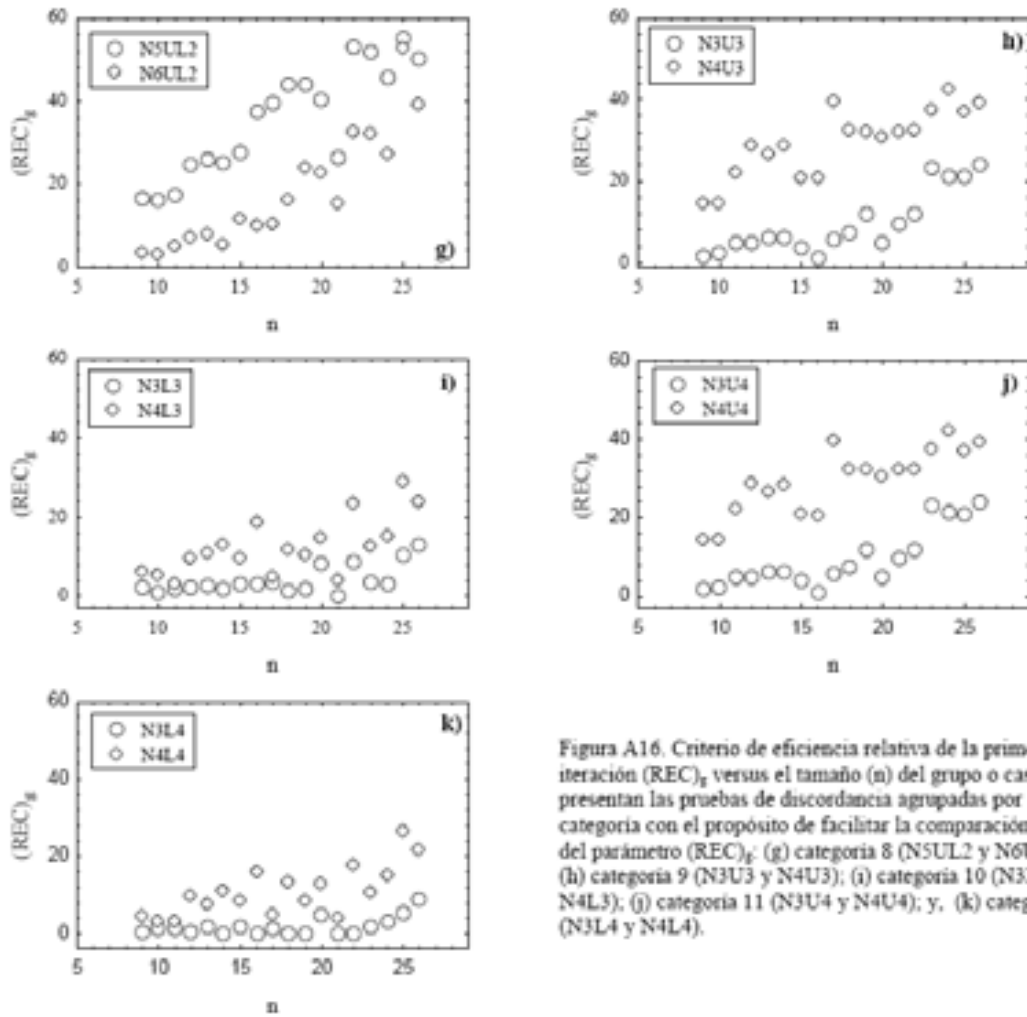


Figura A16. Criterio de eficiencia relativa de la primera iteración $(REC)_g$ versus el tamaño (n) del grupo o caso. Se presentan las pruebas de discordancia agrupadas por categoría con el propósito de facilitar la comparación visual del parámetro $(REC)_g$: (g) categoría 8 (N5UL2 y N6UL2); (h) categoría 9 (N3U3 y N4U3); (i) categoría 10 (N3L3 y N4L3); (j) categoría 11 (N3U4 y N4U4); y, (k) categoría 12 (N3L4 y N4L4).

APÉNDICE 3

Verma, S.P., González-Ramírez R., Rodríguez-Rios R. (2009a)

Comparison of two sample preparation methods in x-ray fluorescence spectrometry for the analysis of Ni and Cr. Geostandards and Geoanalytical Research (enviado: 10 de mayo, 2009; revisado: 10 de agosto, 2009; se espera respuesta del editor).

Geostandards and Geoanalytical Research

**GEOSTANDARDS and
GEOANALYTICAL RESEARCH****Comparison of two sample preparation methods in x-ray
fluorescence spectrometry for the analysis of Ni and Cr**

Journal:	<i>Geostandards and Geoanalytical Research</i>
Manuscript ID:	GGR-0047.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	
Complete List of Authors:	Verma, Surendra; UNAM, CIE Ramírez, Rosalinda Ríos, Rodolfo
Keywords:	X-ray fluorescence spectrometry, statistics, technique intercomparison, uncertainty estimation, trace elements, wavelength dispersive X-ray fluorescence



10-Aug-2009

Dear Verma:

Your manuscript entitled "Comparison of two sample preparation methods in x-ray fluorescence spectrometry for the analysis of Ni and Cr" by Verma, Surendra; Ramírez, Rosalinda; Ríos, Rodolfo, has been successfully submitted online and is presently being given consideration for publication in Geostandards and Geoanalytical Research.

Co-authors: Please contact the Editorial Office as soon as possible if you disagree with being listed as a co-author for this manuscript.

Your manuscript ID is GGR-0047.R1.

Please mention the above manuscript ID in all future correspondence or when calling the office for questions. If there are any changes in your postal address or e-mail address, please log in to Manuscript Central at <http://mc.manuscriptcentral.com/ggr> and edit your user information as appropriate.

You can also view the status of your manuscript at any time by checking your Author Centre after logging in to <http://mc.manuscriptcentral.com/ggr> .

Thank you for submitting your manuscript to Geostandards and Geoanalytical Research.

Yours faithfully,
Geostandards and Geoanalytical Research Editorial Office

**Comparison of two sample preparation methods in x-ray fluorescence
spectrometry for the analysis of Ni and Cr**

Surendra P. Verma ^{1,*}, Rosalinda González-Ramírez ², and Rodolfo Rodríguez-Ríos ^{3,†}

¹Departamento de Sistemas Energéticos, ² Posgrado en Ingeniería,
Centro de Investigación en Energía, Universidad Nacional Autónoma de México, Priv.
Xochicalco s/no., Col Centro, Apartado Postal 34, Temixco 62580, Mexico

³ Facultad de Ingeniería e Instituto de Geología, Universidad Autónoma de San Luis
Potosí, Av. Dr. Manuel Nava No. 8, Zona Universitaria, San Luis Potosí, S.L.P. 78240,
Mexico

* Corresponding author: Phone: +52-55-56229745; Fax number: +52-55-56229766; E-mail address: spv@cie.unam.mx

† Deceased on July 27, 2009

Manuscript submitted to: **Geostandards and Geoanalytical Research**
(Co-Editor: Prof. P.J. Potts), May 10, 2009; Revised August 10, 2009

Abstract

Pressed powder pellets and fused beads or glass disks are routinely used in x-ray fluorescence spectrometry for the analysis of major- and trace-elements, respectively, in geological materials. In order to evaluate the performance of these two sample preparation methods, we determined Ni and Cr concentrations in both powder pellets and glass beads of 14 reference materials (RMs) from Japan, France and South Africa, and 85 igneous and 3 sedimentary rock samples from Mexico. We also computed new values of statistical parameters for RMs from outlier-based multiple-test method and compared them with the literature mean and confidence limit values. The results show that the multiple-test method provided more reliable central tendency and dispersion parameters for RMs than those obtained previously from the two or three standard deviation method, or from robust methods. The powder pellet and fused bead sample preparation methods provided consistent results for greater Ni and Cr concentration levels ($> 50 \mu\text{g g}^{-1}$ in this application); for smaller concentration levels, however, these methods showed somewhat greater differences. For quantitative comparisons, both ordinary and weighted least-squares linear regression models were used to show that the two sample preparation methods provided generally unbiased results.

Keywords: RM, robust methods, outlier-based methods, XRF, x-ray fluorescence spectrometry, WLR, OLR.

Introduction

X-ray fluorescence spectrometry (XRF) is a widely used technique for the analysis of geological materials, in which two different methods of sample preparation (fused beads and pressed powder pellets) are used (e.g., Potts 1987, Potts and Webb 1992). For silicate rock analysis, fused beads or glass disks are generally used for major-elements from SiO₂ to P₂O₅ whereas powder pellets are employed for trace-elements (e.g., Verma *et al.* 1992). Nevertheless, it may be useful to compare the performance of these two sample preparation methods for the analysis of some common elements. We chose Ni and Cr for this purpose because of their petrogenetic importance in geochemical studies (e.g., Rollinson 1993) and because these two elements are relatively easy to analyse in rock samples by XRF. This evaluation was done using 14 reference materials (RMs) and 88 (85 igneous and 3 sedimentary) Mexican rock samples.

Databases and analytical methods

For this study, 14 RMs (most of them could be considered certified reference materials –CRMs– for Cr and Ni, but we are referring to them here as RMs) used for the evaluation of accuracy of our analytical procedure were: andesites JA-2 and JA-3, basalts JB-1a, JB-2, and JB-3, gabbro JGb-1, granite JG-3, and peridotite JP-1 from the Geological Survey of Japan (GSJ), Japan (Imai *et al.* 1995); anorthosite AN-G and granite GS-N from the International Working Group–Groupe International de Travail (IWG-GIT) and Association Nationale de la Recherche Technique (ANRT), France, respectively (Govindaraju 1995); and dunite NIM-D, norite (gabbro) NIM-N, peridotite NIM-P, and

dolorite S-18 from the Council of Mineral Technology (MINTEK), South Africa (Steele *et al.* 1978; Govindaraju 1984).

Individual concentration values for Ni and Cr in 10 of these RMs were compiled from the following references: Steele *et al.* (1972, 1978), Church (1981), Robinson and Bennett (1981), Verma *et al.* (1992), Cantagrel and Pin (1994), Boström and Bach (1995), Han *et al.* (1999), Makishima *et al.* (2002), Naqvi *et al.* (2002), Féménias *et al.* (2003), Hoang and Uto (2003), Kelly *et al.* (2003), Haase *et al.* (2004), Ichiyama and Ishiwatari (2005), Ionov *et al.* (2005), Manikyamba *et al.* (2005), Ishizuka *et al.* (2006), Kurosawa *et al.* (2006), Makishima and Nakamura (2006), Ujiie-Mikoshihira *et al.* (2006), and Hastie *et al.* (2007) as well as from two internet addresses (www.aist.go.jp/RIODB/geost and www.georem.mpch.mainz.gwdg.de). The compiled data for these 10 RMs were processed using DODESYS software developed by S.P. Verma and L. Díaz-González (unpublished). Geochemical data for two RMs (JGb-1 from Japan and GS-N from France) were compiled by Guevara *et al.* (2001) and Marroquín-Guerra *et al.* (2009), respectively, and processed in exactly the same way as done in the present work (using either SIPVADE computer program of Verma *et al.* 1998, or DODESYS). The two remaining RMs (AN-G from France and S-18 from South Africa) were not compiled in this work. The DODESYS software uses new, precise and accurate critical values recently simulated for discordancy tests (Barnett and Lewis 1994) by Verma and Quiroz-Ruiz (2006a, b, 2008) and Verma *et al.* (2008). Discordancy tests of single-outlier type (Verma *et al.* 2009, González-Ramírez *et al.* 2009) were used for processing Ni and Cr data in these RMs. The 95% confidence limits of the mean (CL) were also computed from the final normally distributed RM data

using the best interpolation equations for Student t critical values (Verma 2009), because these values for many sample sizes have not been tabulated in the published literature.

The calculation of CL was as follows: $CL = (t_{95})_{(n-1)} \cdot s / \sqrt{n}$, where t_{95} is the Student t critical value for the sample size n or for $(n-1)$ degrees of freedom and s is the standard deviation of the statistical sample (Verma 2005).

Besides the 14 RMs, 85 samples of igneous and 3 of sedimentary rocks from Mexico were also used for the evaluation of two sample preparation methods (powder pellets and fused beads) for XRF. For 61 samples whole-rock compositions reported by Verma *et al.* (1991), Verma (1999, 2000, 2001a, b, c, 2002, 2003, 2006), and Verma and Carrasco-Núñez (2003), as well as their new Ni and Cr concentrations were also compiled. For the remaining 27 samples, the database included unpublished major-element, Ni and Cr data from the first author. For all samples the adjusted SiO_2 concentrations at an anhydrous basis— $(\text{SiO}_2)_{\text{an}}$ —were also calculated using SINCLAS computer program (Verma *et al.* 2002), which were used in later interpretations.

Pressed powder pellets were prepared by thoroughly mixing about 6 g of sample powder with 0.42 g epoxy (0.29 g of Scandiplex A and 0.13 g of Scandiplex B) pre-diluted with acetone. Hydraulic pressure of about 400 bars was applied for 1 min. The pellets were dried overnight in an oven set at 60°C. The dilution of the sample powder amounted to about 7% for these pellets. The pellets were removed from the dye and stored in a desiccator while not in use for XRF measurements.

Fused beads were prepared by mixing in a platinum crucible accurately weighed about 0.8 g of sample powder with accurately weighed about 4.8 g of $\text{Li}_2\text{B}_4\text{O}_7$ and using a

well-tested heating routine of 21.5 min on a Philips SOL'X. The program included shaking of the crucible.

Duplicate pellets and beads were made in all cases and each pellet or bead was counted four times, thus resulting in a total of 8 measurements. The samples were measured on a Philips XRF spectrometer PE 1404 equipped with a Sc/Mo tube of 3 KW power and a PC. This wavelength-dispersive XRF equipment was previously calibrated for Ni and Cr using fused beads as well as powder pellets. The calibration was achieved from independently made fused beads and powder pellets for 29 RMs (MAG-1, RGM-1, SCo-1, BIR-1, GSP-1, W-2, G-1, W-1, G-2, BHVO-1, BR, AGV-1, SY-2, DR-N, QLO-1, BE-N, DNC-1, GM, BM, MA-N, NIM-G, SDC-1, BCR-1, MRG-1, GH, JB-1, Gn-A, TB, TS, NIM-S, and NIM-L). The measurement conditions for Cr and Ni were 100 kV, 30 mA, and crystal LiF220.

Results and discussion

For Ni and Cr in 12 RMs (all those listed above except AN-G and S-18) the statistical results of our data processing outlier-based multiple-test method initially proposed by Verma (1997, 1998) are compared with the statistics reported in the literature for these RMs (Table 1; Figure 1) reported by Imai *et al.* (1995) for eight RMs from Japan, Steele *et al.* (1978) for three RMs from South Africa, and Govindaraju (1995) for two RMs from France. Different statistical methods had been applied to these RMs in the literature that included the so-called two standard deviation (2s) and three standard deviation (3s) outlier-based methods for the Japanese and South African RMs (Imai *et al.* 1995, Steele *et al.* 1978), respectively, and robust methods for French RMs (Govindaraju 1995).

In Figure 1, the dashed diagonal line in (a) shows equi-values of both x-y parameters. With only one exception, the 95% confidence limits of the mean reported in the literature RM compilations for both elements (Ni and Cr) are consistently greater than those obtained in the present work (Figure 1a), which implies that the mean values computed here for these 12 RMs are statistically better (show smaller confidence intervals of the mean) than those reported in the literature (Steele *et al.* 1978, Govindaraju 1995, Imai *et al.* 1995). A comparison of these mean values (Figure 1b) shows that although for most RMs the literature mean values are within a few % difference with respect to the present mean values, some of them (for Ni in JB-2 and for Cr in NIM-N and NIM-P) show much greater differences (%). Thus, the superiority of the outlier-based multiple-test method for processing RM data is once again documented in the present work (Figure 1a; for earlier comparisons see Verma 1997, 1998, Verma *et al.* 1998). Furthermore, the statistical methods of 2s or 3s used for processing the Japanese and South African RMs have been shown to be erroneous by several authors (Barnett and Lewis 1994, Verma 1998, Verma and Quiroz-Ruiz 2006b, Hays *et al.* 2007) and, therefore, are recommended to be abandoned.

These 14 RMs were also used for the comparison of the two sample preparation methods (powder pellets and fused beads); the results are summarised in Figure 2. The dashed diagonal line in Figure 2a, b, c shows equi-values of both x-y parameters. The mean values obtained from powder pellets are consistent with the mean values from the DODESYS software (Figure 2a). The fused beads also, in general, provided mean concentrations of Ni and Cr consistent with the powder pellets (Figure 2b). However, for

fused beads the 95% confidence limits were greater than those for powder pellets, with only a few exceptions (Figure 2c). Finally, although the mean values for most samples by both sample preparation methods were similar (differences <10%), some data showed somewhat greater differences of up to about 16% (Figure 2d).

A large number of samples (86) were used to evaluate the performance of the two sample preparation methods and the results are presented in Figure 3. The dashed diagonal line in Figure 3a, b shows equi-values of both x-y parameters. The mean concentrations obtained by both methods are generally consistent for Ni or Cr concentrations greater than $50 \mu\text{g g}^{-1}$ (data lie close to the diagonal line in Figure 3a). For smaller concentrations ($<50 \mu\text{g g}^{-1}$), significantly greater differences were observed (see the data more scattered around the diagonal line in Figure 3a). The 95% confidence limits of the mean values for fused beads are, in general, greater than those for powder pellets (Figure 3b), which is fully consistent with the results obtained for RMs (Figure 2c). The % differences in the mean values for glass beads are small ($<25\%$) for greater concentrations ($> 50 \mu\text{g g}^{-1}$), but become much greater (up to about 90%) for smaller ($< 50 \mu\text{g g}^{-1}$) concentrations (Figure 3c).

In order to explore further these % differences, we plotted them against the $(\text{SiO}_2)_{\text{adj}}$ concentrations of these rock samples (Figure 3d), from which it becomes clear that for intermediate and felsic rocks with 52-78% $(\text{SiO}_2)_{\text{adj}}$, these differences become much greater than for mafic rocks with lesser $(\text{SiO}_2)_{\text{adj}}$. This rock-type dependent behaviour is related to generally smaller Ni and Cr contents of intermediate and felsic

rocks than those of mafic ones, in other words, as $(\text{SiO}_2)_{\text{adj}}$ increases in igneous rocks, both Ni and Cr concentrations generally decrease.

The possible bias in Ni and Cr analyses for the two sample preparation methods of powder pellets and fused beads was also explored through both ordinary and weighted least-squares linear regression models (OLR and WLR models, respectively; Baumann 1997, Bevington and Robinson 2003, Sayago *et al.* 2004, Verma 2005, Asuero and González 2007). The results are summarized in Table 2. For 12 RMs the data were from the outlier-based multiple-test method and for AN-G from Govindaraju (1995). S-18 was not used because only mean values with no dispersion estimates were available for this RM (Govindaraju 1984) and we wanted to use the same RMs for both OLR and WLR models in order to make the results directly comparable.

The first set of two equations is for the comparison of powder pellets ($_{pp}$) and this work using DODESYS ($_{m}$) for the Ni data plotted in Figure 2a whereas the second set is for Cr data. For calculations of weights in the WLR model, 95% confidence limits for $_{pp}$ and $_{m}$ used were from Figure 2c and 1a, respectively. The linear correlation coefficient (r) values for both Ni (0.99966 and 0.99890 respectively for OLR and WLR) and Cr (0.99998 and 0.94059 respectively for OLR and WLR) are high, justifying thus a linear fit to the data (Figure 2a). The intercepts of the regression equations are small (from 1.61 ± 1.00 to 10.8 ± 6.7 , i.e., practically close to zero) as expected for unbiased analyses of powder pellets (see the scales approaching $50000 \mu\text{g g}^{-1}$ in Figure 2a). For Ni, the slopes are also close to 1 whereas for Cr these are somewhat lower implying that the XRF method using powder pellets gave slightly lower Cr values than those estimated from the compilation of

literature data and use of DODESYS. Nevertheless, the greater dispersion of slopes from WLR model as compared to the OLR is also observed from these equations (Table 2).

A similar comparison of the two XRF sample preparation methods for the RMs is presented in the next two sets of rows (Table 2). Once again, the intercepts are close to zero for both Ni and Cr and the slopes close to 1 for Cr. The slopes for Ni are somewhat smaller (about 0.96) implying that fused beads provided slightly lower Ni concentrations as compared to powder pellets.

Finally, the comparison of these two preparation methods using a large number of samples (88; see final two rows in Table 2) shows that the intercepts are close to zero and the slopes close to 1. For Cr, the fused bead concentrations might be somewhat greater than those for powder pellets.

Nevertheless, the statistical results of Table 2 imply that the two sample preparation methods for XRF provided unbiased Ni and Cr concentrations.

Conclusions

The multiple-test outlier-based method using DODESYS computer program provided more consistent results for processing RM data than the other outlier-based or robust methods. Both powder pellet and fused bead sample preparation methods provided consistent XRF results for Ni and Cr in RMs and Mexican rocks when the concentrations were greater than about $50 \mu\text{g g}^{-1}$. For smaller concentrations, the powder pellet method performed better for XRF determinations of Ni and Cr than the fused bead method.

Acknowledgements

The second author (RGR) is grateful to SEP for granting permission to carry out doctoral studies at CIE-UNAM. The third author (RRR) thanks his university for granting him a sabbatical leave at CIE-UNAM which enabled his participation in the present research. We are extremely sorrowful to report untimely death of our co-author Rodolfo Rodríguez-Ríos, and dedicate this paper to his memory. We thank the anonymous reviewer and the editor Phil Potts for suggestions on an earlier version of this paper.

References

- Asuero A.G. and González G. (2007)**
Fitting straight lines with replicated observations by linear regression. III. Weighting data. *Critical Reviews in Analytical Chemistry*, 37, 143-172.
- Barnett V. and Lewis T. (1994)**
Outliers in statistical data, John Wiley & Sons, Chichester, 584 pp.
- Baumann K. (1997)**
Regression and calibration for analytical separation techniques. Part II: Validation, weighted and robust regression. *Process Control and Quality*, 10, 75-112.
- Bevington P.R. and Robinson D.K. (2003)**
Data reduction and error analysis for the physical sciences, McGrawHill, NewYork, 320 pp.
- Boström K. and Bach W. (1995)**
Trace Element Determinations by X-ray fluorescence analysis: advantages, limitations, and alternatives. *Proceedings of the Ocean Drilling Program*, 142, 61-68.
- Cantagrel F. and Pin C. (1994)**
Major, minor and rare-earth element determinations in 25 rock standards by ICP-atomic emission spectrometry. *Geostandards Newsletter*, 18, 123-138.
- Church S.E. (1981)**
Multi-element analysis of fifty-four geochemical reference samples using inductively coupled plasma-atomic emission spectrometry. *Geostandards Newsletter*, 5, 133-160.
- Féménias O., Coussaert N., Bingen B., Whitehouse M., Mercier J.-A.C. and Demaiffe D. (2003)**

- Féménias O., Coussaert N., Bingen B., Whitehouse M., Mercier J.-A.C. and Demaiffe D. (2003)**
A Permian underplating event in late- to post-orogenic tectonic setting. Evidence from the mafic-ultramafic layered xenoliths from Beaumont (Frech Massif Central). *Chemical Geology*, 199, 293-315.
- González-Ramírez R., Díaz-González L. and Verma S.P. (2009)**
Eficiencia relativa de 15 pruebas de discordancia con 33 variantes aplicadas al procesamiento de datos geoquímicos. *Revista Mexicana de Ciencias Geológicas*, 26, 501-515.
- Govindaraju K. (1984)**
1984 compilation of working values for 170 international reference samples of mainly silicate rocks and minerals: main text and tables. *Geostandards Newsletter*, 8 (Special Issue), 3-39.
- Govindaraju K. (1995)**
1995 Working values with confidence limits for twenty-six CRPG, ANRT and IWG-GIT geostandards. *Geostandards Newsletter*, 19 (Special Issue), 1-32.
- Guevara M., Verma S.P. and Velasco-Tapia F. (2001)**
Evaluation of GSJ intrusive rocks JG1, JG2, JG3, JG1a, and JG1b. *Revista Mexicana de Ciencias Geológicas*, 18, 74-88.
- Haase K.M., Goldschmidt B. and Garbe-Schönberg C.-D. (2004)**
Petrogenesis of Tertiary continental intra-plate lavas from the Westerwald region, Germany. *Journal of Petrology*, 45, 883-905.
- Han B.-f., Wang S.-g. and Kagami H. (1999)**
Trace element and Nd-Sr isotope constraints on origin of the Chifeng flood basalts, North China. *Chemical Geology*, 155, 187-199.
- Hastie A.R., Kerr A.C., Pearce J.A. and Mitchell S.F. (2007)**
Classification of altered volcanic island rocks using immobile trace elements: development of the Ti-Co discrimination diagram. *Journal of Petrology*, 48, 2341-2357.
- Hayes K., Kinsella A. and Coffey N. (2007)**
A note on the use of outlier criteria in Ontario laboratory quality control schemes. *Clinical Biochemistry*, 40, 147-152.
- Hoang N. and Uto K. (2003)**
Geochemistry of Cenozoic basalts in the Fukuoka district (northern Kyushu, Japan): implications for asthenosphere and lithospheric mantle interaction. *Chemical Geology*, 198, 249-268.
- Ichiyama Y. and Ishiwatari A. (2005)**
HFSE-rich picritic rocks from the Mino accretionary complex, southwestern Japan. *Contrib Mineral Petrol*, 149, 373-387.
- Imai N., Terashima S., Itoh S. and Ando A. (1995)**
1994 compilation of analytical data for minor and trace elements in seventeen GSJ geochemical reference samples, "igneous rock series". *Geostandards Newsletter*, 19, 135-213.

Ionov D.A., Ashchepkov I. and Jagoutz E. (2005)

The provenance of fertile off-craton lithospheric mantle: Sr-Nd isotope and chemical composition of garnet and spinel peridotite xenoliths from Vitim, Siberia. *Chemical Geology*, 217, 41-75.

Ishizuka O., Taylor R.N., Milton J.A., Nesbitt R.W., Yuasa M. and Sakamoto I. (2006)

Variation in the mantle sources of the northern Izu arc with time and space - constraints from high-precision Pb isotopes. *Journal of Volcanology and Geothermal Research*, 156, 266-290.

Kelley K.A., Plank T., Ludden J. and Staudigel H. (2003)

Composition of altered oceanic crust at ODP sites 801 and 1149. *Geochemistry Geophysics Geosystems*, 4, 1-21.

Kurosawa M., Shima K., Ishii S. and Sasa K. (2006)

Trace element analysis of fused whole-rock glasses by laser ablation-ICP-MS and PIXE. *Geostandards and Geoanalytical Research*, 30, 17-30.

Makishima A. and Nakamura E. (2006)

Determination of major, minor and trace elements in silicate samples by ICP-QMS and ICP-SFMS applying isotope dilution-internal standardisation. *Geostandards and Geoanalytical Research*, 30, 245-271.

Makishima A., Kobayashi K. and Nakamura E. (2002)

Determination of chromium, nickel, copper and zinc in milligram samples of geological materials using isotope dilution high resolution inductively coupled plasma-mass spectrometry. *Geostandards Newsletter: The Journal of Geostandards and Geoanalysis*, 26, 41-51.

Manikyamba C., Naqvi S.M., Rao D.V.S., Mohan M.R., Khanna T.C., Rao T.G. and Reddy G.L.N. (2005)

Boninites from the Neoproterozoic Gadwal Greenstone belt, Eastern Dharwar Craton, India: implications for Archaean subduction processes. *Earth and Planetary Science Letters*, 230, 65-83.

Marroquín-Guerra S.G., Velasco-Tapia F., Díaz-González L. (2009)

Evaluación estadística de Materiales de Referencia Geoquímica del Centre de Recherches Pétrographiques et Géochimiques (Francia) aplicando un esquema de detección y eliminación de valores desviados. *Revista Mexicana de Ciencias Geológicas*, 26, 530-542.

Naqvi S.M., Raj B.U., Rao D.V.S., Manikyamba C., Charan S.N., Balaram V. and Sarma D.S. (2002)

Geology and geochemistry of arenite-quartzwacke from the Late Archean Sandur schist belt - implications for provenance and accretion processes. *Precambrian Research*, 114, 177-197.

Potts P.J. (1987)

A handbook of silicate rock analysis. Blackie (Glasgow), 622 pp.

Potts P.J. and Webb P.C. (1992)

X-ray fluorescence spectrometry. *Journal of Geochemical Exploration*, 44, 251-296.

Robinson D. and Bennett M.C. (1981)

XRF determination of 19 trace elements in international geochemical reference samples. *Geostandards Newsletter*, 5, 175-181.

Rollinson H.R. (1993)

Using geochemical data: evaluation, presentation, interpretation, Longman Scientific Technical, Essex, 344 pp.

Sayago A., Boccio M. and Asnero A.G. (2004)

Fitting straight lines with replicated observations by linear regression: the least squares postulates. *Critical Reviews in Analytical Chemistry*, 34, 39-50.

Steele T.W., Russell B.G., Goudvis R.G., Domel G. and Levin J. (1972)

Preliminary report on the analysis of the six NIMROC geochemical standard samples, Johannesburg, National Institute for Metallurgy.

Steele T.W., Wilson A., Goudvis R., Ellis P.J. and Radford A.J. (1978)

Analyses of the NIMROC reference samples for minor and trace elements, Randsburg, South Africa, National Institute for Metallurgy, 218 p.

Ujlie-Mikoshiba M., Imai N., Terashima S., Tachibana Y. and Okai T. (2006)

Geochemical mapping in northern Honshu, Japan. *Applied Geochemistry*, 21, 492-514.

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. The Journal of Geostandards and Geoanalysis*, 21, 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: 215-250.

Verma S.P. (1999)

Geochemistry of evolved magmas and their relationship to subduction-unrelated mafic volcanism at the volcanic front of the central Mexican Volcanic Belt. *Journal of Volcanology and Geothermal Research*, 93, 151-171.

Verma S.P. (2000)

Geochemistry of the subducting Cocos plate and the origin of subduction-unrelated mafic volcanism at the volcanic front of the central Mexican Volcanic Belt. in: H. Delgado-Granados, G. Aguirre-Diaz and J.M. Stock (Eds.), *Cenozoic tectonics and volcanism of Mexico*, Geological Society of America: 195-222.

Verma S.P. (2001a)

Geochemical evidence for a lithospheric source for magmas from Acozulco caldera, eastern Mexican Volcanic Belt. *International Geology Review*, 43, 31-51.

Verma S.P. (2001b)

Geochemical and Sr-Nd-Pb isotopic evidence for a combined assimilation and fractional crystallisation process for volcanic rocks from the Huichapan caldera, Hidalgo, Mexico. *Lithos*, 56, 141-164.

Verma S.P. (2001c)

Geochemical evidence for a rift-related origin of bimodal volcanism at Meseta Río San Juan, North-Central Mexican volcanic belt. *International Geology Review*, 43, 475-493.

Verma S.P. (2002)

Absence of Cocos plate subduction-related basic volcanism in southern Mexico: a unique case on Earth? *Geology*, 30, 1095-1098.

Verma S.P. (2003)

Geochemical and Sr-Nd isotopic evidence for a rift-related origin of magmas in Tizayuca volcanic field, Central Mexican Volcanic Belt. *Journal of the Geological Society of India*, 61, 257-276.

Verma S.P. (2005)

Estadística básica para el manejo de datos experimentales: aplicación en la geoquímica (geoquimiometría), UNAM, Mexico, D.F., 186 pp.

Verma S.P. (2006)

Extension related origin of magmas from a garnet-bearing source in the Los Tuxtlas volcanic field, Mexico. *International Journal of Earth Sciences*, 95, 871-901.

Verma S.P. (2009)

Evaluation of polynomial regression models for the Student t and Fisher F critical values, the best interpolation equations from double and triple natural logarithm transformation of degrees of freedom up to 1000, and their applications to quality control in science and engineering. *Revista Mexicana de Ciencias Geológicas*, 26, 79-92.

Verma S.P. and Carrasco-Núñez G. (2003)

Reappraisal of the geology and geochemistry of Volcán Zamorano, central Mexico: implications for discriminating the Sierra Madre Occidental and Mexican Volcanic Belt provinces. *International Geology Review*, 45, 724-752.

Verma S.P. and 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, 133-161.

Verma S.P. and 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, 302-319.

Verma S.P. and Quiroz-Ruiz A. (2008)

Critical values for 33 discordancy test variants for outliers in normal samples for very large sizes of 1,000 to 30,000. *Revista Mexicana de Ciencias Geológicas*, 25, 82-96.

Verma S.P., Carrasco-Núñez G. and Milán M. (1991)

Geology and geochemistry of Amealco caldera, Qro., Mexico. *Journal of Volcanology and Geothermal Research*, 47, 105-127.

Verma S.P., Besch T., Guevara M. and Schulz-Dobrich B. (1992)

Determination of twelve trace elements in twenty-seven and ten major elements in twenty-three geochemical reference samples by X-Ray fluorescence spectrometry. *Geostandards Newsletter*, 16, 301-309.

Verma S.P., Orduña-Galván L.J. and 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: The Journal of Geostandards and Geoanalysis*, 22, 209-234.

Verma S.P., Torres-Alvarado I.S. and Sotelo-Rodríguez Z.T. (2002)

SINCLAS: standard igneous norm and volcanic rock classification system. *Computers & Geosciences*, 28, 711-715.

Verma S.P., Quiroz-Ruiz A. and Díaz-González L. (2008)

Critical values for 33 discordancy test variants for outliers in normal samples up to sizes 1000, and applications in quality control in Earth Sciences. *Revista Mexicana de Ciencias Geológicas*, 25, 82-96.

Verma S.P., Díaz-González L. and González-Ramírez R. (2009)

Relative efficiency of single-outlier discordancy tests for processing geochemical data on reference materials. *Geostandards and Geoanalytical Research*, 33, 29-49.

Figure legends

Figure 1. Statistical evaluation of literature compilations with the present work. The symbols used in this and later figures are explained as inset in (a). (a) Comparison of 95% confidence limits for this work $(CL)_m$ with 95% confidence limits calculated for the literature $(CL)_l$ compilations; (b) The % difference $(\{(x_l - x_m)/x_m\} * 100)$ of literature mean (x_l) values as a function of the mean values (x_m) for this work.

Figure 2. Statistical comparison of two XRF sample preparation methods (pressed powder pellets – pp and fused beads – b) using reference materials. (a) Comparison of mean values obtained in this work (x_m) with those from powder pellets (x_{pp}) ; (b) Comparison of mean values obtained from powder pellets (x_{pp}) with those for the fused beads (x_b) ; (c) Comparison of 95% confidence limits for powder pellets $(CL)_{pp}$ with 95% confidence limits for fused beads $(CL)_b$; (d) The % difference $(\{(x_b - x_{pp})/x_{pp}\} * 100)$ of mean (x_b) values for fused beads as a function of the mean values (x_{pp}) for powder pellets for the RMs analysed in this study.

Figure 3. Statistical comparison of two XRF sample preparation methods (pressed powder pellets – pp and fused beads – b) using Mexican igneous and sedimentary rocks. The symbols used are explained as inset in (a) where (ig) stands for igneous rocks and

(sed) for sedimentary rocks. (a) Comparison of mean values obtained from powder pellets (x_{pp}) with those for the fused beads (x_b); (b) Comparison of 95% confidence limits for powder pellets $(CL)_{pp}$ with 95% confidence limits for fused beads $(CL)_b$; (c) The % difference ($\{(x_b - x_{pp})/x_{pp}\} * 100$) of mean (x_b) values for fused beads as a function of the mean values (x_{pp}) for powder pellets; (d) The % difference ($\{(x_b - x_{pp})/x_{pp}\} * 100$) of mean (x_b) values for fused beads as a function of the silica concentrations $(SiO_2)_{ab}$ for the samples analysed in this study.

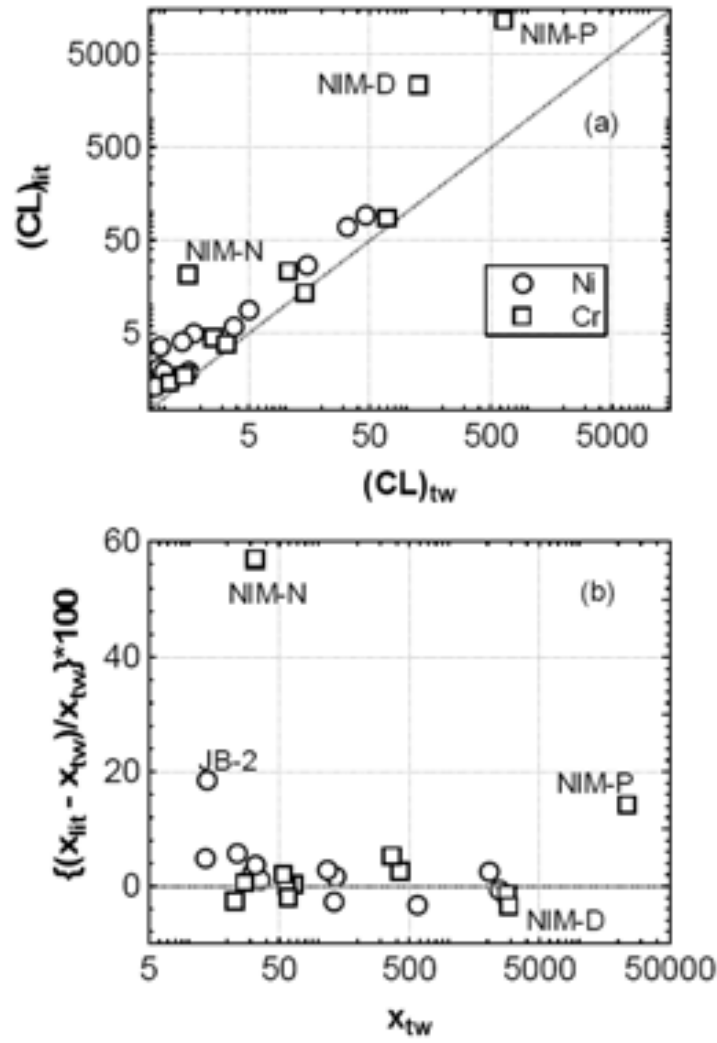


Figure 1. (Verma *et al.*)

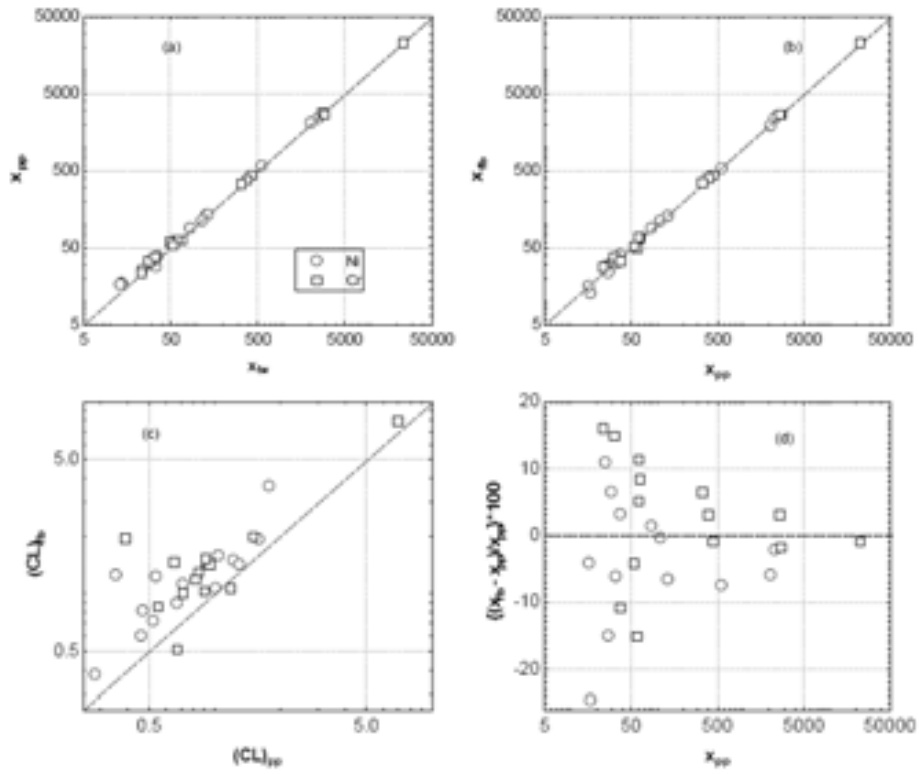


Figure 2. (Verma *et al.*)

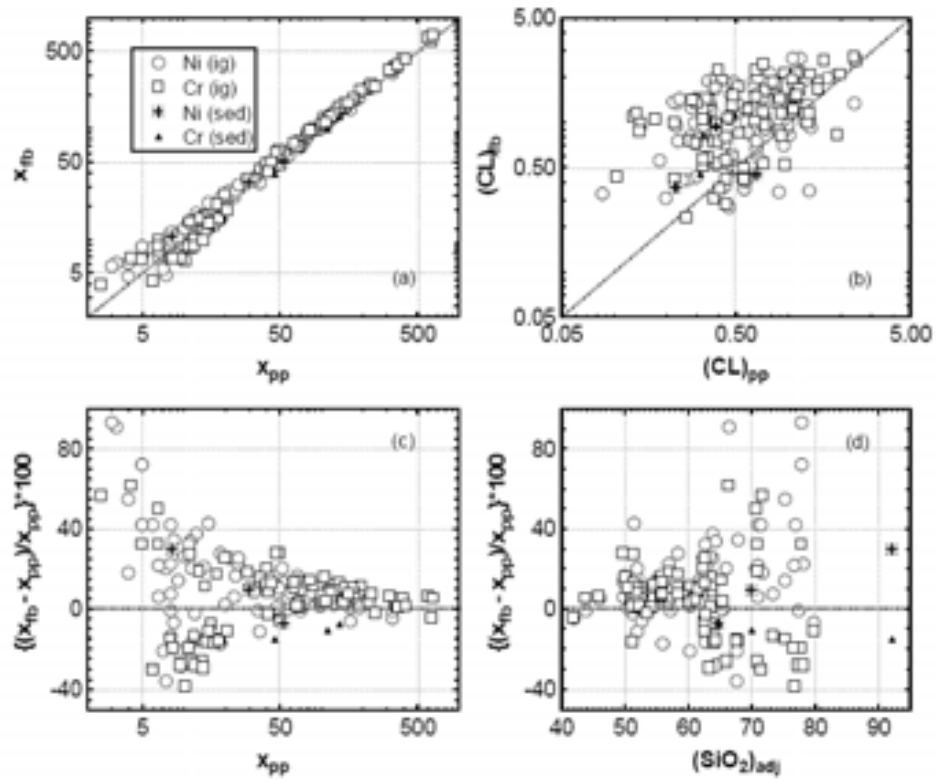


Figure 3. (Verma *et al.*)

Table 1.
Ni and Cr concentrations ($\mu\text{g g}^{-1}$) in RMs used in this study

RM	Literature						This work (DODESYS)							
	Ni			Cr			Ni			Cr				
Code	Type	n_{lit}	\bar{X}_{lit}	s_{lit}	n_{lit}	\bar{X}_{lit}	s_{lit}	R_{lit}	n_{tw}	\bar{X}_{lit}	s_{lit}	n_{tw}	\bar{X}_{tw}	s_{tw}
JA-2	Andesite	25	130	14	29	436	59	*	48	133.6	12.9	52	425.9	38.9
JA-3	Andesite	16	32.2	6.6	23	66.2	10	*	22	31.54	2.12	41	66.0	7.9
JB-1a	Basalt	37	139	15	36	392	40	*	39	136.9	5.5	40	373	47
JB-2	Basalt	41	16.6	6.7	54	28.1	5.3	*	52	14.01	3.25	77	28.0	4.9
JB-3	Basalt	42	36.2	6.1	43	58.1	5.6	*	66	35.83	4.12	68	58.2	6.2
JGb-1	Gabbro	37	25.4	5.7	38	57.8	13.5	*	40	24	5	40	59	8
JG-3	Granite	13	14.3	2.2	21	22.4	3.5	*	28	13.64	2.22	32	23.03	1.63
JP-1	Peridotite	28	2458	177	26	2807	211	*	35	2474	97	42	2852	227
AN-G	Anorthosite	63	35	9.75	68	50	10.4	**	---	---	---	---	---	---
GS-N	Granite	31	34	11.08	33	55	10.49	**	26	32.8	3.5	30	54	9
NIM-D	Dunite	52	2091	334.9	52	2813	8125	***	49	2038	162	53	2916	471
NIM-N	Norite	46	122.4	30.22	40	51.3	64.4	***	51	119.1	17.7	37	32.72	4.83
NIM-P	Peridotite	52	536	96.31	52	27069	41036	****	61	574	60	41	23710	2060
S-18	Dolerite	---	---	86	---	---	325	*****	---	---	---	---	---	---

n is the number of observations; \bar{X} is the central tendency parameter; s is the standard deviation (dispersion) parameter; R_{lit} is the literature reference code (* Imai *et al.* 1995; ** Govindaraju 1995; *** Steele *et al.* 1978; **** Govindaraju 1984); the subscripts lit and tw are literature and this work, respectively; --- information not available. For this work, the computer program DODESYS (Verma and Diaz-González, in preparation) was used; the processed parameters for JGb-1 and GS-N were taken, respectively, from Guevara *et al.* (2001) and Marroquin-Guerra *et al.* (2009), who also used DODESYS or similar computer program for processing their database. The new compiled data are reported as rounded using less strict rules as suggested by Verma (2005).

Table 2.
Statistical information about the relationships presented in Figures 2 and 3

x-y variables (Figure) Element	n_{par}	OLR model (r)	WLR model (r)
$(x_{\text{rw}} - x_{\text{pp}})$ (Fig. 2a, 2c, 1a) Ni (RMs)	13	$x_{\text{pp}} = (1.7 \pm 7.0) + (0.9790 \pm 0.0077)x_{\text{rw}}$ (0.99966)	$x_{\text{pp}} = (-0.61 \pm 1.00) + (0.9724 \pm 0.0138)x_{\text{rw}}$ (0.99890)
$(x_{\text{rw}} - x_{\text{pp}})$ (Fig. 2a) Cr (RMs)	13	$x_{\text{pp}} = (3.8 \pm 13.4) + (0.95250 \pm 0.00207)x_{\text{rw}}$ (0.99998)	$x_{\text{pp}} = (10.8 \pm 6.7) + (0.916 \pm 0.100)x_{\text{rw}}$ (0.94059)
$(x_{\text{pp}} - x_{\text{fb}})$ (Fig. 2b) Ni (RMs)	13	$x_{\text{fb}} = (-1.7 \pm 5.6) + (0.9614 \pm 0.0066)x_{\text{pp}}$ (0.99972)	$x_{\text{fb}} = (0.08 \pm 0.95) + (0.9570 \pm 0.0078)x_{\text{pp}}$ (0.99630)
$(x_{\text{pp}} - x_{\text{fb}})$ (Fig. 2b) Cr (RMs)	13	$x_{\text{fb}} = (9.8 \pm 8.8) + (0.99104 \pm 0.00144)x_{\text{pp}}$ (0.9999)	$x_{\text{fb}} = (0.460 \pm 2.33) + (1.0129 \pm 0.0098)x_{\text{pp}}$ (0.99949)
$(x_{\text{pp}} - x_{\text{fb}})$ (Fig. 3a, 3b) Ni (Mexico)	88	$x_{\text{fb}} = (2.97 \pm 0.62) + (0.9939 \pm 0.0073)x_{\text{pp}}$ (0.99769)	$x_{\text{fb}} = (3.22 \pm 0.73) + (0.9940 \pm 0.0066)x_{\text{pp}}$ (0.99809)
$(x_{\text{pp}} - x_{\text{fb}})$ (Fig. 3a, 3b) Cr (Mexico)	88	$x_{\text{fb}} = (1.86 \pm 1.17) + (1.027 \pm 0.0066)x_{\text{pp}}$ (0.99821)	$x_{\text{fb}} = (-0.68 \pm 0.65) + (1.0340 \pm 0.0070)x_{\text{pp}}$ (0.99805)

x-y variables refer to the Figure #; n_{par} -number of x-y paired data used in the regressions; OLR-ordinary least-squares linear regression model; WLR-weighted least-squares linear regression model; r-linear correlation coefficient; the quoted errors on regression coefficients are one standard error values. The subscripts rw , pp , and fb are for this work, powder pellets, and fused beads, respectively; x in these equations refer to mean values. The first term on the right hand side of each equation is the intercept and the second term gives the slope of the regression line.

APÉNDICE 4

Verma S.P., Rodriguez-Rios, R., González-Ramírez R. (2009b)

Statistical evaluation of classification diagrams for altered igneous rocks. Turkish Journal of Earth Sciences (enviado: 22 de enero, 2009; aceptado con cambios menores: 3 de Julio, 2009; revisado: 15 de julio, 2009; disponible en línea de la revista: 14 de agosto, 2009)

Statistical Evaluation of Classification Diagrams for Altered Igneous Rocks

SURENDRA P. VERMA^{1*}, RODOLFO RODRÍGUEZ-RÍOS^{2,1,†}, ROSALINDA GONZÁLEZ-
RAMÍREZ³

¹ *Departamento de Sistemas Energéticos, Centro de Investigación en Energía, Universidad Nacional Autónoma de México, Temixco, Mor. 62580, Mexico.*

² *(on sabbatical leave from) Facultad de Ingeniería e Instituto de Geología, Universidad Autónoma de San Luis Potosí, Av. Dr. Manuel Nava No. 8, Zona Universitaria, San Luis Potosí, S.L.P. 78240, Mexico.*

³ *Posgrado en Ingeniería – Energía, Centro de Investigación en Energía, Universidad Nacional Autónoma de México, Temixco, Mor. 62580, Mexico.*

[†] Deceased on July 27, 2009.

* Corresponding author; email spv@cie.unam.mx

Manuscript submitted to Turkish Journal of Earth Sciences, 22 January, 2009; revised manuscript submitted on: 15 July, 2009

Abstract:

The International Union of Geological Sciences (IUGS) has proposed recommendations for the classification of relatively fresh volcanic rocks, with no specific indications for altered volcanic rocks, besides discouraging the use of the total alkalis *versus* silica diagram. The Nb/Y–Zr/TiO₂ diagram has been in use for the classification of altered rocks now for over 30 years. Recently (during 2007) another diagram (Co–Th) has been proposed to replace this old diagram, particularly for altered arc rocks. Using an extensive database of all kinds of relatively fresh rocks from four tectonic settings (island arc, continental rift, ocean island, and mid-ocean ridge) as well as from three settings excluding island arc, we carried out an objective evaluation of the old Nb/Y–Zr/TiO₂ diagram for rock classification. Similarly, for the evaluation of the new Co–Th diagram, an extensive database of similar rocks from island arcs, continental arc of the Andes, continental rifts, ocean islands, and the Mexican Volcanic Belt, was used. Statistical parameters of correct classification or success rate and minimum mis-classification defined in this work, respectively, were used to evaluate these diagrams. Our results of the quantification of these parameters showed that none of these diagrams seems to work properly for the classification of fresh rocks. Then, it is difficult to imagine that they would work well for the classification of altered rocks. Thus, an urgent need of applying correct statistical methodology for handling compositional data exists for proposing new classification diagrams that could provide classification and nomenclature to altered volcanic rocks fully consistent with the IUGS nomenclature for fresh rocks.

Keywords: TAS classification, volcanic rocks, plutonic rocks, chemical classification, correct statistical analysis of compositional data.

Introduction

The classification and nomenclature in any science are fundamental issues because the accuracy of the language used for communication in that particular science depends on them. For the classification of relatively fresh igneous rocks, the International Union of Geological Sciences (IUGS) has made specific recommendations for assigning rock names that depend on their mineralogical and chemical characteristics (Le Bas *et al.* 1986; Le Bas 2000; Le Maitre *et al.* 2002). The well-known TAS (total alkalis *versus* silica; Le Bas *et al.* 1986) diagram seems to be the most popular and widely used for the classification of volcanic rocks. Verma *et al.* (2002) presented a computer program (SINCLAS) to be used for the IUGS volcanic rock classification scheme, which facilitated the application of the TAS diagram as well as provided a standard way of calculating the CIPW norm (Verma *et al.* 2003). In fact, the classification of volcanic rocks and their nomenclature depend on both concepts—the TAS diagram and the CIPW norm (Le Maitre *et al.* 2002; Verma *et al.* 2002). However, the IUGS failed to provide any specific recommendations for the classification of altered rocks, besides discouraging the use of their procedure of relatively fresh rocks for this purpose (Le Bas *et al.* 1986).

In the published literature, some diagrams (alternative to the TAS diagram) have long been proposed using the so-called immobile elements (Floyd & Winchester 1975, 1978; Winchester & Floyd 1976, 1977), which have been cited in thousands of published papers. In fact, these diagrams, particularly the Nb/Y–Zr/TiO₂ diagram of Winchester & Floyd (1977), have been in wide use even today. Just to name a few references during 2007-2008, we can cite: Gokten & Floyd (2007); Shekhawat *et al.* (2007); Ahmad *et al.* (2008); Bagci *et al.* (2008);

Gladkochub *et al.* (2008); Gursu (2008); Kadir *et al.* (2008); Keskin *et al.* (2008); Kalmar & Kovacs-Palffy (2008); Kaygusuz *et al.* (2008); Mondal *et al.* (2008); Nardi *et al.* (2008); Pandarinath *et al.* (2008); Sheth & Melluso (2008); Wang *et al.* (2008); Yigitbas *et al.* (2008); and Zheng *et al.* (2008).

More recently, the subject of the classification of altered rocks has been revived through the publication of a paper by Hastie *et al.* (2007) who stated that the existing diagrams did not work well for arc rocks and proposed, more specifically, the use of Co-Th diagram for the classification of altered rocks from volcanic arcs.

The question arises if these older (Floyd & Winchester 1975, 1978; Winchester & Floyd 1976, 1977) and the most recent (Hastie *et al.* 2007) diagrams “correctly” classify altered rocks. We cannot answer for sure this question by studying altered rocks because we do not know up to what extent their chemical composition was modified by alteration processes in the field. We could, of course, resort to experimental laboratory-controlled work to answer it, which would also be costly, time consuming, and difficult due to the multivariate nature of this problem. Therefore, we adopted the philosophy of objectively testing the functioning of these diagrams using data for fresh volcanic rocks from different areas and tectonic settings. If the classification diagrams were shown to work well for fresh rocks, i.e., if they showed that high percentages of fresh rocks are named correctly and consistently with the IUGS classification scheme (combination of the TAS diagram and CIPW norm), we could expect that they might work well for altered rocks as well, provided the concentrations of the chemical elements used in these diagrams were not significantly modified during the alteration processes. Thus, the percentages

of correct classification in such diagrams would probably represent approximately the maximum percentages of correct classification for altered rocks.

With this philosophy in mind, the following methodology was applied for the present evaluation: (a) compile databases for fresh volcanic rocks from different tectonic settings; (b) separate samples of a given rock type from the compiled databases; (c) plot these samples of a particular rock type in the diagram to be evaluated and determine the new rock names; (d) count samples of each new rock name as classified in the evaluated diagram; (e) calculate statistical information as the percentages of each new rock type in terms of the original samples of that particular rock type being evaluated; (f) repeat this procedure for all rock types from the IUGS classification scheme; and (g) report the results in figures and tables and point out their implications.

Specifically, two diagrams –the old Nb/Y–Zr/TiO₂ diagram of Winchester & Floyd (1977) and the new Co–Th diagram of Hastie *et al.* (2007)– were evaluated in detail. The results clearly show that none of them works satisfactorily, highlighting thus the urgent need of proposing new, more efficient diagrams, for which the statistically correct methodology for handling compositional data must be used.

Databases

The data were compiled for all rock types of Miocene to Recent age from different areas of known, uncontroversial tectonic settings from all over the world. Initially, databases from island

arcs, continental rifts, ocean islands, and mid-ocean ridges, as well as the Mexican Volcanic Belt (MVB) and the continental arc of the Andes, were established and used by Verma & Aguilar-Y-Vargas (1988); Verma (1997, 2000, 2002, 2004, 2006, 2009a, 2009b); Vasconcelos-F. *et al.* (1998, 2001), Agrawal *et al.* (2004, 2008); Verma *et al.* (2006); and Agrawal & Verma (2007). An updated version of these databases was prepared and used for the present work. Specifically, Verma *et al.* (2006) presented the information on the number of samples, their tectonic setting and location coordinates, and literature references. Later, Agrawal *et al.* (2008) stated that Electronic Annexure EA-1, with such information on additional samples compiled by them, is available upon request from the authors. Additional details are given in a companion paper by Verma (2009b). Therefore, these details are omitted from the present paper to avoid repetition.

All data, except those from the MVB and the Andes, were used for evaluation of the old Nb/Y–Zr/TiO₂ diagram by Winchester & Floyd (1977). Furthermore, a second evaluation of this old diagram was also carried out using rocks from only three tectonic settings of continental rift, ocean island and mid-ocean ridge.

For the evaluation of the new Co–Th diagram by Hastie *et al.* (2007), data from island arcs, continental rifts, and ocean islands as well as the MVB and the continental arc of the Andes were used separately. MORB data were not used here because, as expected, our compilation for this setting was mostly of basic rocks, and we wanted to cover all types of rocks from a given tectonic setting. The Andes data were an updated version of the compilation by Verma *et al.* (2006).

The rock names of all compiled rocks were ascertained using SINCLAS computer program (Verma *et al.* 2002, 2003), which also provided standard igneous norm according to the IUGS recommendations (Le Bas *et al.* 1986; Le Bas 2000; Le Maitre *et al.* 2002). Note that SINCLAS also provides adjusted data (identified here as the subscript _{adj}) on an anhydrous 100% basis with a prior adjustment of Fe-oxidation ratio. The rest of the methodology was the same as outlined above in the Introduction section.

Results

The results are arranged in two following subsections.

1. Old classification diagrams

Floyd & Winchester (1975, 1978) and Winchester & Floyd (1976, 1977) presented several diagrams for the classification of altered rocks. These were of the following three types: (1) element-element (i) Zr-P₂O₅; and (ii) Zr-TiO₂; (2) element-element ratio (iii) Ce-Zr/TiO₂; (iv) Ga-Zr/TiO₂; (v) Zr/TiO₂-SiO₂; (vi) Nb/Y-SiO₂; (vii) Y/Nb-TiO₂; and (viii) Zr/P₂O₅-TiO₂; and (3) element ratio-element ratio (ix) Nb/Y-Zr/TiO₂; (x) Nb/Y-Ga/Sc; and (xi) Zr/P₂O₅-Nb/Y.

Several diagrams –(i) Zr-P₂O₅; (ii) Zr-TiO₂; (vii) Y/Nb-TiO₂; (viii) Zr/P₂O₅-TiO₂; (ix) Nb/Y-Zr/TiO₂; and Zr/P₂O₅-Nb/Y– were proposed (Floyd & Winchester 1975; Winchester & Floyd 1976) to classify only two types of basaltic rocks – tholeiitic and alkali. The term tholeiite has not been recommended by the IUGS (Le Bas *et al.* 1986; Le Bas 2000; Le Maitre *et al.*

2002). Because for this evaluation we wanted to strictly follow the IUGS recommendations for the rock classification and nomenclature, it was not possible to separate tholeiites from alkali basalt in our database by the IUGS scheme. Therefore, these diagrams cannot be evaluated if we were to use the IUGS nomenclature as the reference frame for our work.

The diagrams (v) Zr/TiO_2-SiO_2 and (vi) $Nb/Y-SiO_2$ (Winchester & Floyd 1977; Floyd & Winchester 1978), both involving SiO_2 and having been proposed to classify all volcanic rock types, are also not worth evaluating for several reasons. First, they involve as one of the axes, viz., SiO_2 , of the TAS diagram. The names inferred from Zr/TiO_2-SiO_2 and $Nb/Y-SiO_2$ are likely to be similar to the TAS diagram, because in both the subdivision basalt-andesite-dacite-rhyolite depends on the SiO_2 content. However, the subdivision proposed by Winchester & Floyd (1977) does not fully match with that of the IUGS (Le Bas *et al.* 1986), for example, in the former, basaltic andesite is absent and rhyodacite is present. These differences will be simply reflected in the evaluation. Second, SiO_2 may also be somewhat variable under alteration processes, for example, under geothermal conditions (e.g., Fournier & Potter II 1982; Verma & Santoyo 1997; M.P. Verma 2000; Torres-Alvarado 2002; Pandarinath *et al.* 2006; Torres-Alvarado *et al.* 2007). Silica is known to dissolve from rocks –especially from basic rocks– from interaction with water at greater temperatures than those of the surface ambient conditions. This is why the well known silica geothermometers actually work for inferring subsurface temperatures in geothermal systems (e.g., Fournier & Potter II 1982; Verma & Santoyo 1997; Díaz-González *et al.* 2008; Palabiyik & Serpen 2008; Verma *et al.* 2008a). Finally, because the classification depends on both axes, the other parameter – Zr/TiO_2 or Nb/Y – might affect the rock names if they are not perfect proxies for total alkalis. The behaviour of these two ratio variables can be better evaluated in the $Nb/Y-Zr/TiO_2$ diagram (see below).

Winchester & Floyd (1977) also presented (iii) Ce–Zr/TiO₂ and (iv) Ga–Zr/TiO₂ diagrams for rock classification. However, they also noted that these diagrams did not perform so well as the Nb/Y–Zr/TiO₂ diagram, because different basalt types and basanite were not clearly distinguished and, for subalkaline magmas, neither Ce nor Ga showed any significant increase with differentiation, i.e., no significant change with increasing SiO₂. Consequently, although these authors presented these two diagrams, they did not recommend their use for rock classification purposes.

The 10th diagram –(x) Nb/Y–Ga/Sc– proposed by Winchester & Floyd (1977), was also not evaluated because the authors noted that the data, on which this diagram was based, were particularly scarce and the classification boundaries were not definitive. No new boundaries were later proposed by these authors. Besides, the functioning of the Nb/Y parameter will be evaluated in the Nb/Y–Zr/TiO₂ diagram.

Thus, in spite of almost a dozen of these old diagrams, only Nb/Y–Zr/TiO₂ diagram of Winchester & Floyd (1977) –henceforth called, for simplicity, the W&F diagram– was evaluated in this work. The results are presented in Figures 1-7. The numbers of the IUGS (TAS+CIPW norm) classified samples for each rock type as well as those of the W&F diagram classified samples were calculated. Thus, for a given IUGS rock name, the total number of samples was identified and assumed to represent 100%. The numbers of samples plotting in all fields of the W&F diagram were divided by the initial number of samples of that particular rock type used for the evaluation and the ratios were expressed as percentages of W&F classification. When the W&F field had the same name as the initial IUGS rock name, it was said to represent correct

classification or correct success rate (identified as *italic boldface* in Tables 1 and 2), whereas when the W&F field name differed from the IUGS, it was said to quantify mis-classification (expressed as simple numbers –without highlighting– in Tables 1 and 2). All statistical information including the number of samples and the calculated percentages are included in Tables 1 and 2, respectively, for all data from four tectonic settings and those from three tectonic settings except island arc. For the IUGS rock names not present in the W&F diagram (second part of Tables 1 and 2), the highest percentage of the resulting rock W&F types was highlighted in *italics*.

We start the discussion with those rock types that exist in both the TAS and W&F classification. Then, those rock names absent from the W&F classification will be mentioned. The results of three tectonic settings –without arc rocks– will be discussed at the end of this subsection. In order to help the reader better understand our evaluation procedure, the results for alkali basalt and subalkali basalt samples (Table 1) are presented at greater details than the remaining rock types.

In our database there were 650 samples of alkali basalt and 1200 of subalkali basalt as classified from the IUGS nomenclature (the combination of TAS diagram and CIPW norm; Le Bas *et al.* 1986; Le Bas 2000; Le Maitre *et al.* 2002; Verma *et al.* 2002) –alkali basalt being a nepheline normative rock and subalkali basalt a hypersthene normative rock, both of them with adjusted silica $(\text{SiO}_2)_{\text{adj}}$ between 45% and 52% and adjusted total alkalis $(\text{Na}_2\text{O}+\text{K}_2\text{O})_{\text{adj}}$ up to 5%. The corresponding TAS diagrams showing these alkali basalt and subalkali basalt samples are given in Figures 1a and 1b, respectively. Now, if these fresh rocks were to be classified correctly in the W&F diagram (Figure 1c, d; Nb/Y–Zr/TiO₂ diagram of Winchester & Floyd

1977), most of them (a high percentage) should be classified as alkali basalt and sub-alkaline basalt, respectively.

For alkali basalt, we observed (Table 1) that out of 650 (designated as 100%) samples from our databases, the correct classification according to the W&F diagram amounted to 480 (about 73.8%) samples. The mis-classification of 170 (about 26.2%) samples (Table 1; Figure 1c) was as follows: 81 (about 12.5%) samples as sub-alkaline basalt; 59 (about 9.1%) as andesite/basalt; 20 (about 3.1%) as basanite/nephelinite; 9 (about 1.4%) as andesite; and 1 (about 0.1%) as trachyandesite. For subalkali basalt, on the other hand, we observed that out of 1200 (100%) samples, only 384 (about 32.0%) were sub-alkaline basalt and most of them, i.e., the remaining 816 (about 68.0%) were mis-classified (Figure 1d; Table 1). The mis-classification for subalkali basalt (Table 1) ranged as follows: 470 (about 39.2%) samples as of ambiguous type andesite/basalt; 313 (about 26.1%) as alkali-basalt; 23 (about 1.9%) as andesite; 9 (about 0.7%) as basanite/nephelinite; and 1 (about 0.1%) as rhyodacite/dacite. We also note that none of the alkali basalt samples was mis-classified as trachyte, phonolite, comendite/pantellerite, rhyodacite/dacite, or rhyolite. Similarly, none of the subalkali basalt was mis-classified as trachyandesite, trachyte, phonolite, comendite/pantellerite, or rhyolite.

The results of basanite and andesite samples are plotted in Figure 2a-d. A total of 541 samples were separated as basanite from our database (Figure 2a; Table 1). The IUGS makes a distinction or sub-classification of basanite as basanite, melanephelinite and nephelinite, depending on the relative proportions of normative olivine, albite, and nepheline minerals (see Verma *et al.* 2002 for details). In the W&F diagram both basanite and nephelinite occupy exactly the same field (Figure 2c). Therefore, we did not make any further distinction of IUGS

classification of basanite. Most of these 541 (100%) samples of basanite were mis-classified by the W&F scheme, with 322 (59.5%) of them being mis-classified as alkali-basalt (Figure 2c; Table 1). Only 161 (29.8%) samples were correctly classified as basanite/nephelinite. The remaining basanite samples were mis-classified as trachyte (47 samples) and sub-alkaline basalt (11 samples). Our database included 941 andesite samples according to the IUGS classification scheme (Figure 2b), which were plotted in Figure 2d of W&F diagram. Of these, 706 (75.0%) samples were correctly identified as andesite (Table 1) followed by 87 (9.3%) as ambiguous andesite/basalt. The remaining mis-classification consisted of 40 (4.3%) samples as alkali-basalt, 39 (4.1%) as rhyodacite/dacite, 35 (3.7%) as trachyandesite, and 34 (3.6%) as sub-alkaline basalt.

Evaluation of the W&F diagram for trachyandesite, trachyte and phonolite is presented in Figure 3a-f and summarised in Table 1. Our testing through 222 trachyandesite samples from our database (Figure 3a) revealed that this type of rock was very poorly classified in the W&F diagram (Figure 3b), with only 65 (29.3%) samples correctly classified as such (Table 1). Most samples (85 representing 38.3%) were mis-classified as andesite (Table 1). This mis-classification was followed by 40 (18.0%) samples as ambiguous types (rhyodacite/dacite) and 16 (7.2%) samples as alkali-basalt (Table 1), with the remaining (16) samples as other rock types. The evaluation of the W&F diagram using 81 trachyte samples (Figure 3c) revealed that only 24 (about 29.6%) samples were correctly classified as trachyte, with the remaining mostly mis-classified as rhyodacite/dacite and trachyandesite (30 samples–37.1% and 15 samples–18.5%, respectively; Figure 3d; Table 1). Only 49 samples were classified as phonolite in our database (Figure 3e). According to W&F, the correct classification for them amounted to 37 (about 76%) as phonolite, with most (11) remaining samples (22%) being mis-classified as trachyte (Figure 3f; Table 1).

Finally, our database had a fairly large number of samples of dacite and rhyolite (524 and 350, respectively) as determined by the IUGS classification (Table 1; Figure 4a, b). Their correct classification by the W&F diagram was very poor (Figure 4c, d), with only 133 (25.4%) samples as rhyodacite/dacite and 58 (16.6%) as rhyolite, respectively. The majority of samples, therefore, were mis-classified (Table 1). Dacite samples were mis-classified mostly as andesite (314 samples; 59.9%) and rhyolite samples as comendite/pantellerite (109 samples; 31.1%) and rhyodacite/dacite (94 samples; 26.8%).

The remaining rock names of the IUGS classification were not actually included in the W&F classification. These are briefly treated approximately from high-Mg varieties and ultrabasic to acid types. The lower part of Table 1 also provides statistical information on the W&F classification of these samples. The rock names comendite and pantellerite used by Winchester & Floyd (1977) were not incorporated by the IUGS scheme (Le Bas *et al.* 1986; Le Bas 2000; Le Maitre *et al.* 2002) and, therefore, could not be evaluated.

For high-Mg picrite magmas, our database provided 157 samples. Note that picrites are not classified by the TAS diagram (Le Bas 2000). However, to continue to explore the relationship of the TAS diagram with the W&F diagram, we made an artificial distinction of picrites according to the TAS field in which they would plot (Figure 5a). Thus, these 157 samples of picrites were sub-divided as: 6 samples only as picrite (foidite); 45 as picrite (microbasalt); and 106 as picrite (alkali basalt). The W&F classification of these six picrite (foidite) samples was not considered statistically significant. The 45 picrite samples of microbasalt type were classified (Table 1; Figure 5b) as alkali-basalt (27 samples), basanite/nephelinite (16 samples) and sub-

alkaline basalt (2 samples). The 106 picrite samples of alkali basalt type, on the other hand, were classified as andesite/basalt (43 samples), alkali-basalt (34 samples) and sub-alkaline basalt (29 samples). Our databases included 118 samples (Figure 5c) of foidite, an ultrabasic rock. When these samples were plotted in the W&F diagram (Figure 5d), we observed that 45 samples (about 38.1%) were classified as alkali-basalt, 37 (31.4%) as basanite/nephelinite, 33 (28.0%) as andesite, 2 (1.7%) as sub-alkaline basalt, and 1 (0.8%) as andesite/basalt. Only 30 samples of picrobasalt were compiled in our database (Figure 5e), which were classified as alkali-basalt and sub-alkaline basalt in the W&F diagram (20 and 10 respectively; Figure 5f).

For tephrite (155 samples), trachybasalt (314 samples), and phonotephrite (73 samples), the results summarised in Figure 6a-f and Table 1 showed that these rock types were mostly classified by the W&F diagram as alkali-basalt (66, 214, and 53 samples, respectively). Additionally, for these three rock types a significant number of samples (20, 49, and 9, respectively) were recognised as sub-alkaline basalt (Table 1). For tephrite samples, basanite/nephelinite also represented an important W&F classification (50 samples; 32.3%). A significant number of tephrite samples (18; 11.6%) were classified as trachyandesite in the W&F scheme.

Finally, we present the remaining three important rock types (basaltic andesite, basaltic trachyandesite, and trachydacite) according to the IUGS classification (Table 1; Figure 7a, c, e), but not included as such in the W&F diagram (Figure 7b, d, f). For basaltic andesite a very large number of samples (1239) were present in our database (Figure 7a). These were classified (Figure 7b) mainly as andesite/basalt (614 samples; 49.6%), sub-alkaline basalt (262 samples; 21.1%), and andesite (259 samples; 20.9%). Similarly, Figure 7c, 7d and Table 1 show that 392 samples

of basaltic trachyandesite were classified mainly as alkali-basalt (154 samples; 39.3%), andesite (94 samples; 23.9%), and sub-alkaline basalt (62 samples; 15.8%). Our final rock type trachydacite was represented by 69 samples (Figure 7e), which were classified mainly as trachyte and rhyodacite/dacite (36 samples–52.2% and 14 samples–20.3%, respectively; Figure 7f; Table 1).

In summary, the correct classification by the W&F diagram (Winchester & Floyd 1977) ranged from very low values of about 16.6% to reasonably high values of 76%. Alkali basalt, andesite and phonolite were best classified as such (about 73.8%-76%). The classification for subalkali basalt, basanite, trachyandesite, dacite, and rhyolite (with 16.6%-32.0%) was simply not acceptable. The remaining nine rock types included in the IUGS classification (Figures 5-7) also did not provide any one coherent rock name in the W&F scheme; the highest percentages ranged from 38.1% to 72.6%. Therefore, the wide use of this Nb/Y–Zr/TiO₂ diagram currently in practice is not particularly justified.

In order to obtain a totally unbiased evaluation of the W&F diagram, which is not particularly recommended for classifying arc rocks, we prepared a selected database by excluding all arc rocks and once again evaluated this diagram. The results are summarised in Table 2. To limit the space of this paper, no new diagrams are presented (in fact, all these samples are already included in Figures 1-7). As expected, the total number of samples of a given rock type generally decreased in Table 2 as compared to Table 1, and this decrease was more pronounced for sub-alkaline varieties than for alkaline types, but with exactly the same number of samples of picrite and foidite.

For alkali basalt samples, the correct classification by the W&F diagram increased from 73.8% to 84.5% (compare Tables 1 and 2), but for subalkali basalt it still remained unacceptably low (39.2%; Table 2). Basanite, trachyandesite, trachyte, and phonolite rocks were also not satisfactorily classified by the W&F diagram (correct classification of only about 30%, 36.5%, 36.9%, and 22.9%, respectively). Although andesite rock samples were fairly well classified as such (77.6%), neither dacite nor rhyolite samples had acceptable correct classifications (27.6% classified as ambiguous rhyodacite/dacite and 18.5% as rhyolite, respectively; Table 2). For rock types not included in the W&F classification (the second part of Table 2), the classification results remained practically the same as in Table 1.

From this work, therefore, an urgent need of adequate classification scheme for altered rocks is clearly established.

2. New classification diagram

Recently Hastie *et al.* (2007) recognised from an all together different approach (analysis of the chemical effects of alteration) that there is need for “a reliable way to classify rocks from the geological record”. They stated that none of the frequently used $\text{SiO}_2\text{-K}_2\text{O}$ (Peccerillo & Taylor 1976) and the IUGS recommended TAS (Le Bas *et al.* 1986) diagrams are appropriate for this purpose. They also argued that although Winchester & Floyd (1977) developed immobile element proxies for the TAS diagram, the need still existed for proxies of $\text{SiO}_2\text{-K}_2\text{O}$ diagram. Hastie *et al.* (2007) proposed the use of two proxy elements –Co for SiO_2 and Th for K_2O – in a new Co–Th bivariate diagram. Their contention that the W&F diagram is appropriate to replace

the TAS diagram has already been shown to be deficient in the present work (see the earlier subsection).

We now attempt to evaluate the new Co–Th diagram for classification purposes (Hastie *et al.* 2007). The same database as the one used for the evaluation of the W&F diagram was employed, with the addition of the Andes and the Mexican Volcanic Belt. We decided not to use the SiO₂–K₂O plot for processing our database because we wanted to strictly follow the IUGS recommendations for rock names (Le Bas *et al.* 1986; Le Bas 2000), in which the SiO₂–K₂O scheme (e.g., Peccerillo & Taylor 1976) was not included. Further, an evaluation of Co–Th diagram in terms of SiO₂–K₂O scheme was already provided by the original authors (Hastie *et al.* 2007) who concluded that success rates ranged only up to about 80% and, therefore, were not particularly high.

In this context, the subdivision of rock types in the Co–Th diagram is rather poor (Hastie *et al.* 2007). Actually only three independent regions for rock names were proposed in this diagram: basalt; an overlap region of basaltic andesite with andesite (referred to in this work as “basaltic andesite/andesite”); and an overlap region of dacite with rhyolite that also includes latite and trachyte (“dacite/rhyolite/latite/trachyte”). Each of these three regions is further subdivided into tholeiitic, calc-alkaline, and high-K₂O/shoshonitic. The correspondence of this three-fold subdivision of alkali-enrichment with the IUGS classification is difficult to establish. In the later scheme, only two-fold subdivision –subalkali and alkali– may be implicitly imagined for basalt. Certainly, there are many rock names depending on the contents of total alkalis at any given silica level. For example, at the same silica level the IUGS rock names can vary from basaltic andesite, basaltic trachyandesite, tephrite to phonolite, or from andesite, trachyandesite to phonolite (Le

Bas *et al.* 1986; Verma *et al.* 2002). For basaltic magmas, we could have arbitrarily decided the adjective subalkali to correspond to both tholeiitic and calc-alkaline divisions and alkali to high-K₂O/shoshonite. However, in our evaluation of the Co–Th diagram we decided not to enter into this oversimplification or assumptions regarding the IUGS nomenclature. Instead, separate identities of the three-fold subdivision of the Co–Th diagram were maintained.

Nevertheless, some clarification must be made for the correct classification (also called success rate) and consequently for the mis-classification. As an example, when a single basic rock type, such as basalt (IUGS nomenclature), was used for evaluation and if the rock name resulting from the Co–Th diagram for a basalt sample was not basalt, this particular sample represented an obvious mis-classification. Thus, for a group of samples of a given rock type it was easier to determine the “obvious mis-classification” or also called here as the “minimum mis-classification”. The later term will be used and highlighted in our following presentation. We will not explicitly refer to correct classification or success rate, because it will be more difficult to determine for the Co–Th diagram of Hastie *et al.* (2007) than for the old W&F diagram. In the former, the rock names are limited to only three separate fields.

Hastie *et al.* (2007) mentioned that their diagram is especially useful for the classification of arc rocks. Therefore, we maintained the identity of the tectonic setting or volcanic provinces used in this evaluation. These were: island arc; continental arc of the Andes; MVB; continental rift; and for some rock types, additionally ocean island. The results of our evaluation are presented in Figures 8-10 and summarised in Table 3. The “minimum mis-classification” is shown in **boldface** for rock names common to both classifications (IUGS and Co–Th diagram), or in *italics* for rock names not included in the IUGS classification (Table 3). Importantly, note

that this highlighting is reverse to that used for the W&F classification (Tables 1 and 2), in which the correct classification was shown in *italic boldface*.

For subalkali basalt from island arcs (237 samples with Co and Th data compiled in our database; Figure 8a), the minimum mis-classification amounted to 44 samples (18.6%) as basaltic andesite/andesite (Figure 8b; Table 3). A total of 193 remaining samples were, therefore, identified as the three varieties of basalt (see the columns of Thol., CA, and SHO in Table 3). The 8 shoshonite samples could also be considered as obvious mis-classification, but we decided not to discuss such finer details. For subalkali basalt (47 samples) from the Andes, 4 samples amounting to about 8% were obviously mis-classified as basaltic andesite or andesite. For subalkali basalt samples from the MVB and continental rift setting, no obvious mis-classification was observed. This does not mean, however, that the complement was the correct classification or success rate, we simply cannot clearly define it as discussed earlier.

The minimum mis-classification of alkali basalt samples (Figure 8c) was much less, with only a few samples mis-classified as basaltic andesite/andesite (Figure 8d; Table 3).

For the total of 206 samples of basaltic andesite from island arc setting (Figure 9a), a larger number of them (99 samples amounting to about 48.1%) represented the minimum mis-classification –93 samples (45.2%) as basalt and 6 samples (2.9%) as dacite/rhyolite/latite/trachyte (Figure 9b; Table 3). An even greater extent of minimum mis-classification was observed for this kind of rock samples from the Andes (Figure 9b), because 87 samples out of 138 (63.1%) were mis-classified as basalt and only the remaining 51 samples (36.9%) were shown to be basaltic andesite/andesite (Table 3). As for the Andes, the minimum

mis-classification for basaltic andesite from the MVB was also significant –about 50.4% (Table 3). For continental rift, this mis-classification was extremely high (about 93%), with 37 out of 40 samples mis-classified as basalt (Figure 9b).

Andesite (Figure 9c) can only be classified ambiguously as basaltic andesite/andesite (Figure 9d; Table 3). The minimum mis-classification of andesite samples from island arc, the Andes, and MVB as basalt was, respectively, about 12.8%, 23.7%, and 14.6% dacite/rhyolite/latite/trachyte (Figure 9d; Table 3).

Dacite, trachyte, trachydacite and rhyolite (Figure 9e) were considered together because they could only be classified as a group by the Co–Th diagram (Figure 9f). Such rock samples from island arc and the Andes showed 25.3% and 29.5% minimum mis-classification, respectively (Table 3). Similarly, the MVB and continental rift setting had these values of 30.1% and 24.0%, respectively.

The remaining results synthesised in Table 3 are for rock names (from the IUGS nomenclature) that were not included in the Co–Th diagram of Hastie *et al.* (2007). These rocks (Figure 10a-f) are mostly more alkalic than the earlier rocks already evaluated and presented in the first part of Table 3. In this second part of Table 3, combined rock types are arranged in the following order: ultrabasic (two groups; Figure 10a, b), basic (four groups; Figure 10c, d), and intermediate to acid (five groups; Figure 10e, f).

As expected, ultrabasic magmas are rather scarce in island arcs (no arc samples in Figure 10a; Table 3) and alkali-rich rocks, such as basaltic trachyandesite, trachybasalt, trachyte and

trachyandesite, are much less abundant in island arc setting than in other tectonic areas, including the MVB and continental arc of the Andes (Figure 10c, e; Table 3). Three out of 19 samples of such basic magmas from island arc were mis-classified as basaltic andesite/andesite (Figure 10d), and 9 out of 18 of intermediate magmas did so as basalt (Figure 10f).

Statistically significant number of samples of ultrabasic magmas having both Co and Th data were from continental rift and ocean island settings only (Figure 10a). They were mostly classified as shoshonitic basalt (Figure 10b; Table 3), with a considerable number (22 out of 86 samples; 25.6%) from continental rift also being classified as calc-alkaline basalt. The minimum mis-classification was only three samples (1.7%) mis-classified as basaltic andesite/andesite for the ocean island setting (Table 3). The minimum mis-classification for basic magmas is represented by the few samples that are classified as basaltic andesite/andesite in the Co–Th diagram (Figure 10d; Table 3). They represent only about 4.4% to 7.9%. Similarly, for intermediate to acid magmas, the mis-classified samples are those that result as basalt (Figure 10f; Table 3). However, these are much more numerous than those for the earlier two categories. The minimum mis-classification amounted to about 11.5%, 24.5%, 50.8%, and 52%, respectively, for the Andes, MVB, continental rift, and ocean island (Figure 10f; Table 3).

In summary, the minimum mis-classification for those rock samples that have names common in the IUGS scheme and Co–Th diagram of Hastie *et al.* (2007) ranged as follows: (i) two cases of low values of about 1.2% and 8%; (ii) five cases of intermediate values of about 12.8%-24.0%; (iii) seven cases of unacceptably high values of about 25.3%-93% (see the upper part of Table 3). For those rocks with names in the IUGS scheme and not in the Co–Th diagram

of Hastie *et al.* (2007), the minimum mis-classification were generally small (1.7%-11.5%), but with some greater values in the range of 24.5%-52%.

Discussion: the need of still newer classification diagram

On the basis of unacceptably low success rates for correct classification by the old W&F diagram and unacceptably high minimum mis-classification registered for the new Hastie *et al.* (2007) diagram, we can safely conclude that both diagrams are not recommended for the classification of altered rocks.

Concerning the Hastie *et al.* (2007) proposal of the Co–Th diagram for classification purposes, not much used so far by other researchers, it is not clear to us how one would actually employ it for classification purposes. The diagram is supposedly proposed for altered arc volcanic rocks. For older terrains where there is much need of classifying altered rocks, it seems to us that the user must then ascertain that the rocks actually belonged to an arc before using this diagram. If so, how is it to going to be done; by plate tectonic reconstructions or by discrimination diagrams? Plate reconstructions are largely model-based. Should we use discrimination diagrams? But, most of the existing discrimination diagrams do not properly work as shown by one of us in a companion paper (Verma 2009b). Then, should we use only the new discriminant function discrimination diagrams (Agrawal *et al.* 2008) based on natural log-ratio transformation of relatively immobile elements?

The other question that comes to one's mind is the following: Is there any use in knowing three classes of rock names, two of these classes being largely indeterminate (“basaltic andesite/andesite” and “dacite/rhyolite/latite/trachyte”), with two or more rock names in each of them? For fresh rocks there are a dozen of these individual names proposed by the IUGS (Le Bas *et al.* 1986; Le Bas 2000). The greater utility of this new Co–Th diagram probably resides in its capacity to assign tholeiitic, calc-alkaline and shoshonitic adjectives to basalt, basaltic andesite/andesite, and dacite/rhyolite/latite/trachyte.

From this discussion, it becomes clear to us that considerable amount of work will have to be done to reach the goal of altered rock classification system that would better match the IUGS scheme. Floyd & Winchester (1975, 1978) and Winchester & Floyd (1976, 1977) did try this goal even before the IUGS acceptance of the TAS diagram (Le Bas *et al.* 1986). However, the recent effort (Hastie *et al.* 2007) does not provide us a much needed working framework for classification of altered volcanic rocks from all tectonic settings. Unfortunately, we are probably far from having “a reliable way to classify rocks from the geological record”. Thus, new proposals for altered rock classification are very much required to solve the classification and nomenclature problems in earth sciences.

Statistical requirements for new proposals

A major problem with bivariate (such as the Co–Th diagram of Hastie *et al.* 2007), ternary diagrams (such as those evaluated by Verma 2009b) and even the element ratio/ratio plot (such as the Nb/Y–Zr/TiO₂ diagram of Winchester & Floyd 1977), is that they probably represent too few

chemical variables to correctly handle the multivariate problem of altered rock classification. The closure and constant sum problem is another factor that inhibits correct statistical treatment (Chayes 1960, 1965, 1978; Aitchison 1984, 1986, 1989). The representativeness of the initial datasets used for proposing these various diagrams may be another factor that affects their application. In such diagrams the dividing boundaries are also subjective, generally drawn by eye. An objective way to replace them by probability-based boundaries has already been proposed (Agrawal 1999) and practiced in discrimination diagrams (Agrawal *et al.* 2004, 2008; Verma *et al.* 2006; Agrawal & Verma, 2007).

For the future work we propose to follow the example of Verma *et al.* (2006) and Agrawal *et al.* (2008), who solved most of these problems for proposing new natural logarithm-ratio based discriminant function diagrams for tectonomagmatic discrimination using major- and trace-elements, respectively (Verma 2009b). In addition to the above problems, it is important to note that the statistical analysis also requires that the data be normally distributed without any discernible statistical contamination. Fortunately, this final problem can also be solved by the methodology proposed by Barnett & Lewis (1994) and considerably improved recently by Verma & Quiroz-Ruiz (2006a, 2006b, 2008), Verma *et al.* (2008b), and Verma (2009c). Such a statistical procedure of discordancy tests has been already extensively employed by numerous researchers (e.g., Castrellon-Urbe *et al.* 2008; Díaz-González *et al.* 2008; Jafarzadeh & Hosseini-Barzi 2008; Nagarajan *et al.* 2008; Obeidat *et al.* 2008; Palabiyik & Serpen 2008; Vargas-Rodríguez *et al.* 2008; Vattuone *et al.* 2008; Armstrong-Altrin 2009; Marroquín-Guerra *et al.* 2009; Pandarinath 2009a, 2009b; Gómez-Arias *et al.* 2009; González-Márquez & Hansen 2009; González-Ramírez *et al.* 2009; Madhavaraju & Lee 2009; Ostrooumov *et al.* 2009; Rodríguez-Ríos & Torres-Aguilera, 2009; Verma *et al.* 2009a, 2009b). Therefore, it will not be difficult to

incorporate this procedure in the proposal of new classification diagrams for altered volcanic rocks. Such combined methodology of extensive database, discordant outlier tests, and linear discriminant analysis could therefore be easily extended for the proposal of much needed new diagrams for plutonic rocks as well. This work is currently under progress.

Conclusions

From the statistical evaluations using fresh rocks presented in this work, we conclude that none of the existing classification diagrams works well for altered volcanic rocks. Therefore, an urgent need exists to explore this field of earth sciences and fulfil the much needed altered rock classification system. Significant progress has been achieved in the statistical methods of identifying discordant outliers and linear discriminant analysis based on natural logarithm-ratio transformations, whose application would facilitate the proposal of new classification diagrams consistent with the IUGS nomenclature.

Acknowledgements

R. Rodríguez-Ríos is grateful to the Universidad Autónoma de San Luis Potosí for granting him a sabbatical leave to work at the Centro de Investigación en Energía (Universidad Nacional Autónoma de México). R. González-Ramírez expresses her gratefulness to the Secretaría de Educación Pública for the permission to carry out doctoral studies in Universidad Nacional Autónoma de México. We are grateful to Samuele Agostini, an anonymous individual, and the

editor Erdin Bozkurt, all of them, while highly appreciating our work, provided suggestions that helped improve our presentation.

References

- AGRAWAL, S., 1999. Geochemical discrimination diagrams: a simple way of replacing eye-fitted boundaries with probability based classifier surfaces. *Journal of the Geological Society of India* **54**, 335-346.
- AGRAWAL, S. & VERMA, S.P. 2007. Comment on "Tectonic classification of basalts with classification trees" by Pieter Vermeesch (2006). *Geochimica et Cosmochimica Acta* **71**, 3388-3390.
- AGRAWAL, S., GUEVARA, M. & VERMA, S.P. 2004. Discriminant analysis applied to establish major-element field boundaries for tectonic varieties of basic rocks. *International Geology Review* **46**, 575-594.
- AGRAWAL, S., GUEVARA, M. & VERMA, S.P. 2008. Tectonic discrimination of basic and ultrabasic rocks through log-transformed ratios of immobile trace elements. *International Geology Review* **50**, 1057-1079.
- AHMAD, T., TANAKA, T., SACHAN, H.K., ASAHARA, Y., ISLAM, R. & KHANNA, P.P. 2008. Geochemical and isotopic constraints on the age and origin of the Nidar Ophiolitic Complex, Ladakh, India: Implications for the Neo-Tethyan subduction along the Indus suture zone. *Tectonophysics* **451**, 206-224.
- AITCHISON, J. 1984. Reducing the dimensionality of compositional data set. *Mathematical Geology* **16**, 617-635.
- AITCHISON, J. 1986. The statistical analysis of compositional data. Chapman and Hall, London New York, pp. 416.
- AITCHISON, J. 1989. Measures of location of compositional data sets. *Mathematical Geology* **21**, 787-790.
- Armstrong-Altrin, J.S. 2009. Provenance of felsic beach sands from three areas of Mexico. *Revista Mexicana de Ciencias Geológicas* (in press).
- BAGCI, U., PARLAK, O. & HOCK, V. 2008. Geochemistry and tectonic environment of diverse magma generations forming the crustal units of the Kizildag (Hatay) ophiolite, southern Turkey. *Turkish Journal of Earth Sciences* **17**, 43-71.
- BARNETT, V. & LEWIS, T. 1994. Outliers in statistical data. John Wiley & Sons, Chichester, pp. 584.
- CASTRELLON-URIBE, J., CUEVAS-ARTEAGA, C. & TRUJILLO-ESTRADA, A. 2008. Corrosion monitoring of stainless steel 304L in lithium bromide aqueous solution using transmittance optical detection technique. *Optics and Lasers in Engineering* **46**, 469-476.

- CHAYES, F. 1960. On correlation between variables of constant sum. *Journal of Geophysical Research* **65**, 4185-4193.
- CHAYES, F., 1965. Classification in a ternary-diagram by means of discriminant functions. *American Mineralogist* **50**, 1618-1633.
- CHAYES, F. 1978. Ratio correlation. A manual for students of petrology and geochemistry. The University of Chicago Press, Chicago and London, p. 99.
- DÍAZ-GONZÁLEZ, L., SANTOYO, E. & REYES-REYES J. 2008. Tres nuevos geotermómetros mejorados de Na/K usando herramientas computacionales y geoquímicas: aplicación a la predicción de temperaturas de sistemas geotérmicos. *Revista Mexicana de Ciencias Geológicas* **25**, 465-482.
- FLOYD, P.A. & WINCHESTER, J.A. 1975. Magma type and tectonic setting discrimination using immobile elements. *Earth and Planetary Science Letters* **27**, 211-218.
- FLOYD, P.A. & WINCHESTER, J.A. 1978. Identification and discrimination of altered and meta-morphosed volcanic rocks using immobile elements. *Chemical Geology* **21**, 291-306.
- FOURNIER, R.O. & POTTER II, R.W. 1982. A revised and expanded silica (quartz) geothermometer. *Geothermal Resources Council Bulletin* **11**, 3-12.
- GLADKOCHUB, D.P., MAZUKABZOV, A.M., DONSKAYA, T.V., DE WAELE, B., STANEVICH, M. & PISAREVSKY, S.A. 2008. The age and origin of volcanics in the Riphean section of the Siberian craton (western Baikal area). *Russian Geology and Geophysics* **49**, 749-758.
- GOKTEN, E. & FLOYD, P.A. 2007. Stratigraphy and geochemistry of pillow basalts within the ophiolitic melange of the Izmir-Ankara-Erzincan suture zone: implications for the geotectonic character of the northern branch of Neotethys. *International Journal of Earth Sciences* **96**, 725-741.
- GÓMEZ-ARIAS, E., ANDAVERDE, J., SANTOYO, E. & URQUIZA, G. 2009. Determinación de la viscosidad y su incertidumbre en fluidos de perforación usados en la construcción de pozos geotérmicos: aplicación en el campo de Los Humeros, Puebla, México. *Revista Mexicana de Ciencias Geológicas* **26**, 516-529.
- GONZÁLEZ-MÁRQUEZ, L.C. & HANSEN, A.M. 2009. Adsorción y mineralización de atrazina y relación con parámetros de suelos del DR. 063 Guasave, Sinaloa. *Revista Mexicana de Ciencias Geológicas* **26**, (in press).
- GONZÁLEZ-RAMÍREZ, R., DÍAZ-GONZÁLEZ, L. & VERMA, S.P. 2009. Eficiencia relativa de las 33 pruebas de discordancia para valores desviados basada en datos geoquímicos de materiales de referencia. *Revista Mexicana de Ciencias Geológicas* **26**, 501-515.
- GURSU, S. 2008. Petrogenetic and tectonic significance of rift-related pre-Early Cambrian mafic dikes, Central Taurides, Turkey. *International Geology Review* **50**, 895-913.
- HASTIE, A.R., KERR, A.C., PEARCE, J.A. & MITCHELL, S.F. 2007. Classification of altered volcanic island rocks using immobile trace elements: development of the Th-Co discrimination diagram. *Journal of Petrology* **48**, 2341-2357.

- JAFARZADEH, M. & HOSSEINI-BARZI, M. 2008. Petrography and geochemistry of Ahwaz Sandstone Member of Asmari Formation, Zagros, Iran: implications on provenance and tectonic setting. *Revista Mexicana de Ciencias Geológicas* **25**, 247-260.
- KADER, S., ONEN-HALL, A.P., AYDIN, S.N., YAKICIER, C., AKARSU, N. & TUNCER, M. 2008. Environmental effect and genetic influence: a regional cancer predisposition survey in the Zonguldak region of Northwest Turkey. *Environmental Geology* **54**, 391-409.
- KESKIN, M., GENÇ, S.C. & TUYSUZ, O. 2008. Petrology and geochemistry of post-collisional Middle Eocene volcanic units in North-Central Turkey: Evidence for magma generation by slab breakoff following the closure of the Northern Neotethys Ocean. *Lithos* **104**, 267-305.
- KALMAR, J. & KOVACS-PALFFY, P. 2008. Geochemical study of leptinites from Stejera (Romania). *Carpathian Journal of Earth and Environmental Sciences* **3**, 49-64.
- KAYGUSUZ, A., SIEBEL, W., SEN, C. & SATIR, M. 2008. Petrochemistry and petrology of I-type granitoids in an arc setting: the composite Torul pluton, Eastern Pontides, NE Turkey. *International Journal of Earth Sciences* **97**, 739-764.
- LE BAS, M.J. 2000. IUGS reclassification of the high-Mg and picritic volcanic rocks. *Journal of Petrology* **41**, 1467-1470.
- LE BAS, M.J., LE MAITRE, R.W., STRECKEISEN, A. & ZANETTIN, B. 1986. A chemical classification of volcanic rocks based on the total alkali-silica diagram. *Journal of Petrology* **27**, 745-750.
- LE MAITRE, R.W., STRECKEISEN, A., ZANETTIN, B., LE BAS, M.J., BONIN, B., BATEMAN, P., BELLINI, G., DUDEK, A., SCHMID, R., SORESENSEN, H. & WOOLLEY, A.R. 2002. Igneous rocks. A classification and glossary of terms: recommendations of the International Union of Geological Sciences Subcommittee of the Systematics of Igneous Rocks. Cambridge University Press, Cambridge, pp. 236.
- MADHAVARAJU, J. & LEE, Y.I. 2009. Geochemistry of the Dalmiapuram Formation of the Uttatur Group (Early Cretaceous), Cauvery basin, southeastern India: Implications on Provenance and Paleoredox conditions. *Revista Mexicana de Ciencias Geológicas* **26**, 380-394.
- MARROQUÍN-GUERRA, S.G., VELASCO-TAPIA, F. & DÍAZ-GONZÁLEZ, L. 2009. Evaluación estadística de Materiales de Referencia Geoquímica del Centre de Recherches Pétrographiques et Géochimiques (Francia) aplicando un esquema de detección y eliminación de valores desviados y su posible aplicación en el control de calidad de datos geoquímicas. *Revista Mexicana de Ciencias Geológicas* **26**, 530-542.
- MONDAL, M.E.A., CHANDRA, R. & AHMAD, T. 2008. Precambrian mafic magmatism in Bundelkhand craton. *Journal of the Geological Society of India* **72**, 113-122.
- NAGARAJAN, R., SIAL, A.N., ARMSTRONG-ALTRIN, J.S., MADHAVARAJU, J. & NAGENDRA, R. 2008. Carbon and oxygen isotope geochemistry of Neoproterozoic limestones of the Shahabad Formation, Bhima basin, Karnataka, southern India. *Revista Mexicana de Ciencias Geológicas* **25**, 225-235.

- NARDI, L.V.S., PLA-CID, J., BITENCOURT, M.D. & STABEL, L.Z. 2008. Geochemistry and petrogenesis of post-collisional ultrapotassic syenites and granites from southernmost Brazil: the Piquiri Syenite Massif. *Anais da Academia Brasileira de Ciencias* **80**, 353-371.
- OBEIDAT, M.M., AHMAD, F.Y., HAMOURI, N.A.A., MASSADEH, A.M. & ATHAMNEH, F.S. 2008. Assessment of Nitrate Contamination of Karst Springs, Bani Kanana, Northern Jordan. *Revista Mexicana de Ciencias Geológicas* **25**, 426-437.
- Ostrooumov, M., Taran, Y., Arellano-Jiménez, M., Ponce, A. & Reyes-Gasga, J. 2009. Colimaite, K3VS4 – a new potassium-vanadium sulfide mineral from the Colima volcano, State of Colima (Mexico). *Revista Mexicana de Ciencias Geológicas* (in press).
- PALABIYIK, Y. & SERPEN, U. 2008. Geochemical Assessment of Simav Geothermal Field, Turkey. *Revista Mexicana de Ciencias Geológicas* **25**, 408-425.
- PANDARINATH, K. 2009a. Clay minerals in SW Indian continental shelf sediments cores as indicators of provenance and paleomonsoonal conditions: a statistical approach. *International Geology Review* (in press).
- PANDARINATH, K. 2009b. Evaluation of geochemical sedimentary reference materials of the Geological Society of Japan (GSJ) by an objective outlier rejection statistical method. *Revista Mexicana de Ciencias Geológicas* (in press).
- PANDARINATH, K., PUSHPARANI, D.E., TORRES-ALVARADO, I. & VERMA, S.P. 2006. X-ray diffraction analysis of hydrothermal minerals from the Los Azufres geothermal system, Mexico. *International Geology Review* **48**, 174-190.
- PANDARINATH, K., DUSKI, P., TORRES-ALVARADO, I.S. & VERMA, S.P. 2008. Element mobility during the hydrothermal alteration of rhyolitic rocks of the Los Azufres geothermal field, Mexico. *Geothermics* **37**, 53-72.
- PECCERILLO, A. & TAYLOR, S.R., 1976. Geochemistry of Eocene calc-alkaline volcanic rocks from the Kastamonu area, Northern Turkey. *Contributions to Mineralogy and Petrology* **58**, 63-81.
- Rodriguez-Rios, R. & Torres-Aguilera, J.M. 2009. Evolución petrológica y geoquímica de un vulcanismo bimodal oligocénico en el Campo Volcánico de San Luis Potosí (México). *Revista Mexicana de Ciencias Geológicas* (in press).
- SHEKHAWAT, L.S., PANDIT, M.K. & JOSHI, D.W. 2007. Geology and geochemistry of palaeoproterozoic low-grade metabasic volcanic rocks from Salumber area, Aravalli Supergroup, NW India. *Journal of Earth System Science* **116**, 511-524.
- SHEIH, H.C. & MELLUSO, L. 2008. The Mount Pavagadh volcanic suite, Deccan Traps: Geochemical stratigraphy and magmatic evolution. *Journal of Asian Earth Sciences* **32**, 5-21.
- TORRES-ALVARADO, I.S. 2002. Chemical equilibrium in hydrothermal systems: the case of Los Azufres geothermal field, Mexico. *International Geology Review* **44**, 639-652.
- TORRES-ALVARADO, I.S., PANDARINATH, K., VERMA, S.P. & DULSKI, P. 2007. Mineralogical and geochemical effects due to hydrothermal alteration in the Los Azufres geothermal field, Mexico. *Revista Mexicana de Ciencias Geológicas* **24**, 15-24.

- VARGAS-RODRÍGUEZ, Y.M., GÓMEZ-VIDALES, V., VÁZQUEZ-LABASTIDA E., GARCÍA-BORQUEZ, A., AGUILAR-SAHAGUN, G., MURRIETA-SÁNCHEZ, H. & SALMON, M. 2008. Caracterización espectroscópica, química y morfológica y propiedades superficiales de una montmorillonita Mexicana. *Revista Mexicana de Ciencias Geológicas* **25**, 135-144.
- VASCONCELOS-F., M., VERMA, S.P. & RODRÍGUEZ-G., J.F. 1998. Discriminación tectónica: nuevo diagrama Nb-Ba para arcos continentales, arcos insulares, "rifts" e islas oceánicas en rocas máficas. *Boletín de la Sociedad Española de Mineralogía* **21**, 129-146.
- VASCONCELOS-F., M., VERMA, S.P. & VARGAS-B., R.C. 2001. Diagrama Ti-V: una nueva propuesta de discriminación para magmas básicos en cinco ambientes tectónicos. *Revista Mexicana de Ciencias Geológicas* **18**, 162-174.
- VATTUONE, M.E., LEAL, P.R., CROSTA, S., BERBEGLIA, Y., GALLEGOS, E. & MARTÍNEZ-DOPICO, C. 2008. Paragénesis de zeolitas alcalinas en un afloramiento de basaltos olivínicos amigdaloides de Junín de Los Andes, Neuquén, Patagonia, Argentina. *Revista Mexicana de Ciencias Geológicas* **25**, 483-493.
- Verma, M.P., 2000. Revised quartz solubility temperature dependence equation along the water-vapor saturation curve. Proceedings World Geothermal Congress, Kyushu-Tohoku, Japan, pp. 1927-1932.
- VERMA, S.P. 1997. Estado actual de los diagramas de clasificación magmática y de discriminación tectónica. *Actas INAGEQ* **3**, 49-78.
- VERMA, S.P. 2000. Geochemistry of the subducting Cocos plate and the origin of subduction-unrelated mafic volcanism at the volcanic front of the central Mexican Volcanic Belt. In: DELGADO-GRANADOS, H., AGUIRRE-DÍAZ, G. & STOCK, J. M. (eds), *Cenozoic tectonics and volcanism of Mexico*. Geological Society of America, Geological Society of America Special paper 334, 195-222.
- VERMA, S.P. 2002. Absence of Cocos plate subduction-related basic volcanism in southern Mexico: a unique case on Earth? *Geology* **30**, 1095-1098.
- VERMA, S.P. 2004. Solely extension-related origin of the eastern to west-central Mexican Volcanic Belt (Mexico) from partial melting inversion model. *Current Science* **86**, 713-719.
- VERMA, S.P. 2006. Extension related origin of magmas from a garnet-bearing source in the Los Tuxtlas volcanic field, Mexico. *International Journal of Earth Sciences* **95**, 871-901.
- VERMA, S.P. 2009a. Continental rift setting for the central part of the Mexican Volcanic Belt: A Statistical Approach. *The Open Geology Journal* **3**, 8-29.
- VERMA, S.P. 2009b. Statistical Evaluation of Bivariate, Ternary and Discriminant Function Tectonomagmatic Discrimination Diagrams. *Turkish Journal of Earth Sciences* (in press).
- VERMA, S.P. 2009c. Evaluation of polynomial regression models for the Student t and Fisher F critical values, the best interpolation equations from double and triple natural logarithm transformation of degrees of freedom up to 1000, and their applications to quality control in science and engineering. *Revista Mexicana de Ciencias Geológicas* **26**, 79-92.

- VERMA, S.P. & AGUILAR-Y-VARGAS, V.H. 1988. Bulk chemical composition of magmas in the Mexican Volcanic Belt (Mexico) and inapplicability of generalized arc-models. *Chemie der Erde* **48**, 203-221.
- 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**, 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**, 302-319.
- VERMA, S.P. & QUIROZ-RUIZ, A. 2008. Critical values for 33 discordancy test variants for outliers in normal samples for very large sizes of 1,000 to 30,000 and evaluation of different regression models for the interpolation and extrapolation of critical values. *Revista Mexicana de Ciencias Geológicas* **25**, 369-381.
- VERMA, S.P. & SANTOYO, E. 1997. New improved equations for Na/K, Na/Li and SiO₂ geothermometers by outlier detection and rejection. *Journal of Volcanology and Geothermal Research* **79**, 9-23.
- VERMA, S.P., GUEVARA, M. & AGRAWAL, S. 2006. Discriminating four tectonic settings: five new geochemical diagrams for basic and ultrabasic volcanic rocks based on log-ratio transformation of major-element data. *Journal of Earth System Science* **115**, 485-528.
- VERMA, S.P., TORRES-ALVARADO, I.S. & SOTELO-RODRÍGUEZ, Z.T. 2002. SINCLAS: standard igneous norm and volcanic rock classification system. *Computers & Geosciences* **28**, 711-715.
- VERMA, S.P., TORRES-ALVARADO, I.S. & VELASCO-TAPIA, F. 2003. A revised CIPW norm. *Schweizerische Mineralogische und Petrographische Mitteilungen* **83**, 197-216.
- VERMA, S.P., GUEVARA, M. & AGRAWAL, S. 2006. Discriminating four tectonic settings: five new geochemical diagrams for basic and ultrabasic volcanic rocks based on log-ratio transformation of major-element data. *Journal of Earth System Science* **115**, 485-528.
- VERMA, S.P., PANDARINATH, K. & SANTOYO, E. 2008a. SolGeo: A new computer program for solute geothermometers and its application to Mexican geothermal fields. *Geothermic* **37**, 597-621.
- VERMA, S.P. & QUIROZ-RUIZ, A. & DÍAZ-GONZÁLEZ, L. 2008b. Critical values for 33 discordancy test variants for outliers in normal samples up to sizes 1000, and applications in quality control in Earth Sciences. *Revista Mexicana de Ciencias Geológicas* **25**, 82-96.
- VERMA, S.P., DÍAZ-GONZÁLEZ, L. & GONZÁLEZ-RAMÍREZ, R. 2009a. Relative efficiency of single-outlier discordancy tests for processing geochemical data on reference materials and application to instrumental calibrations by a weighted least-squares linear regression model. *Geostandards and Geoanalytical Research* **33**, 29-49.
- VERMA, S.P., PANDARINATH, K., VELASCO-TAPIA, F. & RODRÍGUEZ-RÍOS, R. 2009b. Evaluation of the odd-even effect in limits of detection for electron microprobe analysis of natural minerals. *Analytica Chimica Acta* **638**, 126-132.

- WANG, X.C., LI, X.H., LI, W.X., LI, Z.X., LIU, Y., YANG, Y.H., LIANG, X.R. & TU, X.L. 2008. The Bikou basalts in the northwestern Yangtze block, South China: Remnants of 820-810 Ma continental flood basalts? *Geological Society of America Bulletin* **120**, 1478-1492.
- WINCHESTER, J.A. & FLOYD, P.A. 1976. Geochemical magma type discrimination: application to altered and metamorphosed basic igneous rocks. *Earth and Planetary Science Letters* **28**, 459-469.
- WINCHESTER, J.A. & FLOYD, P.A. 1977. Geochemical discrimination of different magma series and their differentiation products using immobile elements. *Chemical Geology* **20**, 325-343.
- YIGITBAS, E., WINCHESTER, J.A. & OTTLEY, C.J. 2008. The geochemistry and setting of the Demirci paragneisses of the Sunnice Massif, NW Turkey. *Turkish Journal of Earth Sciences* **17**, 421-431.
- ZHENG, Y., GU, L., TANG, X., LI, C., LIU, S., & CHANGZHI, W. 2008. Geological and geochemical signature of sea-floor alteration rocks of the highly metamorphosed Hongtoushan massive sulfide deposit, Liaoning. *Acta Petrologica Sinica* **24**, 1928-1936.

Figure legends

Figure 1. Statistical evaluation of the Nb/Y–Zr/TiO₂ diagram of Winchester & Floyd (1977) – called the W&F diagram in this work– in reference to the TAS (total alkalis *versus* silica) diagram (Le Bas *et al.* 1986; Verma *et al.* 2002) of the IUGS classification scheme, using basaltic rocks from our database. Note also that the IUGS recommendation to use adjusted data in the TAS diagram was strictly followed (Verma *et al.* 2002). The field names in the TAS diagram, viz., (a) and (c), are: PB–microbasalt; B–basalt; BA–basaltic andesite; BSN–basanite; TEP–tephrite; TB–trachybasalt; BTA–basaltic trachyandesite; TA–trachy andesite; FOI–foiidite; PHT–phonotephrite; and TPH–tephriphonolite. Only part of the TAS diagram is shown. Other TAS rock names not included in this diagram, but present in some later Figures are: PH–phonolite; A–andesite; D–dacite; TD–trachydacite; T–trachyte; R–rhyolite. Similarly, for the W&F diagram, viz., (b) and (d), the field names are: B,Alk–alkali-basalt; B,Sub-Alk–Sub-alkaline basalt; BSN/NPH–basanite/nephelinite; B/A–basalt/andesite; A–andesite; TA–trachyandesite; T–trachyte; PH–phonolite; COM/PAN–comendite/pantellerite; RD/D–rhyodacite/dacite; and R–rhyolite. The same symbols are used in the W&F diagram as in the corresponding TAS diagram, i.e., the symbols are the same in the (a) and (c) pairs of diagrams and (b) and (d) pairs. (a) Alkali basalt (650) samples according to the TAS diagram; (b) Subalkali basalt (1200) samples according to the TAS diagram; (c) The same alkali basalt (650) samples of the TAS diagram plotted in the W&F diagram; and (d) The same subalkali basalt (1200) samples of the TAS diagram plotted in the W&F diagram.

Figure 2. Statistical evaluation of the W&F diagram in reference to the TAS diagram using basanite and andesite rocks from our database. See Figure 1 for more explanation. (a)

similar to microbasalt and 106 similar to alkali basalt) samples according to the TAS diagram; (b) The same picrite (151) samples of the TAS diagram plotted in the W&F diagram; (c) Foidite (118) samples according to the TAS diagram; (d) The same foidite (118) samples of the TAS diagram plotted in the W&F diagram; (e) Picrobasalt (30) samples according to the TAS diagram; and (f) The same picrobasalt (30) samples of the TAS diagram plotted in the W&F diagram.

Figure 6. Statistical evaluation of the W&F diagram in reference to the TAS diagram using tephrite, trachybasalt and phonotephrite rocks from our database. See Figure 1 for more explanation. (a) Tephrite (155) samples according to the TAS diagram; (b) The same tephrite (155) samples of the TAS diagram plotted in the W&F diagram; (c) Trachybasalt (314) samples according to the TAS diagram; (d) The same trachybasalt (314) samples of the TAS diagram plotted in the W&F diagram; (e) phonotephrite (73) samples according to the TAS diagram; and (f) The same phonotephrite (73) samples of the TAS diagram plotted in the W&F diagram.

Figure 7. Statistical evaluation of the W&F diagram in reference to the TAS diagram using basaltic andesite, basaltic trachyandesite and trachydacite rocks from our database. See Figure 1 for more explanation. (a) Basaltic andesite (1239) samples according to the TAS diagram; (b) The same basaltic andesite (1239) samples of the TAS diagram plotted in the W&F diagram; (c) Basaltic trachyandesite (392) samples according to the TAS diagram; (d) The same basaltic trachyandesite (392) samples of the TAS diagram plotted in the W&F diagram; (e) Trachydacite (69) samples according to the TAS diagram; and (f) The same trachydacite (69) samples of the TAS diagram plotted in the W&F diagram.

Figure 8. Statistical evaluation of the Co–Th diagram of Hastie *et al.* (2007) in reference to the TAS (total alkalis *versus* silica) diagram (Le Bas *et al.* 1986; Verma *et al.* 2002) of the IUGS classification scheme, using basaltic rocks from our database. Note separate identity of tectonic settings was maintained in this and later diagrams. See Figure 1 for more explanation on the TAS diagram. The rock type abbreviations in the Co–Th diagram are: B–basalt; BA/A–basaltic andesite/andesite; D/R*–dacite/rhyolite/latite/trachyte; IAT–island arc tholeiite; CA–calc-alkaline; and H•K and SHO–high•K and shoshonite. See also Table 3. The same symbols are used in the Co–Th diagram as in the corresponding TAS diagram. Furthermore, they are explained as inset in (a). (a) Subalkali basalt (437 samples; 237 from island arcs, 47 from the Andes, 61 from the Mexican Volcanic Belt–MVB, and 92 from continental rifts) samples according to the TAS diagram; (b) Alkali basalt (167 samples; 22 from island arcs, 12 from the Andes, 48 from the MVB, and 85 from continental rifts) samples according to the TAS diagram; (c) The same subalkali basalt (437) samples of the TAS diagram plotted in the Co–Th diagram; and (d) The same alkali basalt (167) samples of the TAS diagram plotted in the Co–Th diagram.

Figure 9. Statistical evaluation of the Co–Th diagram of Hastie *et al.* (2007) in reference to the TAS (total alkalis *versus* silica) diagram (Le Bas *et al.* 1986; Verma *et al.* 2002), using basaltic andesite, andesite, dacite, trachyte, trachydacite and rhyolite rocks from our database. See Figures 1 and 8 and Table 3 for more explanation. (a) Basaltic andesite (513 samples; 206 from island arcs, 138 from the Andes, 129 from the MVB, and 40 from continental rifts) samples according to the TAS diagram; (b) The same basaltic andesite (513) samples of the TAS diagram plotted in the Co–Th diagram; (c) Andesite (456 samples; 109 from island arcs, 59 from the Andes, and 288 from the MVB) samples

according to the TAS diagram; (d) The same andesite (456) samples of the TAS diagram plotted in the Co–Th diagram; (e) Diverse acid rock types (436 samples; 83 from island arcs, 112 from the Andes, 166 from the MVB, and 75 from continental rifts) samples according to the TAS diagram; and (f) The same (436) samples of the TAS diagram plotted in the Co–Th diagram.

Figure 10. Statistical evaluation of the Co–Th diagram of Hastie *et al.* (2007) in reference to the TAS (total alkalis *versus* silica) diagram (Le Bas *et al.* 1986; Verma *et al.* 2002), using diverse rocks (other than those used in Figures 8 and 9) from our database. See Figures 1 and 8 and Table 3 for more explanation. (a) Different combinations of rock types (see Table 3 for abbreviations; 86 from continental rifts and 180 from ocean islands) samples according to the TAS diagram; (b) The same samples of the TAS diagram (a) plotted in the Co–Th diagram; (c) Different combinations of rock types (see Table 3 for abbreviations; 19 samples from island arcs, 51 from the Andes, 45 from the MVB, and 134 from continental rifts) according to the TAS diagram; (d) The same samples of the TAS diagram (c) plotted in the Co–Th diagram; (e) Different combinations of rock types (see Table 3 for abbreviations; 18 from island arcs, 70 from the Andes, 94 from the MVB, 67 from continental rifts, and 29 from ocean islands) samples according to the TAS diagram; and (f) The same samples of the TAS diagram (e) plotted in the Co–Th diagram.

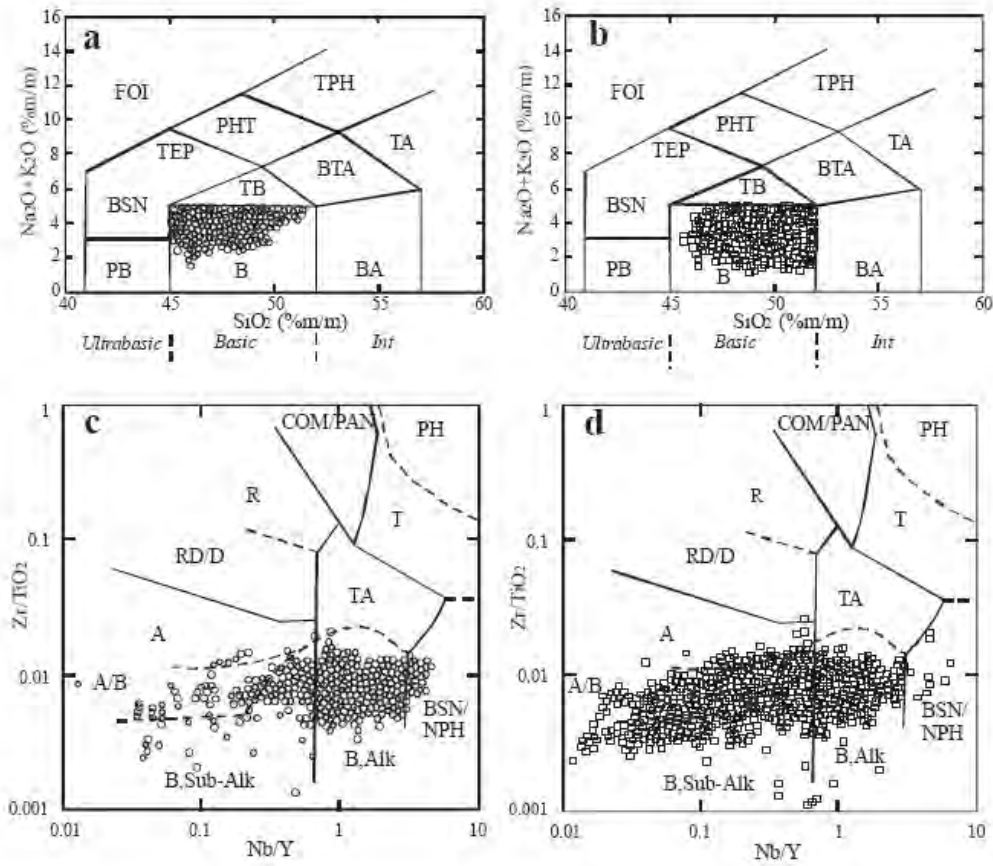


Figura 1

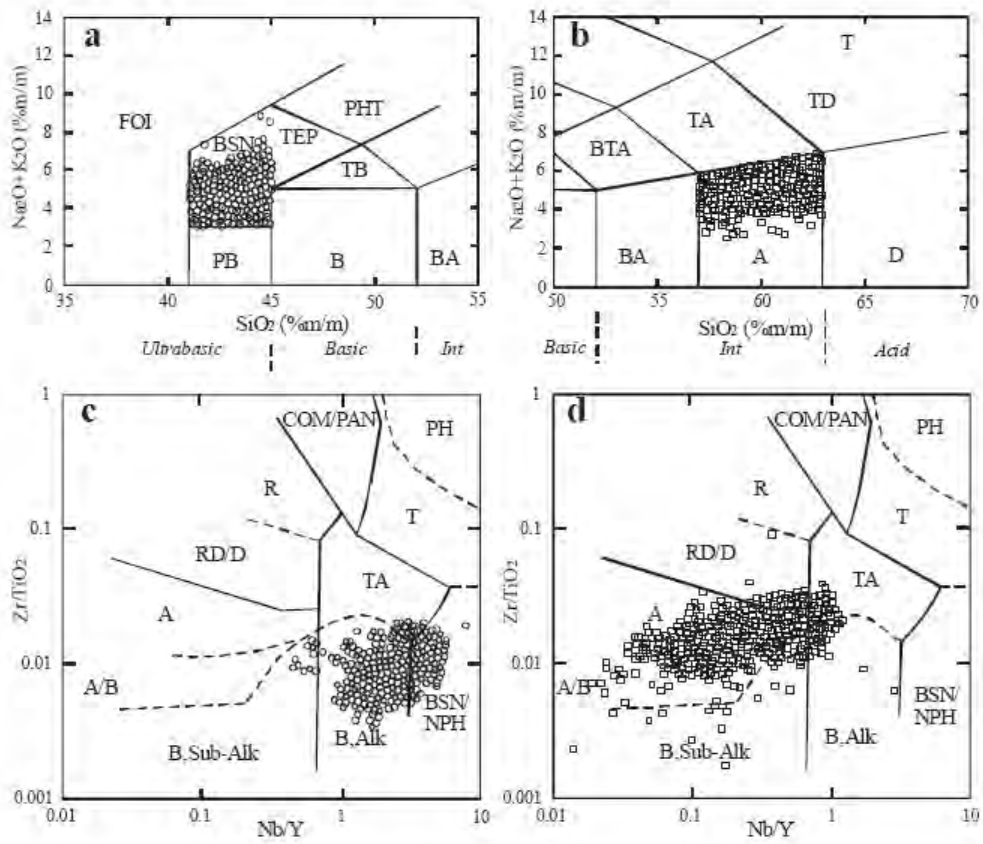


Figura 2

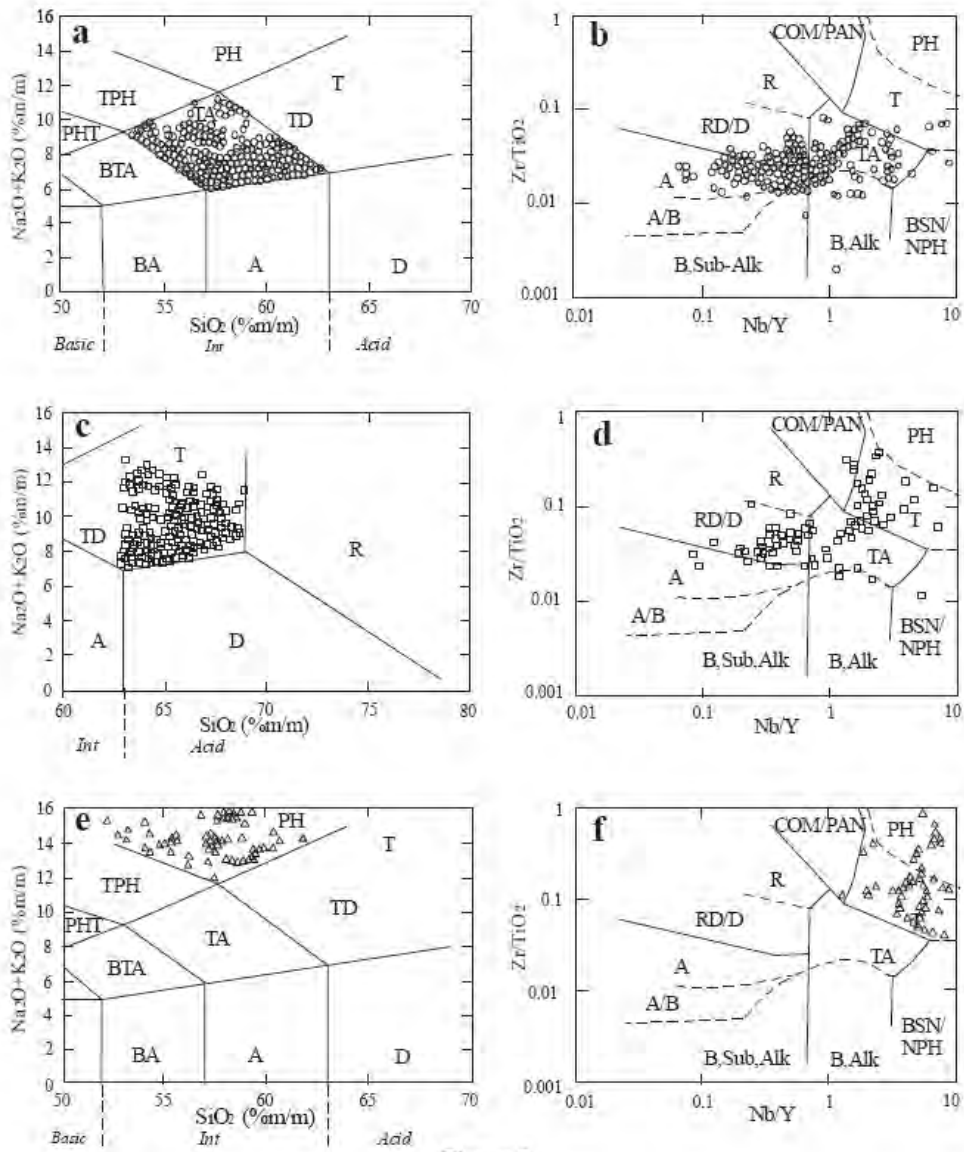


Figura 3

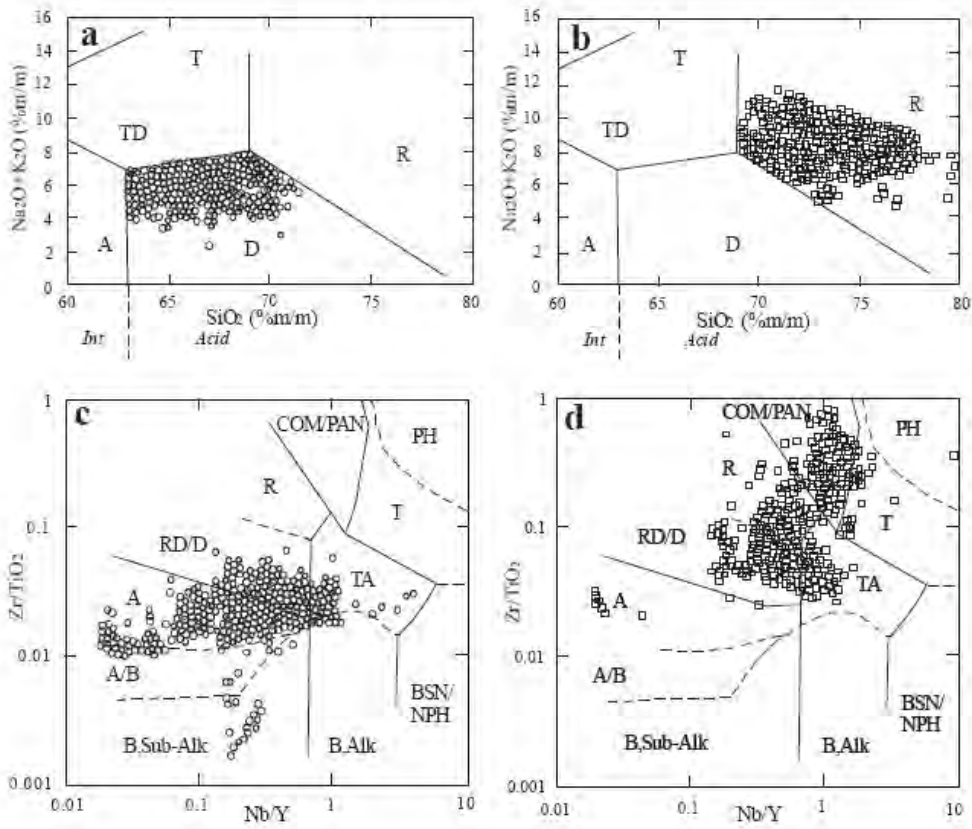


Figura 4

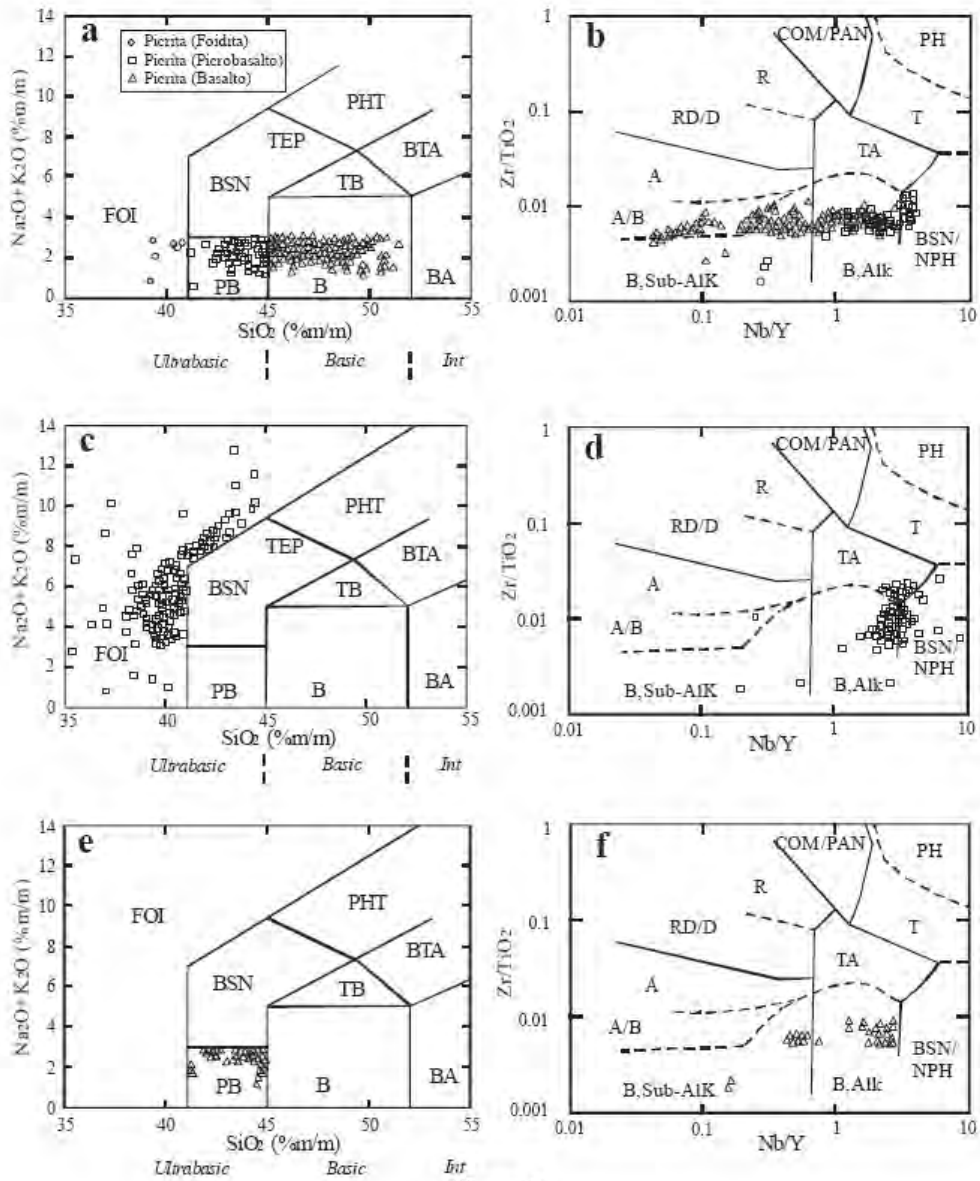


Figura 5

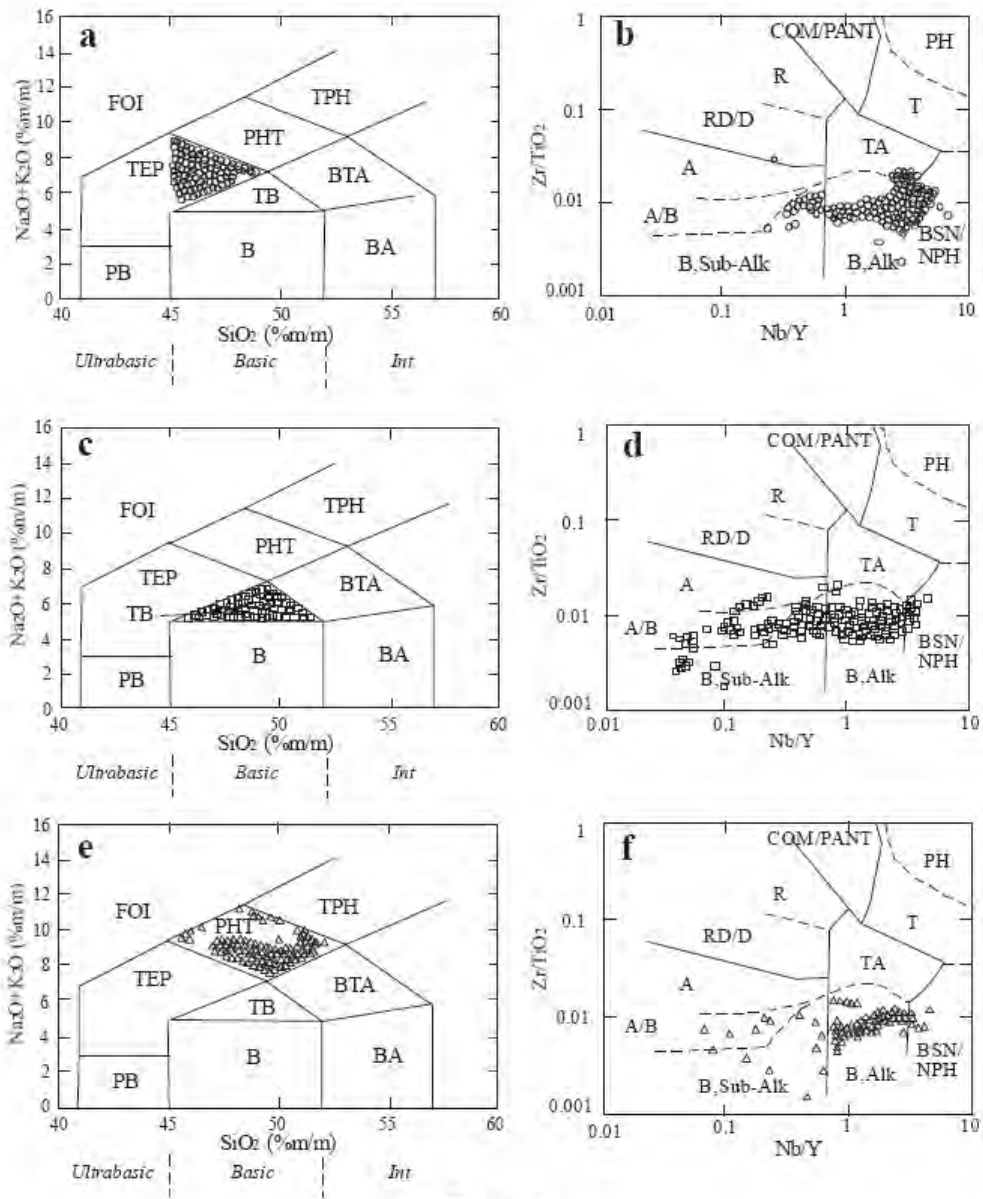


Figura 6

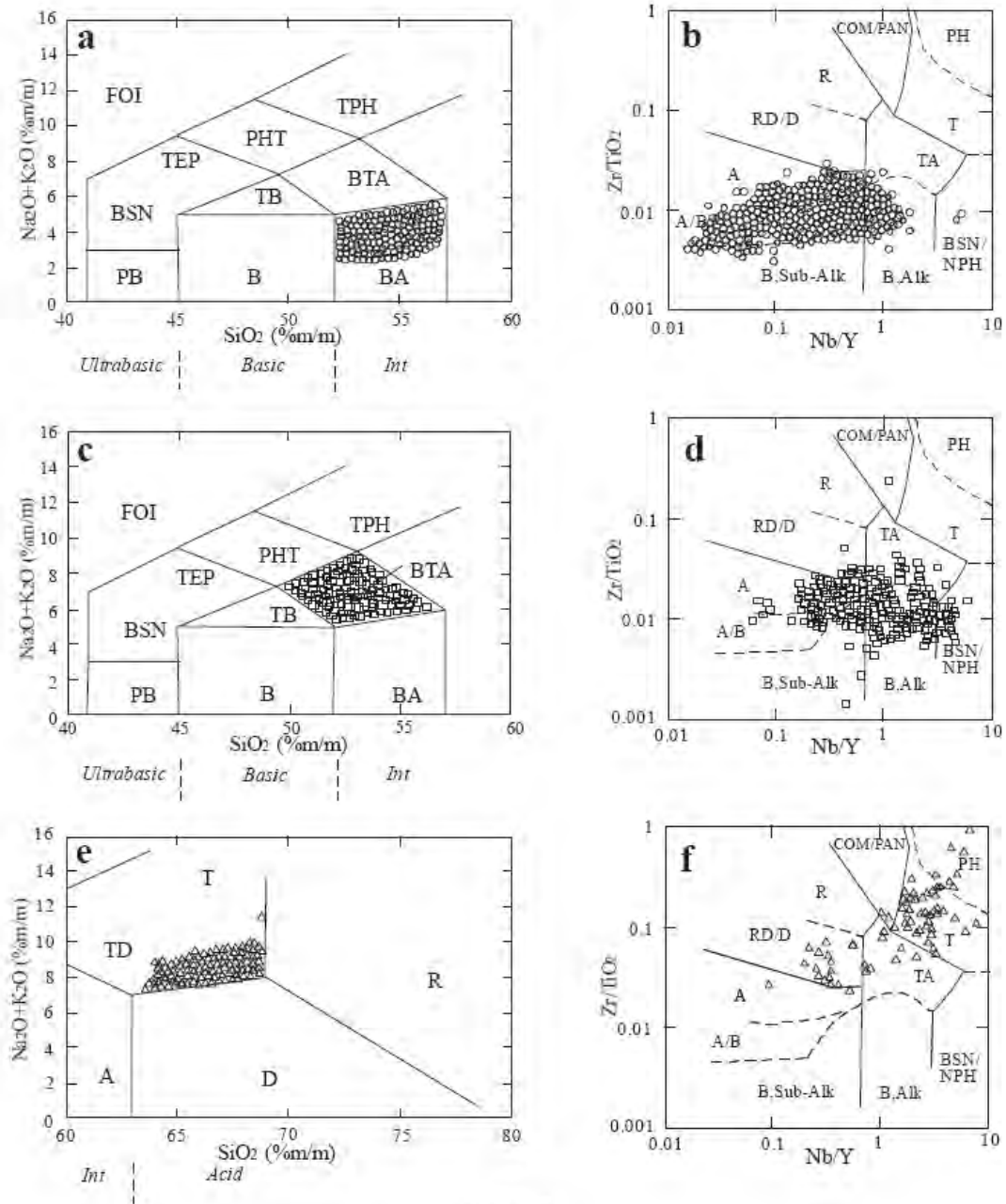


Figura 7

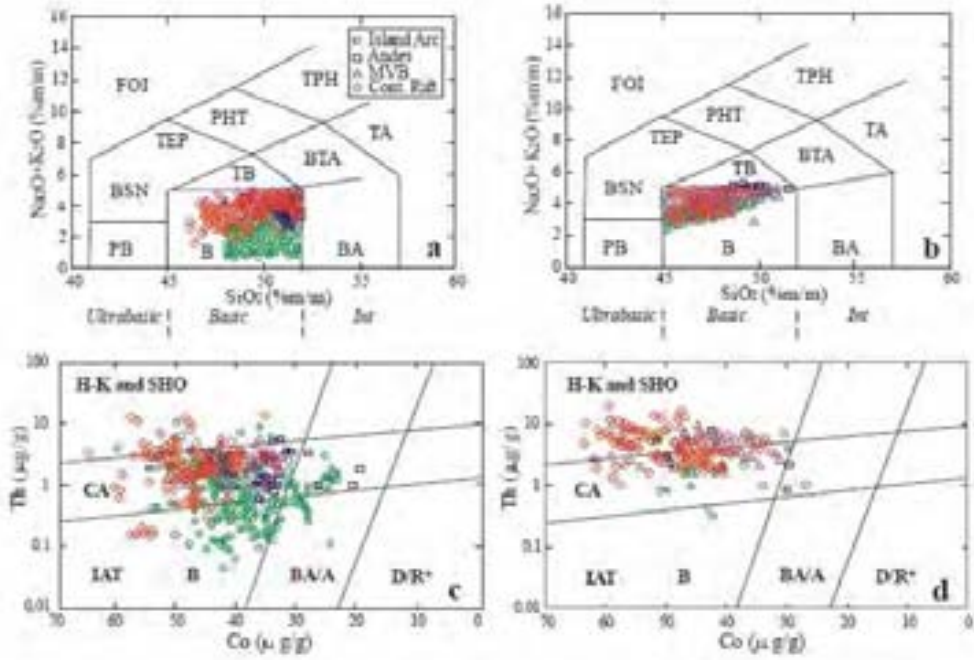


Figura 8

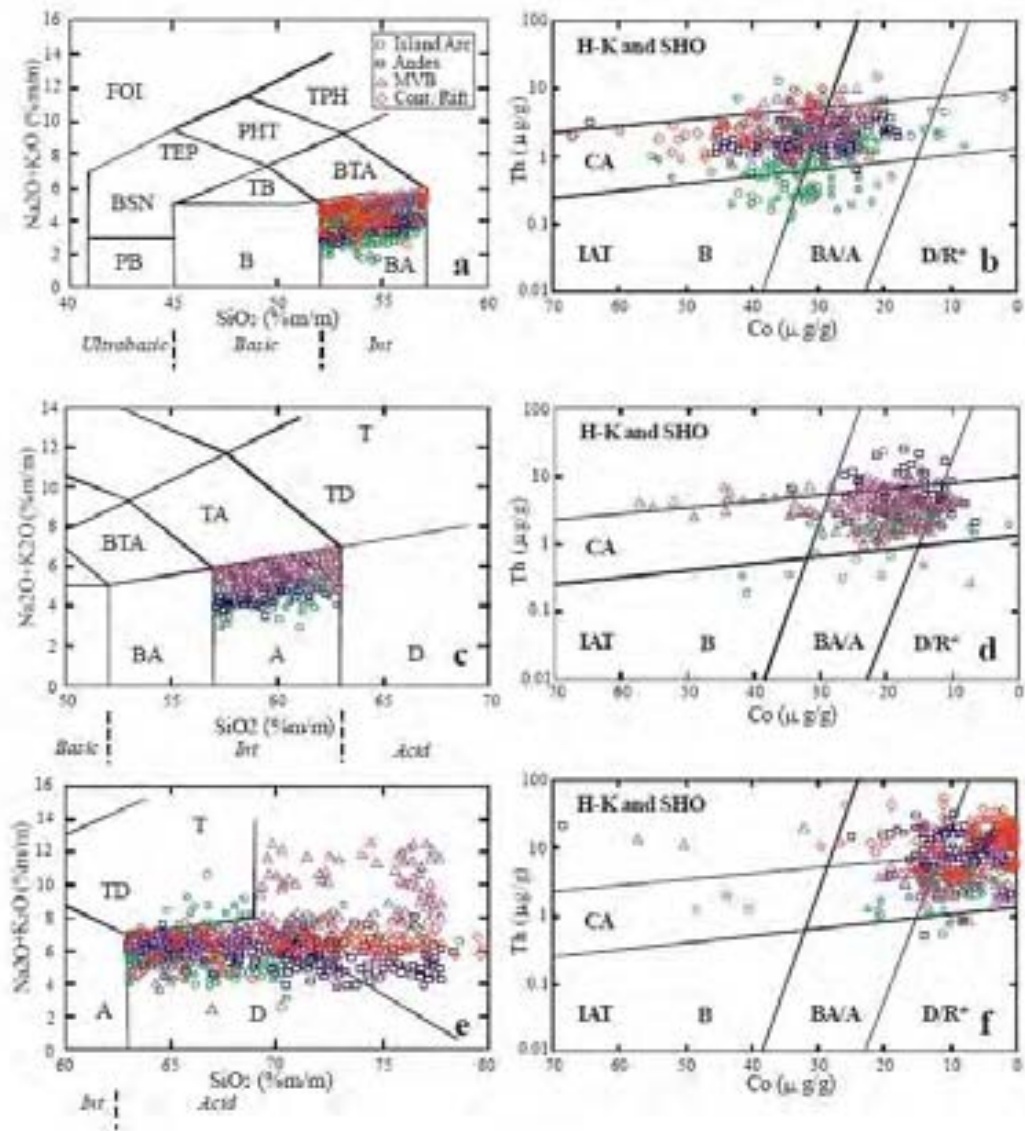


Figura 9

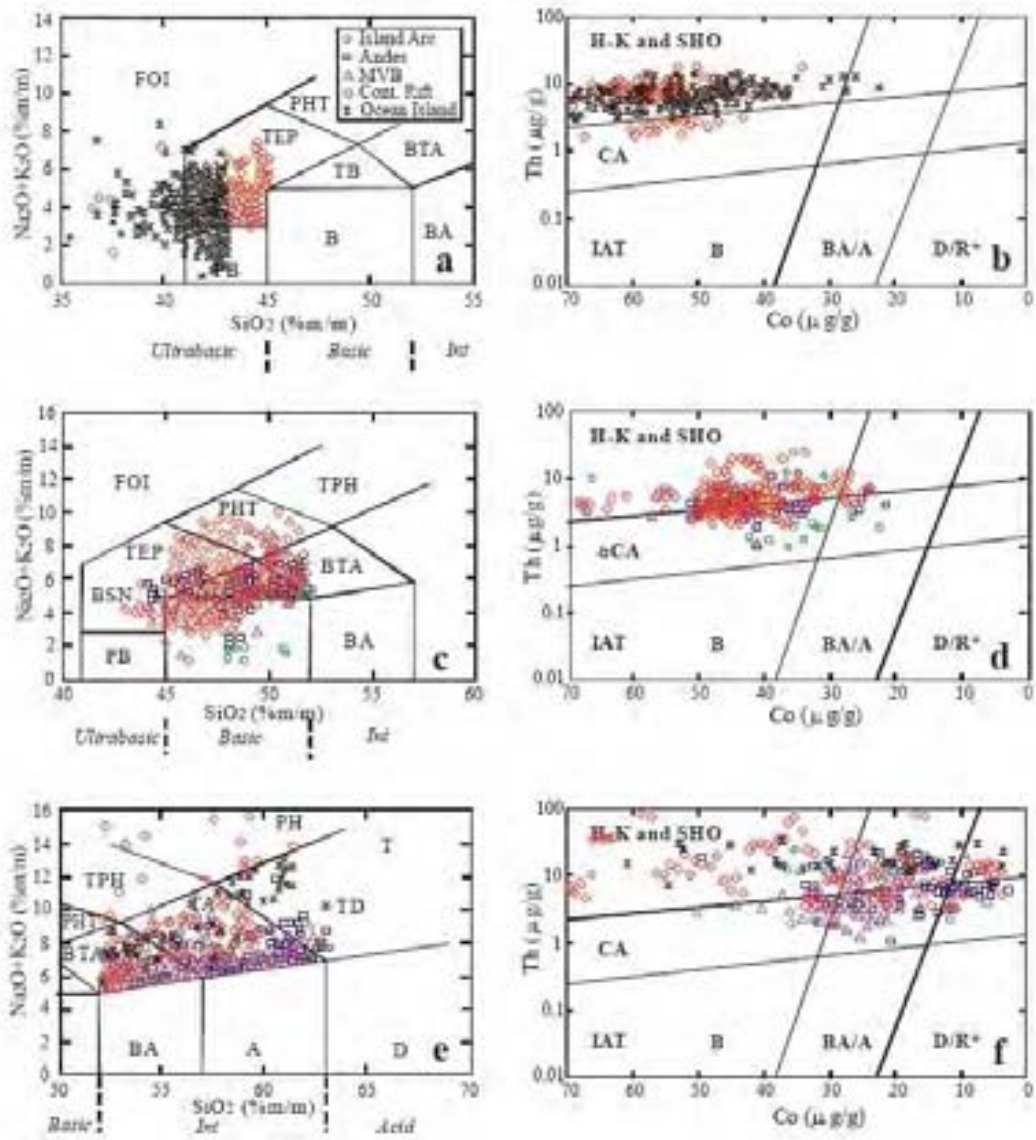


Figura 10

Table 1. Evaluation of Nb/Y–Zr/TiO₂ diagram (Winchester & Floyd 1977; called here W&F diagram) as compared to the IUGS volcanic rock classification (TAS and CIPW norm; Le Bas *et al.* 1986; La Bas 2000; Le Maître *et al.* 2002; Verma *et al.* 2002).

IUGS classification	Total number of Samples (%)	Figure #	Number of classified samples (% of classified samples) according to Nb/Y–Zr/TiO ₂ diagram (Winchester & Floyd 1977) W&F classification										
			Alkali-basalt	Sub-alkaline basalt	Basaltic/Nephelinitic	Andesite/Basalt	Andesite	Trachyandesite	Trachyte	Phonolite	Comendite/Pantellerite	Rhyodacite/Dacite	Rhyolite
Alkali basalt	650 (100)	1a, 1c	488 (75.0)	81 (12.5)	20 (3.1)	59 (9.1)	9 (1.4)	1 (0.1)					
Subalkali basalt	1200 (100)	1b, 1d	313 (26.1)	384 (32.0)	9 (0.7)	470 (39.2)	23 (1.9)					1 (0.1)	
Basaltic	541 (100)	2a, 2c	322 (59.5)	11 (2.0)	161 (29.8)					47 (8.7)			
Andesite	941 (100)	2b, 2d	40 (4.3)	34 (3.6)		87 (9.3)	786 (73.8)	35 (3.7)				29 (3.1)	
Trachyandesite	222 (100)	3a, 3d	35 (7.2)	7 (3.2)	3 (1.3)	2 (0.9)	85 (38.3)	45 (20.3)		4 (1.8)		40 (18.0)	
Trachyte	81 (100)	3b, 3e	2 (2.5)		1 (1.2)		6 (7.4)	15 (18.3)	24 (29.6)		3 (3.7)	30 (37.1)	
Phonolite	49 (100)	3c, 3f							11 (22)	37 (76)	1 (2)		
Dacite	524 (100)	4a, 4c	7 (1.4)	18 (3.4)		21 (4.0)	314 (59.9)	31 (5.9)				153 (29.4)	
Rhyolite	350 (100)	4b, 4d					9 (2.6)	57 (16.3)	1 (0.3)	22 (6.3)	109 (31.1)	94 (26.8)	58 (16.6)
Picrite (picrobasalt)	45 (100)	5a, 5b	27 (60)	2 (4)	16 (36)								
Picrite (alkali basalt)	106 (100)	5a, 5b	34 (32.1)	29 (27.3)		43 (40.6)							
Foidite	118 (100)	5c, 5d	45 (38.1)	2 (1.7)	37 (31.4)	1 (0.8)	33 (28.0)						
Picrobasalt	30 (100)	5e, 5f	20 (67)	10 (33)									
Teplite	155 (100)	6a, 6b	66 (42.6)	20 (12.9)	50 (32.3)			18 (11.6)				1 (0.6)	
Trachybasalt	314 (100)	6c, 6d	214 (68.2)	49 (15.6)	12 (3.8)	31 (9.9)	7 (2.2)	1 (0.3)					
Phonotephrite	73 (100)	6e, 6f	35 (72.6)	9 (12.3)	6 (8.2)	5 (6.9)							
Basaltic andesite	1299 (100)	7a, 7b	98 (7.9)	362 (28.1)	2 (0.2)	614 (49.6)	259 (20.9)	3 (0.2)				1 (0.1)	
Basaltic trachyandesite	392 (100)	7c, 7d	154 (39.3)	62 (15.8)	20 (5.1)	27 (6.9)	94 (23.9)	28 (6.6)		1 (0.3)	1 (0.3)	7 (1.8)	
Trachydacite	69 (100)	7e, 7f					2 (2.9)	8 (11.6)	36 (52.2)	7 (10.1)	2 (2.9)	14 (20.3)	

Numbers in *italic bold face* are for the correct classification; numbers in *italic* show the rock type, in which most samples of rock types not included in W&F diagram were classified.

Table 2. Evaluation of Nb/Y–Zr/TiO₂ diagram (Winchester & Floyd 1977; called here W&F diagram) as compared to the IUGS volcanic rock classification (TAS and CIPW norm; Le Bas *et al.* 1986; La Bas 2000; Le Maitre *et al.* 2002; Verma *et al.* 2002), using rocks from all the tectonic settings except arc rocks (For more explanation, see Table 1).

IUGS classification	Total number of Samples (%)	Figure # (for ref. only)	Number of classified samples (% of classified samples) according to Nb/Y–Zr/TiO ₂ diagram (Winchester & Floyd 1977) W&F classification											
			Alkali-basalt	Sub-alkaline basalt	Basanite/Nephelinite	Andesite/Basalt	Andesite	Trachyandesite	Trachyte	Phonolite	Comendite/Pantellerite	Rhyodacite/Dacite	Rhyolite	
Alkali basalt	556 (100)	1a, 1c	470 (84.5)	43 (7.7)	20 (3.6)	18 (3.2)	5 (0.9)							
Subalkali basalt	597 (100)	1b, 1d	291 (48.7)	234 (39.2)	9 (1.5)	60 (10.1)	2 (0.3)						1 (0.2)	
Basanite	536 (100)	2a, 2c	317 (59.1)	11 (2)	161 (30)				47 (8.8)					
Andesite	546 (100)	2b, 2d	36 (6.6)	16 (2.9)		22 (4)	424 (77.6)	24 (4.4)					24 (4.4)	
Trachyandesite	178 (100)	3a, 3b	10 (5.6)	4 (2.2)	3 (1.7)	2 (1.1)	50 (28.1)	65 (36.5)		4 (2.2)			40 (22.4)	
Trachyte	65 (100)	3c, 3d	2 (3)		1 (1.5)		3 (4.6)	15 (23.1)	24 (36.9)				17 (26.1)	
Phonolite	48 (100)	3e, 3f							36 (75)	11 (22.9)	1 (2)			
Dacite	272 (100)	4a, 4c	7 (2.5)	1 (0.4)		1 (0.4)	175 (64.3)	13 (4.7)					75 (27.6)	
Rhyolite	243 (100)	4b, 4d						21 (8.6)	22 (9)	1 (0.4)	108 (44.4)	46 (18.9)	45 (18.5)	
Picrite (picrobasalt)	45 (100)	5a, 5b	27 (60)	2 (4.4)	16 (35.5)									
Picrite (alkali basalt)	80 (100)	5a, 5b	17 (21.2)	24 (30)		39 (48.7)								
Foidite	118 (100)	5c, 5d	45 (38.1)	2 (1.7)	37 (31.4)	1 (0.8)	33(28)							
Picrobasalt	29 (100)	5e, 5f	19 (65.5)	10 (34.5)										
Tephrite	155 (100)	6a, 6b	66 (42.6)	20 (12.9)	50 (32.3)				18 (11.6)				1 (0.6)	
Trachybasalt	247 (100)	6c, 6d	176 (71.2)	30 (12.1)	12 (4.8)	21 (8.5)	7 (2.8)	1 (0.4)						
Phonotephrite	66 (100)	6e, 6f	48 (72.7)	8 (12.1)	5 (7.5)	5 (7.5)								
Basaltic andesite	500 (100)	7a, 7b	85 (17)	192 (38.4)	1 (0.2)	97 (19.4)	122 (24.4)	3 (0.6)						
Basaltic trachyandesite	351 (100)	7c, 7d	149 (42.4)	57 (16.2)	18 (5.1)	23 (6.5)	70 (19.9)	26 (7.4)		1 (0.3)	1 (0.3)		6 (1.7)	
Trachydacite	65 (100)	7e, 7f					1 (1.5)	8 (12.3)	36 (55.4)	7 (10.7)	2 (3)	11 (16.9)		

Table 3. Evaluation of Co-Th diagram (Hastie *et al.* 2007) for the classification of fresh volcanic rocks for different tectonic settings, as compared to the IUGS volcanic rock classification (TAS and CIPW norm; Le Bas *et al.* 1986; La Bas 2000; Le Maitre *et al.* 2002; Verma *et al.* 2002).

IUGS classification	Tectonic Setting	No. of Samples (%)	Figure #	B			BA/A			D/R*			
				Thol.	CA	SHO	Thol.	CA	SHO	Thol.	CA	SHO	
B _{subalk}	Island arc	237 (100)	8a, 8c	64 (27.0)	121 (51.0)	8 (3.4)	19 (8.0)	25 (10.6)					
B _{subalk}	The Andes	47 (100)	8a, 8c		39 (84)	4 (8)		4 (8)					
B _{subalk}	MVB	61 (100)	8a, 8c		56 (91.8)	5 (8.2)							
B _{subalk}	Continental rift	92 (100)	8a, 8c	8 (8.7)	63 (68.5)	21 (22.8)							
B _{alk}	Island arc	22 (100)	8b, 8d	2 (9)	17 (77)	3 (14)							
B _{alk}	The Andes	12	8b, 8d		4	6		2					
B _{alk}	MVB	48 (100)	8b, 8d	20 (42)		28 (58)							
B _{alk}	Continental rift	85 (100)	8b, 8d		38 (44.7)	46 (54.1)		1 (1.2)					
BA	Island arc	206 (100)	9a, 9b	19 (9.2)	70 (34.0)	4 (2.0)	25 (12.1)	82 (39.8)			6 (2.9)		
BA	The Andes	138 (100)	9a, 9b		83 (60.2)	4 (2.9)	1 (0.7)	49 (35.5)					
BA	MVB	129 (100)	9a, 9b		57 (44.2)	8 (6.2)	1 (0.8)	58 (45.0)	1 (0.7)				
BA	Continental rift	40 (100)	9a, 9b	9 (23)	28 (70)		2 (5)	1 (2)					
A	Island arc	109 (100)	9c, 9d	3 (2.8)	2 (1.8)		4 (3.7)	88 (80.7)	3 (2.8)		2 (1.8)	7 (6.4)	
A	The Andes	59 (100)	9c, 9d			1 (1.7)		26 (44.1)	19 (32.2)			12 (20.3)	
A	MVB	288 (100)	9c, 9d		9 (3.1)	12 (4.2)		230 (79.9)	16 (5.5)			21 (7.3)	
D+T+TD+R	Island arc	83 (100)	9e, 9f		3 (3.6)			18 (21.7)			4 (4.8)	47 (56.6)	11 (13.3)
D+T+TD+R	The Andes	112 (100)	9e, 9f			1 (0.9)		5 (4.5)	27 (24.1)		2 (1.8)	20 (17.8)	57 (50.9)
D+T+R	MVB	166 (100)	9e, 9f			3 (1.8)		41 (24.7)	6 (3.6)		2 (1.2)	75 (45.2)	39 (23.5)
D+T+TD+R	Continental rift	75 (100)	9e, 9f			2 (2.6)		2 (2.6)	14 (18.8)			18 (24.0)	39 (52.0)
BSN+FOI+PIC+TEP	Continental rift	86 (100)	10a, 10b		22 (25.6)	64 (74.4)							
BSN+FOI+PB+PIC+TEP	Ocean Island	180 (100)	10a, 10b			177 (98.3)			3 (1.7)				
BTA ¹ +PIC+TB	Island arc	19	10c, 10d		10	6		3					
BSN+BTA ¹ +TB+PIC	The Andes	51 (100)	10c, 10d		22 (43.1)	25 (49.0)		3 (5.9)	1 (2.0)				
BSN+BTA ¹ +PHT+TB	MVB	45 (100)	10c, 10d		19 (42.2)	24 (53.3)		2 (4.4)					
BSN+BTA ¹ +PHT+PIC+TB	Continental rift	134 (100)	10c, 10d		20 (14.9)	106 (79.1)		2 (1.5)	6 (4.5)				
BTA ² +T+TA	Island arc	18 (100)	10e, 10f		3	6		3	5			1	
BTA ² +T+TA	The Andes	70 (100)	10e, 10f		2 (2.9)	6 (8.6)		22 (31.4)	18 (25.7)			19 (27.1)	3 (4.3)
BTA ² +T+TA	MVB	94 (100)	10e, 10f		14 (14.9)	9 (9.6)		38 (40.4)	24 (25.5)			7 (7.4)	2 (2.1)
BTA ² +PH+PHT+TPH+TA	Continental rift	67 (100)	10e, 10f		4 (6.0)	30 (44.8)		3 (4.5)	17 (25.4)			4 (6.0)	9 (13.4)
BTA ² +T+TA	Ocean Island	29 (100)	10e, 10f			15 (52)			11 (38)				3 (10)

The rock abbreviations are (see Verma *et al.* 2002): B_{subalk}–subalkali basalt; B_{alk}–alkali basalt; BTA¹–basaltic trachyandesite (basic); PIC–picrite; TB–trachybasalt; BSN–basanite; PHT–phonotephrite; FOI–foiidite; TEP–tephrite; PB–picrobasalt; BTA²–basaltic trachyandesite (intermediate); T–trachyte; TA–trachyandesite; PH–phonolite; TPH–tephthiphonolite; BA–basaltic andesite; A–andesite; D–dacite; TD–trachydacite; R–rhyolite. For the Co–Th diagram (Hastie *et al.* 2007), B–basalt; BA/A–basaltic andesite/andesite; D/R*–diortite/rhyolite/latite/trachyte; Thol.–tholeiite; CA–calc-alkaline; SHO–high-K and shoshonite. The percentages for total number of samples < 50 were arbitrarily quoted as integers. Numbers and percentages in boldface are for “obvious mis-classification”, also called “minimum mis-classification” and those without boldface are correct rock name (not necessarily for the correct classification or success rate). The numbers and percentages in italic are for probable “minimum mis-classification”. Note this highlighting is reverse to that used in Tables 1 and 2.