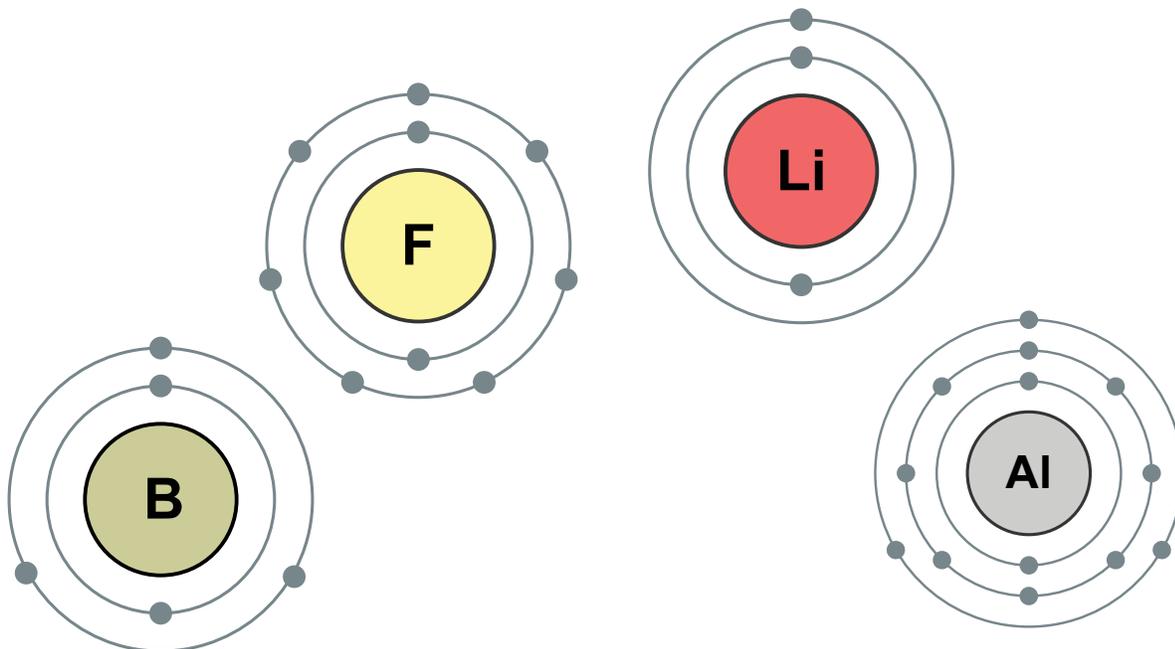


National Water-Quality Assessment Program

Software for Analysis of Chemical Mixtures: Composition, Occurrence, Distribution, and Possible Toxicity



F + Li	Li + B	Li + Al + B
F + Al	Al + B	F + B + Al
Li + Al	F + Li + Al	F + Li + B + Al
F + B	F + Li + B	

Scientific Investigations Report 2013–5030

Software for Analysis of Chemical Mixtures: Composition, Occurrence, Distribution, and Possible Toxicity

By Jonathon C. Scott, Kenneth A. Skach, and Patricia L. Toccalino

National Water-Quality Assessment Program

Scientific Investigations Report 2013–5030

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2013

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

To order USGS information products, visit <http://store.usgs.gov>

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Scott, J.C., Skach, K.A., and Toccalino, P.L., 2013, Software for analysis of chemical mixtures—composition, occurrence, distribution, and possible toxicity: U.S. Geological Survey Scientific Investigations Report 2013–5030, 27 p., 2 apps., <http://pubs.usgs.gov/sir/2013/5030/>.

Contents

Abstract	1
Introduction	1
Purpose and Scope	2
Needs for Mixture Software	2
Selecting Specific Mixtures for Future Research	2
Identifying the Causes for the Occurrence of Specific Mixtures	3
Determining the Effects of Mixtures	4
Mixture Software	4
Installation	4
Data Files	5
Programs	9
Program qwmixtures.pl	9
Program qwmixtrace.pl	12
Program qwmixbq.pl	13
Program qwmixsort.pl	14
Summary and Conclusions	15
References Cited	15
Appendixes	
1. Messages Printed by the Chemical-Mixture Analysis Software	19
2. Example Installation Procedure for the Chemical-Mixture Analysis Software	25

Figures

1. Graph showing number of mixtures resulting from a specified number of detected analytes	2
2. Graph showing number of mixtures resulting from a specified fraction of the detected analytes	3
3. An example relational database (RDB) analytical-data file	7
4. An example relational database (RDB) reference-data file	8
5. Annotated example of a report.txt file from the mixture-enumeration program qwmixtures.pl	11
6. Example of an uncompressed mixtures.rdb file from the mixture-enumeration program qwmixtures.pl	11
7. Graph showing mixture-enumeration-program completion time for all mixtures in a single sample with varying numbers of detected compounds and different computer hardware and operating systems	12

Tables

1. Perl packages used by the mixture software	4
2. Environment variables used by the mixture software	5
3. Coding convention for analytical results using remark codes and numeric concentration values	8
4. Command-line options for the mixture-enumeration program qwmixtures.pl	10
5. Descriptions of files created by the mixture-enumeration program qwmixtures.pl	11
6. Command-line options for the mixture-tracing program qwmixtrace.pl	13
7. Command-line options for the mixture-toxicity-estimation program qwmixbq.pl	13
8. Descriptions of the contents of the compressed file bq.rdb.gz created by the mixture-toxicity-estimation program qwmixbq.pl	14

Software for Analysis of Chemical Mixtures: Composition, Occurrence, Distribution, and Possible Toxicity

By Jonathon C. Scott, Kenneth A. Skach, and Patricia L. Toccalino

Abstract

The composition, occurrence, distribution, and possible toxicity of chemical mixtures in the environment are research concerns of the U.S. Geological Survey and others. The presence of specific chemical mixtures may serve as indicators of natural phenomena or human-caused events. Chemical mixtures may also have ecological, industrial, geochemical, or toxicological effects. Chemical-mixture occurrences vary by analyte composition and concentration. Four related computer programs have been developed by the National Water-Quality Assessment Program of the U.S. Geological Survey for research of chemical-mixture compositions, occurrences, distributions, and possible toxicities. The compositions and occurrences are identified for the user-supplied data, and therefore the resultant counts are constrained by the user's choices for the selection of chemicals, reporting limits for the analytical methods, spatial coverage, and time span for the data supplied. The distribution of chemical mixtures may be spatial, temporal, and (or) related to some other variable, such as chemical usage. Possible toxicities optionally are estimated from user-supplied benchmark data.

The software for the analysis of chemical mixtures described in this report is designed to work with chemical-analysis data files retrieved from the U.S. Geological Survey National Water Information System but can also be used with appropriately formatted data from other sources. Installation and usage of the mixture software are documented. This mixture software was designed to function with minimal changes on a variety of computer-operating systems. To obtain the software described herein and other U.S. Geological Survey software, visit <http://water.usgs.gov/software/>.

Introduction

Mixtures of chemicals are widely recognized as an important topic for environmental research (Schwarzenbach and others, 2006; National Research Council, 2009). Mixture software is needed to assist with mixture research by assisting with prioritization of research into specific mixtures and investigation of the possible causes and effects of mixtures.

A "mixture" is defined as a group of two or more of the detected chemical analytes in a sample. The "mixture components" are all of the chemicals in that particular combination of detected analytes. The "number of mixtures" in a sample is the set of all the pairwise combinations, triplets, quadruplets, and so on, up to the number of detected chemical analytes in a sample, as calculated by using the following equation (Mendenhall, 1975, p. 76).

$$\sum_{r=2, N} N! / r! (N - r)!$$

where

N = the number of detected analytes;
r = the number of analytes in the mixture; and
! = the factorial operator, denoting the product of the integers from one to the preceding integer (for example, 4! = 1 × 2 × 3 × 4 = 24).

Millions or more mixtures of analytes may exist in a single sample; therefore, mixtures in environmental datasets can be difficult to study.

The National Water-Quality Assessment Program of the U.S. Geological Survey (USGS) developed computer programs (mixture software) to facilitate mixture research. The mixture software is used to analyze the composition, occurrence, distribution, and possible toxicity of mixtures in a user-supplied dataset of chemical analytes in samples from water or other media. The mixture software consists of four related computer programs:

1. The `qwmixtures.pl` program is used to study mixture composition and occurrence. The program enumerates (identifies and counts) and sorts by frequency of occurrence the mixtures in user-supplied analytical data.
2. The `qwmixtrace.pl` program is used to study mixture distribution and to trace mixtures to the samples in which the mixtures were detected. These distributions may be spatial, temporal, or characterized by some other variable (such as chemical-usage patterns).
3. The `qwmixbq.pl` program is used to study possible toxicity of mixtures and prepares benchmark-normalized concentration sums for each sample.

2 Software for Analysis of Chemical Mixtures: Composition, Occurrence, Distribution, and Possible Toxicity

4. The qwmixsort.pl program is used by the other programs for sorting files.

Simple command-line interfaces are used to specify user options for the programs. The mixture software is designed to be usable on a variety of computer systems. The time required to run the programs may be hours or days and will vary depending on the size and complexity of the input data and the computing hardware. To obtain the software described herein and other USGS software, visit <http://water.usgs.gov/software/>.

Purpose and Scope

This report describes needs for mixture analysis and the types of scientific questions mixture-analysis software could be used to investigate. The report subsequently documents mixture-analysis software: the set of four computer programs developed and used by the USGS to research mixtures.

The mixture software described in this report was designed to meet the specific needs of USGS scientists to identify and describe mixtures of chemicals in water, sediment, and biological tissue by using data retrieved from the USGS National Water Information System (NWIS). Chemical-analysis data files from other databases may be used with the software if those files are formatted to match the specifications described in this report.

Needs for Mixture Software

Several scientific and engineering professions have reasons to investigate mixtures in environmental settings, including hydrologists, hazardous-waste investigators,

regulatory authorities, water-treatment engineers, biologists, and toxicologists. This software was created to meet the needs of these investigators for selecting specific mixtures for future research, identifying the causes for the occurrence of specific mixtures, and determining the effects of mixtures.

Selecting Specific Mixtures for Future Research

The number of mixtures of chemicals occurring in the environment is very large; therefore, so is the scope of research into mixtures. The number of mixtures present in a sample increases exponentially as the number of co-occurring analytes increases (fig. 1). For example, 30 detected analytes produce 1,073,741,770 mixtures because of the number of combinations of pairs, plus combinations of triples, plus combinations of quadruples, and so on. The greatest number of mixtures resulting from a sample with N detected analytes results from combinations of $N/2$ mixture components (fig. 2). For example, in a sample with 25 detected analytes, most of the mixtures will be enumerated with 13 components. Therefore, the distribution of the number of mixture components for a particular set of samples is entirely dependent on the number of compounds detected in each of the samples (Gilliom and others, 2006). The detection frequency is dependent on the laboratory reporting levels for each of the chemicals, which commonly vary with time and by laboratory.

An occurrence-based screening can be used to limit the scope of future mixture research (Squillace and others, 2002). Various research applications will use occurrence data differently. For geoscience and industrial applications, the most important facts are the chemical properties and most likely reactions of the frequently occurring mixtures. The primary concern about mixtures in an aquatic or human-health context, however, is identification of the possibly

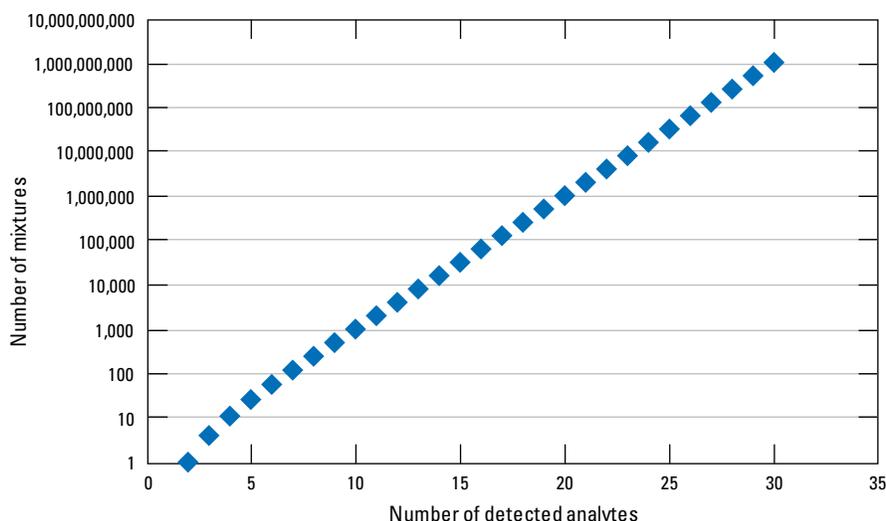


Figure 1. Number of mixtures resulting from a specified number of detected analytes.

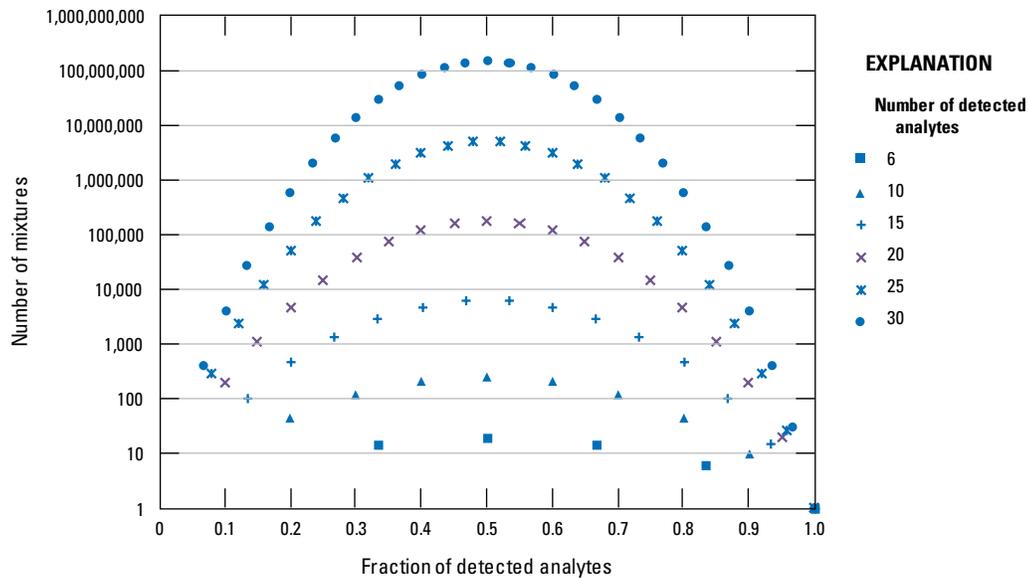


Figure 2. Number of mixtures resulting from a specified fraction of the detected analytes.

most toxic mixtures and the occurrence frequencies of those mixtures. Identification of frequently occurring mixtures, and (or) mixtures with greater possible toxicity, could be used to restrict the scope of future research into specific mixtures.

Concentrations of individual contaminants are commonly compared to human-health or aquatic-life benchmarks to evaluate toxicity. Mixture toxicity is more complex because of differing modes of action among the contaminants, differing exposure routes, and the possibilities for additive, synergistic, antagonistic, or independent combined effects (Pape-Lindstrom and Lydy, 1997; Altenburger and others, 2004; Brian and others, 2005; Lydy and Belden, 2006; Belden and others, 2007; Teuschler, 2007; Gade and others, 2012). Mixture toxicity can depend on many factors, including the target organisms (human or aquatic) and the type of deleterious effect (such as lesions, cancers, death), and is unknown until evaluated. Relative estimates of possible mixture toxicity can be useful for comparative assessment purposes (U.S. Environmental Protection Agency, 1986; U.S. Environmental Protection Agency, 2000) because such estimates can inform other toxicological approaches and can be used to screen candidate mixtures for analysis (Ryker and Small, 2008; Backhaus and Faust, 2012), such as traditional dose-response studies, computational toxicology, or high-speed cellular screening (Ehrenberg, 2008; Abbott, 2009).

The U.S. Environmental Protection Agency (USEPA) has established primary drinking-water standards for some mixtures of contaminants, such as polychlorinated biphenyls, total xylenes, total trihalomethanes, and haloacetic acids (U.S. Environmental Protection Agency, 2012a). Knowledge of mixture occurrence in drinking water and drinking-water sources, and the possible toxicity of those mixtures, may

be useful for setting priorities for future environmental monitoring and toxicity assessments of mixtures in drinking water and drinking-water sources (Toccalino and others, 2012; G. Delzer, U.S. Geological Survey, written commun., 2012).

Identifying the Causes for the Occurrence of Specific Mixtures

Hydrologists and geochemists may want to identify why specific substances occur together in mixtures. The chemicals may already be present in the geologic setting and released by a change in pH or some other factor. Some chemicals are associated with human activities, such as agricultural usage of fertilizer and pesticides. In some cases, the detected mixture compounds may already have been parts of a product, such as gasoline, that was introduced into the environment. In addition, some mixtures occur because of the degradation of a parent compound into degradation products, especially organic compounds and radionuclides (Stackelberg and others, 2001).

Comparison of the expected probability of a mixture resulting from simple random co-occurrence to the actual co-occurrence frequency can provide helpful information when researching mixture distributions and sources (Squillace and Moran, 2006; DeSimone, 2009). Mixtures with nonrandom occurrence frequencies may be traced to sampling locations and investigated for spatial correlations such as hydrogeologic setting, land use, or chemical-usage patterns. Study of the distributions of mixtures may help identify why mixtures occur in specific locations. Some mixtures may be useful as indicators of particular outcomes, such as source tracking of a pollution plume by using a rare combination of substances (Peterson and others, 1985).

Determining the Effects of Mixtures

The effects of mixtures are difficult to determine. Correlation of mixture occurrence with various distributional parameters may provide factors for further study. The mixture distributions may be spatial, temporal, and (or) related to some other variables, such as chemical usage or mixture-component concentrations (Squillace and others, 2002; Phillips and others, 2010; Toccalino and others, 2012). Abrupt ecological changes may be associated with the occurrence of specific mixtures, or the epidemiology of specific health anomalies may correlate with the geographic distribution of specific mixtures (Pickrell, 2002; Kortenkamp, 2008).

Engineers may be interested in the occurrence frequency and geographic distribution of specific mixtures that may be received by industrial plants. Some mixtures may foul an industrial process or may cause a process to produce undesirable byproducts (Mansoori, 2002). Water-treatment plant designs may be altered to remove specific mixtures.

Some investigations of mixtures, such as dose-response studies, need estimates of ambient concentrations. One approach for estimating ambient concentrations could involve computation of minimums, maximums, and percentiles of the concentrations of the mixture components measured in the samples in which the mixtures were observed.

Mixture Software

The mixture software was designed to meet the needs to study the composition, occurrence, distribution, and possible toxicity of mixtures. The mixture software was developed and tested by using the SunOS operating system, version 5.10, with the Perl programming language, version 5.10.1 (Wall and

others, 1996). Although portability was a consideration in the development of the mixture software, usability on other operating systems is mostly unknown. The mixture software requires a customized computing environment for the installation and appropriately formatted data files as input. The software is controlled with command-line options. The installation, formats, and usage of the software are described subsequently. Appendix 1 lists and explains warning and error messages that may be printed by the software programs.

Installation

The Perl programming language is included with several operating systems and can be installed for others (Perl.org, 2012). Each Perl distribution may differ by the software version, as well as the inclusion of different optional Perl “packages.” Perl packages are additional software components that improve the data-processing capabilities for Perl programmers and their programs. The mixture software utilizes and therefore requires the presence of several Perl packages (table 1). Perl packages not included in the distribution commonly are installed by using the Perl package manager (ppm) program (Nesbitt, 2012). Perl and the Perl packages are free software, distributed conditional to an artistic license. Use the command “perldoc perlartistic” to read the terms (Perl.org, 2013).

Three environment variables should be set after the mixture software has been copied to the local computer (table 2). The method used to set environment variables differs depending on the computer-operating system. UNIX users normally set environment variables within shell scripts that run during login. Microsoft Windows users typically set environment variables by using the “Control Panel.”

Table 1. Perl packages used by the mixture software.

[X indicates that the Perl package (row) is used by the program; -- indicates that the Perl package is not used]

Perl package	Programs			
	qwmixtures.pl	qwmixtrace.pl	qwmixbq.pl	qwmixsort.pl
File::Basename	X	X	X	--
File::MergeSort	X	--	--	--
File::Sort	X	--	--	X
File::Spec	X	X	X	X
File::Spec::Functions	--	X	X	--
File::Temp	X	--	X	--
IO::Pipe	X	--	X	X
IO::Zlib	X	X	X	--
Math::BigInt	X	--	--	--
Math::Combinatorics	X	--	--	--
Sort::External	X	--	--	--

Table 2. Environment variables used by the mixture software.

Variable name	Meaning	Default definition
TMPDIR	Directory to store temporary files	The output directory
QWMIXDIR	Directory of mixture software	The current directory
QWMIXLIMIT	One-half of the maximum of number of detections	24

The TMPDIR variable (table 2) may be used to alter the location of where large temporary disk files are stored while the program is running. Moving these files from the default directory may allow solution of more complex problems involving many mixtures and may slightly decrease the time required to solve large problems.

The QWMIXDIR variable (table 2) is used to store the pathname of the directory in which the mixture-software program files are installed. For convenience, the PATH variable may be modified to include the QWMIXDIR directory.

The QWMIXLIMIT variable (table 2) controls the maximum number of detected constituents that the mixture software will solve. The actual limit will be determined by the particular hardware and operating system where the mixture software is installed. The largest number of detected constituents for which mixtures can be enumerated will be two times the selected value of QWMIXLIMIT. Internally, the software splits the enumeration task into two groups of analytes when the number of detections exceeds the QWMIXLIMIT value. The default value for the QWMIXLIMIT variable (24) should work for many installations. Users may evaluate a larger or smaller limit by using the following procedure:

1. Set QWMIXLIMIT=1000.
2. Create a set of files, each with one sample and an increasing number of detections, beginning with a 25-detection sample.
3. Run the mixture-enumeration program (qwmixtures.pl), keeping all mixtures.
4. If the program completes successfully without exceeding any disk or memory limitations, run the program again with a sample containing one more detection (such as 26). If the program fails this attempt, then run the program again with a sample containing one less detection (such as 24) than the previous number.
5. Iterate until the sample with the maximum successful number of detected analytes is determined.
6. Set QWMIXLIMIT to one-half the number of detections in the last successful sample.

Problems involving several hundred or more samples may require alteration of operating-system limits on the number of open files allowed to a process. In the UNIX environment, the “ulimit” or “plimit” commands may be used to increase this limit. The SunOS 5.10 version installed by the USGS sets this limit to 256 open files by default. The sorting and merging operations of the mixture enumeration are then effectively limited to about 250 samples, unless the limit is increased. Example installation procedures are presented in appendix 2.

Data Files

The mixture software works with simple tabular data files. Observations are presented as rows (lines) in each file, and variables are presented in tab-delimited columns. One tab is placed between each variable. The data columns commonly appear to be misaligned when viewed with text-processing software because of the differing number of data characters among rows and columns. Other software, such as spreadsheets, however, display the data aligned in columns and rows. The particular implementation of tab-delimited data files used with the mixture software is known as relational database (RDB) files (Manis and others, 1988; Hobbs, 1992; Jerius, 2004). The USGS has used RDB files since the mid-1990s and more recently as an output format for hydrologic data retrieved from the NWISWeb Web site (Granato, 2009).

Numerous programs have been written and are publicly available for working with RDB files. Use of the mixture software does not require local installation of these additional programs for manipulating RDB files; however, users may find the additional software helpful.

The initial rows in an RDB file may contain free-text comments or metadata and are identified with a pound sign (“#”) as the first character of the row. Comments formatted in this manner may appear among only the initial rows of the data file.

The first noncomment RDB-file row contains column names. There are no inherent restrictions on column names other than avoidance of tab characters. Because RDB files are often used as a simple exchange format among other software packages, however, the limitations of other packages commonly are observed, such as a limited number of characters and avoidance of spaces and most symbols.

The second noncomment row in the RDB file is the column-definition row, which specifies basic column-format information. Only a few data types are explicitly defined for RDB files, each identified with a single character: character (s), numeric (n), date (d), and month (m). The month data type is rarely used and is not used in the mixture software. The column-definition row also consists of maximum-column-length hints, expressed as the number of characters, that precede the data type. These specifications are merely “hints” because an RDB file is still valid if the number of characters in data for the column exceeds the defined length.

6 Software for Analysis of Chemical Mixtures: Composition, Occurrence, Distribution, and Possible Toxicity

The column-length hint and the single-character data-type declaration are concatenated for each data column. The column-definition row consists of a tab-delimited assemblage of these length and type declarations for all of the columns. The remaining rows in the data file consist of the observations. Figure 3 presents an example of an RDB file.

An RDB file can be created by using a variety of methods. Some techniques may create an improperly formatted RDB file, such as one with data rows with inconsistent numbers of tab characters. The RDB “valid” command can be used to verify proper RDB file structure. The mixture software does not verify the input-data format and performs only limited checks on the input data.

The input and output data files used with and created by the mixture software are RDB files. The column-definition row is expected by the mixture software for input data and is produced by the mixture software in the output; however, the mixture software does not use the column-definition data. When the mixture software produces a column identifying the analytes in a mixture, a column named “mixture” is created with a definition of “40s”; however, the actual length of some mixtures may exceed the length hint. Mixtures are identified within the “mixture” column by concatenating the column names of the detected analytes, with a symbol for addition (“+”) inserted between each analyte name.

Analytical-data files can be retrieved from the NWISWeb Web site (U.S. Geological Survey, 2002) and used directly with the mixture software. When the desired set of sites, samples, and constituents have been identified, the NWISWeb user should select the option “tab-separated data, one sample per row with remark codes combined with values.”

USGS users of NWIS (U.S. Geological Survey, 1998) can retrieve analytical-data files that are usable with the mixture software. USGS users of NWIS should retrieve their data in “by-sample” RDB format with “remarks embedded with values” (as described in U.S. Geological Survey, 2006, section 3.4.5.1, “Selecting the Output Format”). Output columns should include any desired site and sample-identification data (such as sampling date), along with columns for the desired constituents.

Several water-quality database systems have used five-digit parameter codes as a concise mechanism for identifying particular types of measurements (Schornick and others, 1989). Historically, the meanings of these codes were coordinated by the USEPA; however, the USGS has maintained the dictionary of code definitions since 1999 (U.S. Geological Survey, 2012).

Both of the retrieval methods described in the preceding paragraphs utilize a parameter-code based naming convention for columns that contain chemical concentrations. The convention creates a column name by concatenating the letter “P” followed by the five-digit parameter code for the constituent. This convention for naming columns containing analytical results has the merits of consistency with past practices (John Booker and David Maddy, U.S. Geological

Survey, written commun., 1982), brevity, and compatibility with other software, such as P-STAT (P-stat.com, 2013) or the Statistical Analysis System (Helwig and Council, 1979).

Initially, much of USGS mixture research concentrated on water, though some work was done with streambed sediments (Phillips and others, 2010). The mixture software is not restricted to any specific sampling medium and could be applied to mixtures in rocks (Eberl, 2008), effluent (Masoner and others, 2012), food (U.S. Department of Agriculture, 2012), air (U.S. Environmental Protection Agency, 2012b), or biologic tissues and fluids (Moulton and others, 2002; Centers for Disease Control and Prevention, 2012).

Data from sources other than NWIS can be formatted into tab-delimited RDB files that are compatible with the mixture software. Each data row in the file should represent one environmental sample and the analytical results for that sample. Column names for analytes preferably should be concise and created by using uppercase and lowercase letters, digits, and the dash (“-”), underscore (“_”), and period (“.”) symbols. Additional columns may be present in the data file; however, the mixture analytes should be in a contiguous block of columns.

Analytical results should be recorded in the data field that intersects the sample (row) and constituent (column) for that measurement (fig. 3). Unqualified detections should be coded numerically. The mixture software does not record the units of measure for the analytes. If the analyte was not measured for that sample, the field should remain empty (delimited by two adjacent tab characters). Remark codes (U.S. Geological Survey, 2006) may be used in conjunction with numeric values or may be used alone to indicate nonquantitative results. A variety of measurement conditions may be expressed by using remark codes and (or) values (table 3). These remark codes indicate to the mixture software a variety of analytical outcomes, especially whether or not the analyte was detected.

One or more additional columns should be supplied in the analytical data. These “metadata” columns provide ancillary information about each sample that can be used to uniquely identify each sample and sampling location and (or) to provide additional data that may be useful in the analysis. Additional data might include geospatial coordinates and information about the setting and conditions where and when the sample was collected.

Users may optionally supply a reference-data file containing human-health or aquatic-life benchmark concentrations and (or) an additional name for the analytes (fig. 4). The reference-data file should be prepared in the RDB format and contain a row for each analyte. Additional rows that describe analytes not included in the current mixture analysis are ignored. One column that lists the column name for each analyte as presented in the analytical-data file must be provided. Additional columns that identify another name for the analyte and (or) a benchmark concentration may be included.

```

#
# File created on 2012-09-18 09:51:24 EDT
#
# site_no          - Station number
# sample_dt       - Sampling date
# p00950          - Fluoride, water, filtered, milligrams per liter
# p01030          - Chromium, water, filtered, micrograms per liter
# p01106          - Aluminum, water, filtered, micrograms per liter
# Description of remark_cd:
# < - less than
# E - estimated
#
site_no           <<> sample_dt           <<> p00950       <<> p01106       <<> p01030
15s              <<> 10d                <<> 12s          <<> 12s          <<> 12s
332611110495101 <<> 1984-11-08        <<>             <<> 302000       <<>
332611110495102 <<> 1984-11-08        <<>             <<> 130000       <<>
332611110495103 <<> 1984-11-08        <<>             <<> 230000       <<>
332611110495104 <<> 1984-11-08        <<>             <<> 17000        <<>
332629110495801 <<> 1984-11-07        <<>             <<> 225000       <<>
332629110495802 <<> 1984-11-07        <<>             <<> 230000       <<>
332629110495803 <<> 1984-11-07        <<>             <<> 141000       <<>
332629110495804 <<> 1984-11-07        <<>             <<> 9620         <<>
332629110495807 <<> 1989-01-10        <<> 10.5         <<> 41500        <<>
332717110501901 <<> 1984-11-07        <<> 0.10         <<> < 100        <<>
332717110501902 <<> 1984-11-06        <<>             <<> 184000       <<> < 100
332717110501903 <<> 1984-11-07        <<>             <<> 3620         <<> < 50
332717110501904 <<> 1986-07-09        <<>             <<> 100000       <<>
332904110504801 <<> 1984-11-08        <<> 9.70         <<> 12000        <<>
332904110504802 <<> 1984-11-08        <<>             <<> 10600        <<> < 20
332904110504803 <<> 1984-11-08        <<>             <<> < 80         <<> < 20
333108110515601 <<> 1989-03-09        <<> 9.40         <<> 9100         <<> 24.0
333108110515602 <<> 1989-01-11        <<> 3.30         <<> 660          <<> < 15.0
333108110515603 <<> 1990-08-25        <<> 4.60         <<> < 1000       <<>
333151110520501 <<> 1986-07-07        <<> 0.10         <<> 60           <<>
333151110520503 <<> 1986-07-07        <<> 0.40         <<> 20           <<>
333151110521201 <<> 1997-05-06        <<> 5.12         <<> 4450         <<> 23.5
333151110521202 <<> 1997-05-06        <<> 3.58         <<> 5720         <<> < 15.0
333204110521701 <<> 1997-05-06        <<> 4.78         <<> 3510         <<> 17.6
333204110521702 <<> 1997-05-06        <<> 5.71         <<> 2680         <<> 24.7
333221110522601 <<> 1999-02-11        <<> 3.35         <<> 1320         <<> < 42.0
333236110523601 <<> 2000-08-30        <<> 2.66         <<> 60           <<> < 42.0
333243110523801 << > 1999-02-11        <<> 1.39         <<> E 20        <<> < 42.0

```

<<>, Tab character

Figure 3. An example relational database (RDB) analytical-data file.

8 Software for Analysis of Chemical Mixtures: Composition, Occurrence, Distribution, and Possible Toxicity

Table 3. Coding convention for analytical results using remark codes and numeric concentration values.

[Italicized *value* indicates a numeric entry. No*, default behavior; --, not applicable]

Measured	Detected	Remark-value syntax	Comments
No	--	Empty cell	No measurement was performed.
Yes	No	U	Not detected or quantified.
Yes	No	< <i>value</i>	Not detected above the reporting limit (<i>value</i>).
Yes	Yes	M or N	Detected but not quantified.
Yes	Yes	<i>value</i>	Quantified detection at <i>value</i> concentration.
Yes	Yes	> <i>value</i>	Detection exceeding the high range (<i>value</i>).
Yes	Yes	E <i>value</i>	Detection at the estimated <i>value</i> .
Yes	Yes	A <i>value</i>	Detection; <i>value</i> is an average.
Yes	Yes	S <i>value</i>	Detection; <i>value</i> is a “most probable value.”
No*	--	V <i>value</i>	<i>Value</i> is affected by contamination. Default interpretation of contaminated results as “not measured” may be overridden by a user-specified program option.

```
# Mixtures software reference-data file, updated 2012-11-18
#
# pcode      - analytical-data column name
# benchmark  - benchmark value
# name       - analyte name
# remarks    - discussion, mg/L milligrams per liter, ug/L micrograms per liter,
#             MCL: Maximum Contaminant Level
#
pcode      <<> benchmark      <<> name      <<> remarks
15s        <<> 10n             <<> 20s        <<> 20s
p00950     <<> 4.             <<> Fluoride   <<> mg/L, MCL
p01106     <<> 99999999          <<> Aluminum  <<> ug/L, placeholder
p01030     <<> 100              <<> Chromium  <<> ug/L, MCL
```

<<>, Tab character

No aluminum benchmark concentration is known, so 99999999 is used as a placeholder

Figure 4. An example relational database (RDB) reference-data file.

Basic mixture analysis consists of identifying and sorting the combinations of the detected analytes in each sample, without regard to concentration. Researchers may evaluate mixture rankings by benchmark-normalized concentration sums. Each contaminant is assigned a benchmark concentration that incorporates an estimate of toxic concentration for a specific route of exposure (for example, ingestion, inhalation, contact, intravenous). When benchmark concentrations are supplied, the mixture software calculates the ratio of the detected concentration to the respective benchmark to compute a normalized concentration. Therefore, when benchmark concentrations are supplied, users should ensure that the benchmark values are expressed in the same units of measure as the analytical data. Alternatively, the user may compute benchmark-normalized concentration ratios and use those data with the mixture software, in lieu of analyte concentrations.

The calculation of a normalized concentration is affected by the presence of a remark code in the analytical data. Remark codes “E” (estimated), “A” (average), and “S” (most probable value) are ignored. Remark codes “M” and “N” (nonquantified detections) are treated as if the concentration is 0.001 reporting units. The remark code “>” (high range exceeded) is treated as a detection at the high range value. Interpretation of the “V” (contamination) remark code depends on the user-specified options but is treated by default as not measured. Normalized concentrations are not computed for analytes qualified with “<” (less than) or “U” (not detected or quantified) remark codes.

Computer-operating systems differ in the method for separating records in text files (Noria, 2006). Sometimes a “carriage return” character separates records, sometimes a “line feed” character does, and other times both do. These three scenarios describe the conventions used by the majority of operating systems currently in use (Net Applications, 2013). The mixture software was designed in anticipation of a potential problem that could be caused by development of data files on a computer with an operating system using one record-separator convention and used as input data to the mixture software on a computer with a different convention; however, files written by the mixture software use the convention of the host computer-operating system.

Some of the files created by the mixture software use data-compression techniques (Ziv and Lempel, 1978). The files are compressed to reduce the disk space needed for using the mixture software. The compressed files are named with a “.gz” suffix (sometimes called a filename “extension”).

Programs

The mixture software consists of four programs: qwmixtures.pl, qwmixtrace.pl, qwmixbq.pl, and qwmixsort.pl. Each program is started with a command line that specifies options to control the behavior of the program. Programs may take hours or days to complete, depending on the size and complexity of the input data.

The programs are designed to write all of the output files associated with a specific analytical dataset within a user-specified directory. The programs will create or overwrite consistent filenames within that directory.

Program qwmixtures.pl

The qwmixtures.pl program is used to investigate mixture occurrences and must be run prior to using the next two subsequently described programs. The qwmixtures.pl program enumerates the mixtures in an analytical-data file and sorts the mixtures by decreasing frequency of occurrence. If alternate analyte names and (or) benchmark-normalized concentrations are desired, a reference-data file containing these data should be supplied when running the program.

The command-line options for the qwmixtures.pl program are presented in table 4. Both the “-f” and “-d” options must be specified. The named directory will be created, if it does not already exist. Any existing files from a previous run of the program in the directory specified by the “-d” option will be overwritten without warning. An example command line is shown next.

```
qwmixtures.pl -f chemdata.rdb -d mixdir
               -m 1 -s 2
```

The options “-s” and “-e” are used to identify the contiguous block of analytical data columns. The time required to run the program may be significantly reduced by using the “-m” option to restrict the minimum number of mixture occurrences to be included in the final tally. When the “-m” option is used, all mixtures are identified and written to an initial unsorted (“raw”) mixture-enumeration file; however, mixtures that are counted less frequently than the “-m” lower limit are excluded from the final report that is sorted by the occurrence frequency. All mixtures will be retained in the final ranked output file if the “-m 1” command-line option is specified.

Some investigations may require recensoring all analytes to a single reporting level. No recensoring is performed unless the “-c” option is specified. Every analyte is recensored by specifying the “-c” option followed by the censoring value. All detections below the censoring value and nonquantified detections are recensored to nondetections when global recensoring is enabled. Detections that exceed the high-range value (accompanied by remark code of “>”) are not recensored. The “-c” option normally will not be used unless all concentrations are measured in the same units.

The “Installation” section of this report describes how the TMPDIR environment variable controls the directory where temporary files are created. The “-T” option overrides the TMPDIR specification for the duration of a single run of the program.

The mixture software normally treats concentrations associated with contamination (with a “V” remark code) as not measured. This treatment can be overridden with the

Table 4. Command-line options for the mixture-enumeration program `qwmixtures.pl`.

[Italicized entries indicate user-specified data]

Option	Default	Description
Required		
<i>-f file</i>	None	File name of the analytical data.
<i>-d directory</i>	None	Directory name for storing output files.
Optional		
<i>-s column</i>	2	Start-constituent column number or name.
<i>-e column</i>	Last column	Ending constituent column number or name.
<i>-m n</i>	2	Minimum number of mixture occurrences.
<i>-c value</i>	No recensoring	Global censoring value.
<i>-T directory</i>	Directory specified by environment variable <code>TMPDIR</code> , else directory specified by <code>“-d”</code>	Directory for creating temporary files.
<i>-v</i>	Missing	Treat contaminated values (V remark) as detections.
<i>-r file</i>	None	File name of the reference-data file.
<i>-p column</i>	None	Column name of the parameter-code column in the reference-data file.
<i>-b column</i>	None	Column name of the benchmark column in the reference-data file.
<i>-n column</i>	None	Column name of the alternate analyte name column in the reference-data file.

`“-v”` option, and the detection will be treated as a measured condition at the specified concentration. When both `“-v”` and `“-c”` are specified, the concentration is also subject to global recensoring.

The `“-r”` option is used to specify the name of a reference-data file containing benchmark concentrations and (or) alternate analyte names. If specified, then the `“-p”` option must also be specified to identify the name of the column in the reference-data file that contains the analytical-data column names. Once this linkage between the two files is established, then the `“-b”` option may be used to specify a column for benchmark concentrations, and (or) the `“-n”` option may be used to specify a column for renaming the analytes. All of the analytical-data columns specified for analysis are required to have a benchmark concentration when the `“-r”` and `“-b”` options are specified.

Many analytes have no respective benchmark. Either of two approaches may be used with the `qwmixtures.pl` program to process a dataset containing analytes both with and without benchmarks. One approach is to make one run with all analytes and not compute normalized concentrations and then make a second normalized-concentration run with only those analytes with benchmarks. An alternative approach is to add the missing benchmark rows to the reference-data file and specify relatively high benchmark concentrations. The normalized sum will be nominally affected if the benchmark concentration is set to a thousand or more times the maximum environmental concentration.

All of the output from the program is placed in a user-specified output directory, except for some temporary files that are placed in a user-specified temporary-file directory. Some files are created for initially enumerating and sorting the mixtures during the duration of the program run and then are deleted after successful completion. If the program is terminated for any reason while running, large temporary files may be left on disk and will need to be removed manually. The files created by the program are described in table 5. Examples of the `“report.txt”` and uncompressed `“mixtures.rdb”` files are presented in figures 5 and 6.

The file `“xlate.rdb”` (table 5) contains one row for each sample, with the same columns as the analytical data and with the rows in the same order; however, the remark-coded concentration-value cells are translated per the requirements of the mixture software. Remark codes are used to adjust concentration values in accordance with user-specified options. Recensoring is applied, if enabled. Nondetected analytes are translated to a zero concentration. Concentrations are normalized to benchmarks if reference values are supplied.

The `“raw.rdb.gz”` RDB-format file contains a column named `“mixture”` followed by all of the metadata columns from the analytical-data file. Every mixture enumerated from every sample is recorded in this file.

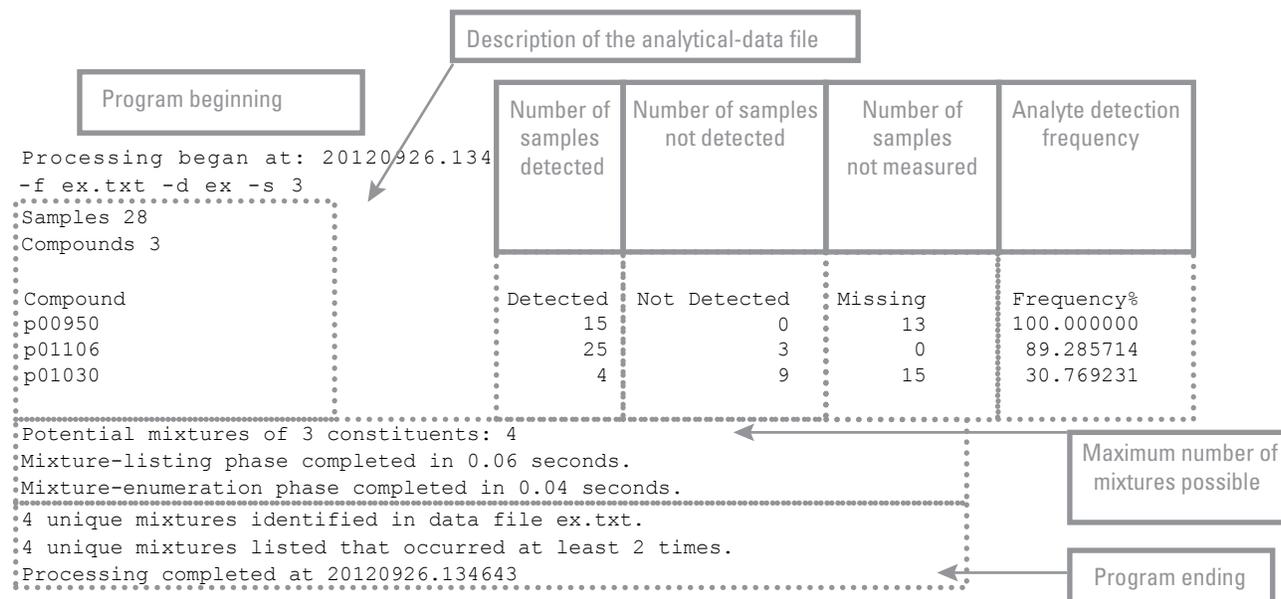
The `“mixtures.rdb.gz”` file contains four columns. The first column is named `“mixture,”` and each row identifies a different mixture. The second column is named `“count”` and contains the number of samples in which each mixture was found. The data rows are sorted in descending order of `“count.”` The third column is named `“expected%”` and contains the expected probability of the mixture occurring. The mixture expected probability is computed as the product of the mixture-component detection frequencies. The fourth column is named `“actual%”` and contains the actual occurrence frequency of each mixture in the analytical data supplied. The detection frequency of an individual analyte is computed as the ratio of the number of detections to the total number of measurements of the analyte. Similarly, mixture-detection frequency is the ratio of the number of detections to the total number of samples in which all mixture components were measured.

The mixture-enumeration program is limited by the hardware configuration of the host computer system. The available memory constrains enumeration of mixtures from

Table 5. Descriptions of files created by the mixture-enumeration program qwmixtures.pl.

[RDB, relational database; "?????" indicates characters in temporary file names that are varied by the program to generate unique file names]

File name	Description
report.txt	Fixed-length report that logs the mixture software run and analyte occurrence frequencies.
xlate.rdb	Translated analytical data.
mixtures.rdb.gz	Compressed RDB-format listing of unique mixtures ranked by mixture occurrence frequencies that is limited by the user minimum-occurrence specification.
raw.rdb.gz	Compressed RDB-format listing of every mixture identified in every input sample.
timing.rdb	RDB-format program-milestone timing data.
sample????	Temporary file used to enumerate mixtures in a sample.
merge????	Temporary file used to merge mixture lists among samples.
count????	Temporary file used to count unique mixtures.

**Figure 5.** Annotated example of a report.txt file from the mixture-enumeration program qwmixtures.pl.

mixture	<> count	<> expected%	<> actual%
40s	<> 7n	<> 9n	<> 9n
p00950+p01106	<> 13	<> 89.285714	<> 86.666667
p00950+p01030	<> 4	<> 30.769231	<> 44.444444
p00950+p01030+p01106	<> 4	<> 27.472527	<> 44.444444
p01030+p01106	<> 4	<> 27.472527	<> 30.769231

<>, Tab character

Figure 6. Example of an uncompressed mixtures.rdb file from the mixture-enumeration program qwmixtures.pl.

samples that exceed some maximum number of co-occurring analytes. The availability of disk space limits the number of mixtures that can be enumerated and sorted. The large number of combinations arising from samples with many detected compounds (fig. 1) causes the mixture software to use large quantities of memory and disk space. The limit on the number of compounds was extended during mixture software design and development by restructuring the computational procedures to conserve memory by shifting some intermediate results to disk files. That restructuring enables solution of problems with more samples and mixtures but requires additional disk space and slows the enumeration procedure for less complex problems. The specific limits and performance will depend on the characteristics of the particular computing platform where the mixture software is installed. Figure 7 presents performance data for some example input data, computer hardware, and operating systems.

The two most time-consuming functions of the mixture-enumeration program are listing all of the mixtures that occur in each sample and sorting the mixtures by decreasing occurrence frequency. The user can recensor detections to a higher detection level to reduce the number of enumerated mixtures. The user-specified recensoring limit (the “-c” option) is applied to all substances. This option might be appropriate if analytic methods provided an exceptionally low reporting level for a few analytes, substantially below concentrations of potential concern. Use of an external program, such as an RDB command, would be appropriate if different recensoring levels need to be applied to individual analytes in the input data. Run times may be improved by recensoring.

The user can reduce the amount of time spent sorting mixtures by eliminating mixtures that occur fewer than N times (by using the “-m” option; table 4) from the final tally.

Many of the water-mixture datasets used with the mixture software present many mixtures that occur only once or twice. Substantial run-time improvement can be achieved by using this option.

Program qwmixtrace.pl

The qwmixtrace.pl program is used to investigate mixture distributions. The program traces user-specified mixtures back to the analytical-sample data in which the mixtures were detected. Tracing mixtures to the source samples can be useful for compiling analyte-concentration ranges, associating the occurrence of specific mixtures with locations and environmental settings, and correlating specific mixtures to particular physical or chemical phenomena. An example command line for the qwmixtrace.pl program is shown next, and the command-line options are presented in table 6.

```
qwmixtrace.pl -f AsCdMix.rdb -d mixdir
               -m "P01002+P01027"
```

The mixtures to be traced are specified by listing the mixtures on the command line with one or more instances of the “-m” option. Alternatively, or in addition, users may trace all mixtures that occurred N or more times by specifying the “-n” option.

The qwtracemix.pl program scans the “raw.rdb.gz” file (created from a previous run of the qwmixtures.pl program; table 5) to find the samples for which the target mixture(s) were identified. If a compressed version of the file is not found, then the program attempts to open an uncompressed version of the file. The “mixtures.rdb.gz” file is read also if the user specifies the “-n” option for selecting mixtures to trace by minimum occurrence frequency.

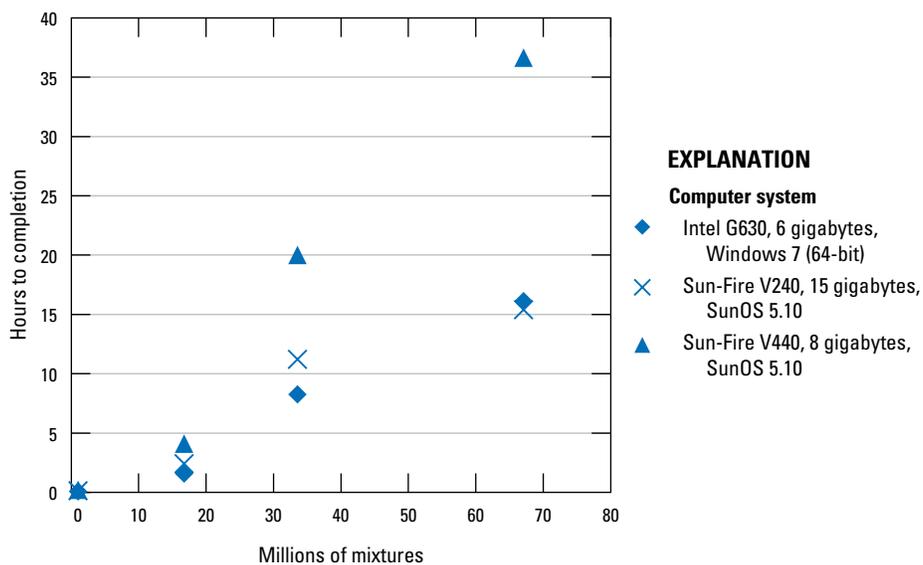


Figure 7. Mixture-enumeration-program completion time for all mixtures in a single sample with varying numbers of detected compounds and different computer hardware and operating systems.

Table 6. Command-line options for the mixture-tracing program `qwmixtrace.pl`.

[Italicized entries indicate user-specified data]

Option	Default	Description
Required		
<i>-f file.gz</i>	None	File name of a compressed RDB file to create containing sample metadata for the traced mixtures and placed in the directory specified by “-d.” The file will be overwritten, if a previous version already exists in the directory.
<i>-d directory</i>	None	Directory name for reading input and writing output files, previously used with the <code>qwmixtures.pl</code> program.
Optional (specify -m, -n, or both options)		
<i>-m “mixture”</i>	None	Mixture to trace in “quotes.” Repeat “-m” as necessary.
<i>-n occurrences</i>	None	Minimum number of mixture occurrences for a mixture to be traced.

The output file created by the `qwtracemix.pl` program is an RDB-formatted file with columns for the mixture name and all of the columns from the analytical-data file supplied to the `qwmixtures.pl` program that were not included in the mixture enumeration. These columns usually are metadata; however, if analytes were excluded from the `qwmixtures.pl` program (with the “-s” or the “-e” option), these analytes will also be written to the output file of the `qwmixtrace.pl` program.

There will be one row in the output file for each sample in which one of the user-specified mixtures has been traced to the analytical data. The user should be able to use the sample metadata columns to perform additional analyses for the target mixtures. The output file will be compressed, and a “.gz” suffix will be added if not already specified. If the output file already exists, the file will be replaced.

If multiple mixtures are traced to the same sample, then the sample metadata will be repeated in the output file on separate rows for each traced mixture. If the minimum (“-n”) occurrence frequency was never observed, and no individual (“-m”) mixtures were specified, an error message is printed and the program stops. If none of the analytical data match a user-specified (“-m”) mixture, however, no warning is printed. When the program completes, a message is printed showing how many mixture occurrences were traced.

Program `qwmixbq.pl`

The `qwmixbq.pl` program is used to evaluate possible mixture toxicity by using a benchmark-normalized additive

model. The approach used by the mixture software allows users to supply their own compound-specific benchmark-comparison concentrations, depending on their applications. The user’s assumptions about route of exposure, mode of action, target organ, species of concern, and other factors are incorporated into the user-supplied benchmark concentrations. A variety of human-health or aquatic-life benchmarks for individual chemicals can be used in the program, depending on the needs of the user. Nowell (2012) summarized the types and sources of water-quality benchmarks available for pesticides, but benchmarks also are available for other kinds of chemicals.

The sum of the benchmark-normalized concentrations is computed for each sample, and greater sums indicate greater relative possible toxicity. The sum can be used to screen for possible mixture toxicity and is useful as a relative index among the enumerated mixtures. Sums can indicate that one sample may be more or less toxic relative to another sample but do not indicate actual toxicity or probability of harm (Toccalino and others, 2012). Each sum is compared to one or more user-specified thresholds. The thresholds default to 0.1 and 1.0 if no specifications are made by the user. An example command line for the `qwmixbq.pl` program is shown next, and the command-line options are presented in table 7.

```
qwmixbq.pl -d mixdir -t 0.5 -t 1.0
```

Table 7. Command-line options for the mixture-toxicity-estimation program `qwmixbq.pl`.

[Italicized entries indicate user-specified data]

Option	Default	Description
Required		
<i>-d directory</i>	None	Directory name for reading input files, “xlate.rdb” and “mixtures.rdb.gz,” previously created by the <code>qwmixtures.pl</code> program, and writing output into file “bq.rdb.gz.”
Optional		
<i>-t threshold</i>	0.1 and 1.0	Exceedance threshold for the normalized-concentration sum. May be specified multiple times for additional thresholds.
<i>-T directory</i>	Directory specified by environment variable <code>TMPDIR</code> , else directory specified by “-d”	Directory for creating temporary files.

The `qwmixbq.pl` program reads the “`xlate.rdb`” and “`mixtures.rdb.gz`” files located in the user-specified (“-d”) directory and created from a previous run of the `qwmixtures.pl` program. The `qwmixtures.pl` program must have been run previously by using options to normalize concentrations to benchmarks (“-r,” “-p,” and “-b”). The “`mixtures.rdb.gz`” file is used as the list of mixtures to analyze. The “`xlate.rdb`” file is used to calculate the normalized-concentration sum for each mixture in every sample and count when the sum exceeds a threshold.

The `qwmixbq.pl` program creates a compressed, RDB-formatted file named “`bq.rdb.gz`” in the user-specified (“-d”) directory. There is one data row in the file for each mixture, and the rows are sorted by descending frequency of exceedance of the final user-specified threshold. The columns are described in table 8.

Three columns of information are computed for each user-supplied threshold (table 8). The first column is the number of samples in which the sum of the normalized concentrations exceeds the threshold. The second column presents the percentage of samples in which the sum exceeded the respective threshold. The third column counts samples

exceeding the threshold, as in the first of the three columns, except that samples are excluded from the count if any single mixture component concentration is greater than the respective benchmark. These counts exclude samples from the sum where threshold exceedance may be caused entirely by a single contaminant concentration. This sum determines the number of samples in which the threshold was exceeded because of the combined effects of several compounds.

The “`report.txt`” file is annotated when the `qwmixbq.pl` program is run, providing a record of the time required for the program to complete.

Program `qwmixsort.pl`

The program `qwmixsort.pl` is supplied as part of the mixture software for sorting a tab-delimited file, but the program is not intended to be run by the user. The program is run by each of the two mixtures programs: `qwmixtures.pl` and `qwmixbq.pl`. The command line for program `qwmixsort.pl` specifies the path name, file name, and sorting options for the file to be sorted.

Table 8. Descriptions of the contents of the compressed file `bq.rdb.gz` created by the mixture-toxicity-estimation program `qwmixbq.pl`.

[“x.x” indicates a numeric exceedance threshold, specified by the user]

Column name	Description
<code>mixture</code>	Mixture name formed by concatenation of the compound names delimited by a plus (“+”) symbol.
<code>analyzed</code>	Number of samples in which all mixture components were analyzed.
<code>present</code>	Number of samples in which all mixture components were detected.
<code>pct_present</code>	Mixture-detection frequency; percentage of samples in which all components were detected among samples in which all components were analyzed.
Three columns for each threshold (x.x)	
<code>count>x.x</code>	Number of samples in which the normalized-concentration sum exceeds the threshold.
<code>pct>x.x</code>	Percentage of samples exceeding the threshold among samples in which all the components were analyzed.
<code>sum>x.x</code>	Number of samples exceeding the threshold and having normalized concentrations below the threshold for all mixture components.

Summary and Conclusions

The composition, occurrence, distribution, and possible toxicity of chemical mixtures in the environment are research concerns of the U.S. Geological Survey and others. The National Water-Quality Assessment Program of the U.S. Geological Survey has developed chemical-mixture analysis software to facilitate assessment of mixtures. The mixture software includes four programs developed for research of the composition, occurrence, distribution, and possible toxicity of mixtures in environmental samples. The software works easily with data files retrieved from the U.S. Geological Survey National Water Information System (NWIS) and NWISWeb systems, as well as with appropriately formatted files from other sources. Installation and usage of the mixture software are documented. The mixture software was designed to function with minimal changes on a variety of computer-operating systems. The time needed for mixture analysis depends on the computer system used and the input data supplied and may require hours or days to complete. To obtain the software described herein and other U.S. Geological Survey software, visit <http://water.usgs.gov/software/>.

Chemical-mixture occurrences vary by analyte composition and concentration. Human-health outcomes ultimately are determined by single or multiple exposure routes from multiple media, such as consumption of food and water; inhalation of air; adsorption through contact with water, air, clothing, and plants; and medical treatments such as blood infusions, injections, and dermal applications. Similar exposures occur in aquatic toxicology. Mixtures ranked by frequency of occurrence or possible toxicity may be used to limit the scope of mixture research.

Mixtures may result from random co-occurrence of the component chemicals; however, the presence of specific chemical mixtures may serve as indicators of natural phenomena or human-caused events. Mixtures may provide key chemical signatures for identifying hydrologic events, geologic configurations, or human activities.

Chemical mixtures may have ecological, industrial, geochemical, or toxicological effects. Mixture analyses may be used as possible factors in epidemiological studies.

Future research opportunities exist for using the chemical-mixture analysis software documented herein. This mixture software can be used for any of these applications for limiting the scope of mixture research and evaluating the possible causes and effects of specific mixtures.

References Cited

Abbott, Alison, 2009, Toxicity testing gets a makeover: *Nature*, v. 461, no. 158, doi:10.1038/461158a, accessed August 17, 2012, at <http://www.nature.com/news/2009/090908/full/461158a.html>.

Altenburger, Rolf, Walter, Helge, and Grote, Matthias, 2004, What contributes to the combined effect of a complex chemical mixture?: *Environmental Science and Technology*, v. 38, no. 23, p. 6353–6362.

Backhaus, Thomas, and Faust, Michael, 2012, Predictive environmental risk assessment of chemical mixtures—A conceptual framework: *Environmental Science & Technology*, v. 46, no. 5, p. 2564–2573, doi:10.1021/es2034125.

Belden, J.B., Gilliom, R.J., and Lydy, M.J., 2007, How well can we predict the toxicity of pesticide mixtures to aquatic life?: *Integrated Environmental Assessment and Management*, v. 3, no. 3, p. 364–372, doi: 10.1002/ieam.5630030307.

Brian, J.V., Harris, C.A., Scholze, Martin, Backhaus, Thomas, Booy, Petra, Lamoree, Marja, Pojana, Guilio, Jonkers, Niels, Runnalls, Tamsin, Bonfá, Angela, Marcomini, Antonio, and Sumpter, J.P., 2005, Accurate prediction of the response of freshwater fish to a mixture of estrogenic chemicals: *Environmental Health Perspectives*, v. 113, no. 6, p. 721–728, doi:10.1289/ehp.7598.

Centers for Disease Control and Prevention, 2012, Fourth National report on human exposure to environmental chemicals updated tables: Centers for Disease Control and Prevention, February 2012, 235 p., accessed August 24, 2012, at http://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Feb2012.pdf.

DeSimone, L.A., 2009, Quality of water from domestic wells in principal aquifers of the United States, 1991–2004: U.S. Geological Survey Scientific Investigations Report 2008–5227, 139 p.

Eberl, D.D., 2008, User's guide to Handlens—A computer program that calculates the chemistry of minerals in mixtures: U.S. Geological Survey Open-File Report 2008–1244, 27 p.

Ehrenberg, Rachel, 2008, Defining toxic—Federal agencies look to cells, animals for chemical testing: *Science News*, v. 173, no. 8, p. 117, accessed November 28, 2008, at http://www.sciencenews.org/view/generic/id/9424/title/Defining_Toxic_Federal_agencies_look_to_cells%2C_not_animals%2C_for_chemical_testing.

Gade, A.L., Heiaas, Harald, Lillicrap, Adam, and Hylland, Ketil, 2012, Ecotoxicity of paint mixtures— Comparison between measured and calculated toxicity: *Science of the Total Environment*, v. 435–436, p. 526–540, doi:10.1016/j.scitotenv.2012.07.011.

- Gilliom, R.J., Barbash, J.E., Crawford, C.G., Hamilton, P.A., Martin, J.D., Nakagaki, Naomi, Nowell, L.H., Scott, J.C., Stackelberg, P.E., Thelin, G.P., and Wolock, D.M., 2006, Pesticides in the Nation's streams and ground water, 1992–2001: U.S. Geological Survey Circular 1291, 172 p.
- Granato, G.E., 2009, Computer programs for obtaining and analyzing daily mean streamflow data from the U.S. Geological Survey National Water Information System Web site: U.S. Geological Survey Open-File Report 2008–1362, 123 p., CD-ROM, 5 appendixes, accessed August 12, 2012, at <http://pubs.usgs.gov/of/2008/1362/>.
- Helwig, J.T., and Council, K.A., eds., 1979, Statistical analysis system user's guide (4th ed.): Raleigh, N.C., SAS Institute, 494 p.
- Hobbs, W.V., [1992?], RDB—A Relational Database management system: accessed August 12, 2012, at <http://compbio.soe.ucsc.edu/rdb/>.
- Jerius, Diab, 2004, RDB: accessed June 27, 2012, at <http://hea-www.harvard.edu/MST/simul/software/docs/rdb.html>.
- Kortenkamp, Andreas, 2008, Low dose mixture effects of endocrine disrupters—Implications for risk assessment and epidemiology: *International Journal of Andrology*, v. 31, p. 233–240, doi: 10.1111/j.1365-2605.2007.00862.x.
- Lydy, M.J., and Belden, J.B., 2006, Assessing potential effects of pesticide mixtures, *in* Pesticides in the Nation's streams and ground water, 1992–2001: U.S. Geological Survey Circular 1291, p. 114–117.
- Manis, Rod, Schaffer, Evan, and Jorgensen, Robert, 1988, UNIX relational database management—Application development in the UNIX environment: Englewood Cliffs, N.J., Prentice Hall, 476 p.
- Mansoori, G.A., 2002, Physicochemical basis of fouling prediction and prevention in the process industry: *Journal of the Chinese Institute of Chemical Engineers*, v. 33, no. 1, p. 25–31.
- Masoner, Jason, Kolpin, Dana, Cozzarelli, Isabelle, Furlong, Edward, Gray, James, Zaugg, Steve, and Foreman, Bill, 2012, Characterizing landfill leachate in the conterminous United States—Contaminants of emerging concern and geochemistry, *in* Proceedings of the American Water Resources Association Annual Water Resources Conference, June 25–27, 2012, Denver, Colo., accessed October 2, 2012, at http://www.awra.org/proceedings/Summer2012/CEC/doc/abs/JasonMasoner_be99917a_8160.pdf.
- Mendenhall, William, 1975, Introduction to probability and statistics (4th ed.): Boston, Mass., Duxbury Press, 460 p.
- Moulton, S.R. II, Kennen, J.G., Goldstein, R.M., and Hambrook, J.A., 2002, Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 02–150, 75 p.
- National Research Council, 2009, Toward a sustainable and secure water future—A leadership role for the U.S. Geological Survey: Washington, D.C., The National Academies Press, 128 p.
- Nesbitt, Murray, 2012, PPM, Perl Package Manager—Locate, install, upgrade software packages: accessed October 3, 2012, at <http://search.cpan.org/~murray/PPM-2.1.9/bin/ppm.pl>.
- Net Applications, 2013, Desktop Operating System Share: accessed February 27, 2013 at <http://www.netmarketshare.com>.
- Noria, Xavier, 2006, Understanding newlines: accessed on August 16, 2012, at <http://onlamp.com/pub/a/onlamp/2006/08/17/understanding-newlines.html>.
- Nowell, Lisa, 2012, Types and sources of water-quality benchmarks for pesticides: U.S. Geological Survey, accessed September 19, 2012, at <http://water.usgs.gov/nawqa/pnsp/benchmarks/source.html>.
- Pape-Lindstrom, P.A., and Lydy, M.J., 1997, Synergistic toxicity of atrazine and organophosphate insecticides contravenes the response addition mixture model: *Environmental Toxicology and Chemistry*, v. 16, no. 11, p. 2415–2420.
- Perl.org, 2012, Perl—Download: accessed on August 15, 2012, at <http://www.perl.org/get.html>.
- Perl.org, 2013, Perl licensing: accessed on February 27, 2013 at <http://dev.perl.org/licenses/>.
- Peterson, B.J., Howarth, R.W., and Garritt, R.H., 1985, Multiple stable isotopes used to trace the flow of organic matter in estuarine food webs: *Science*, v. 227, no. 4692, p. 1361–1363, doi:10.1126/science.227.4692.1361.
- Phillips, P.J., Nowell, L.H., Gilliom, R.J., Nakagaki, Naomi, Murray, K.R., and VanAlstyne, Carolyn, 2010, Composition, distribution, and potential toxicity of organochlorine mixtures in bed sediments of streams: *Science of the Total Environment*, v. 408, no. 3, p. 594–606.
- Pickrell, John, 2002, Killer cocktails—Drug mixes threaten aquatic ecosystems: *Science News*, v. 162, no. 7.
- P-STAT, Inc, 2013, Welcome to P-STAT: accessed February 27, 2013 at <http://www.pstat.com/>.
- Ryker, S.J., and Small, M.J., 2008, Combining occurrence and toxicity information to identify priorities for drinking-water mixture research: *Risk Assessment*, v. 28, no. 3, p. 653–666, doi:10.1111/j.1539-6924.2008.00985.x.
- Schornick, J.C., Pruitt, J.B., Stranathan, H.E., and Duncan, A.C., 1989, An indexing and classification system for earth-science data bases: U.S. Geological Survey Open-File Report 89–47, 56 p.

- Schwarzenbach, R.P., Escher, B.I., Fenner, Kathrin, Hofstetter, T.B., Johnson, C.A., von Gunten, Urs, and Wehrli, Bernhard, 2006, The challenge of micropollutants in aquatic systems: *Science*, v. 313, no. 5790, p. 1072–1077.
- Squillace, P.J., and Moran, M.J., 2006, Factors associated with sources, transport, and fate of volatile organic compounds in aquifers of the United States and implications for ground-water management and assessments: U.S. Geological Survey Scientific Investigations Report 2005–5269, 40 p.
- Squillace, P.J., Scott, J.C., Moran, M.J., Nolan, B.T., and Kolpin, D.W., 2002, VOCs, pesticides, nitrate, and their mixtures in groundwater used for drinking water in the United States: *Environmental Science & Technology*, v. 36, no. 9, p. 1923–1930.
- Stackelberg, P.E., Kauffman, L.J., Ayers, M.A., and Baehr, A.L., 2001, Frequently co-occurring pesticides and volatile organic compounds in public supply and monitoring wells, southern New Jersey, USA: *Environmental Toxicology and Chemistry*, v. 20, no. 4, p. 853–865.
- Teuschler, L.K., 2007, Deciding which chemical mixtures risk assessment methods work best for what mixtures: *Toxicology and Applied Pharmacology*, v. 223, no. 2, p. 139–147.
- Tocalino, P.L., Norman, J.E., and Scott, J.C., 2012, Chemical mixtures in untreated water from public-supply wells in the U.S.—Occurrence, composition, and potential toxicity: *Science of the Total Environment*, v. 431, p. 262–270, doi: 10.1016/j.scitotenv.2012.05.044.
- U.S. Department of Agriculture, 2012, Pesticide Data Program—Annual summary, calendar year 2010: accessed August 23, 2012, at <http://www.ams.usda.gov/AMSV1.0/getfile?dDocName=stelprdc5098550>.
- U.S. Environmental Protection Agency, 1986, Guidelines for the health risk assessment of chemical mixtures: *Federal Register*, v. 51, no. 185, p. 34014–34025, accessed September 26, 2012, at http://www.epa.gov/raf/publications/pdfs/CHEMMIX_1986.PDF.
- U.S. Environmental Protection Agency, 2000, Supplementary guidance for conducting health risk assessment of chemical mixtures: accessed September 26, 2012, at http://www.epa.gov/raf/publications/pdfs/CHEM_MIX_08_2001.PDF.
- U.S. Environmental Protection Agency, 2012a, National Primary Drinking Water Regulations: accessed August 15, 2012, at <http://www.epa.gov/safewater/contaminants/index.html>.
- U.S. Environmental Protection Agency, 2012b, Technology Transfer Network (TTN) Air Quality System (AQS) Data Mart: accessed August 23, 2012, at http://www.epa.gov/ttn/airs/aqsdatamart/basic_info.htm.
- U.S. Geological Survey, 1998, National Water Information System: U.S. Geological Survey Fact Sheet 027–98, 2 p.
- U.S. Geological Survey, 2002, NWISWeb—New site for the Nation’s water data: U.S. Geological Survey Fact Sheet 128–02, 2 p.
- U.S. Geological Survey, 2006, User’s manual for the National Water Information System of the U.S. Geological Survey Water-Quality System: U.S. Geological Survey Open-File Report 2006–1145, 381 p.
- U.S. Geological Survey, 2012, Parameter Code Definition: accessed August 12, 2012, at <http://nwis.waterdata.usgs.gov/usa/nwis/pmcodes>.
- Wall, Larry, Christiansen, Tom, and Schwartz, R.L., 1996, *Programming Perl* (2d ed.): Sebastopol, Calif., O’Reilly and Associates, 646 p.
- Ziv, Jacob, and Lempel, Abraham, 1978, Compression of individual sequences via variable-rate coding: *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, v. 24, no. 5, p. 530–536.

Appendix 1. Messages Printed by the Chemical-Mixture Analysis Software

The chemical-mixture analysis software (subsequently, mixture software) will print three different categories of messages. Error messages will cause the mixture software to stop the processing of mixtures until problems have been corrected. Warning messages will alert the user to possible issues; however, the mixture software will continue to run. Informative messages are printed to provide indicators of the program status; no corrective actions are likely to be necessary, and therefore informative messages are not described herein. Error and warning messages are listed here

(tables 1–1 and 1–2) with possible corrective actions that may be necessary. The mixture software is designed to identify as many possible error and warning conditions that are relevant and then print messages and stop if necessary. The comingling of printed messages with dissimilar causes may be confusing initially; however, the comprehensive approach prevents the need to identify and correct iteratively all sources of error. Text in italics below will be replaced with meaningful text when the message is printed while using the mixture software.

Table 1–1. Error messages printed by the mixture software.

Message:	Benchmark for <i>column</i> is zero.
Action:	Change benchmark concentration for the specified <i>column</i> in the reference data file to a non-zero value, or omit the chemical from the benchmark-normalized analysis.
Programs:	qwmixtures.pl
Message:	Directory is not writeable: <i>directory</i> .
Action:	Change the permissions to the <i>directory</i> or specify a different directory after the “-d” option.
Programs:	qwmixtures.pl, qwmixtrace.pl
Message:	End column not found: <i>column</i> .
Action:	The column specified after “-e” was not found in the analytical-data file. Specify a different column name or input-data file name.
Programs:	qwmixtures.pl
Message:	Environment variable QWMIXDIR is not properly set.
Action:	The environment variable QWMIXDIR is set to a path that is not a directory containing the program qwmixsort.pl. Reset the environment variable.
Programs:	qwmixtures.pl, qwmixbq.pl
Message:	Fatal flaws with reference-data file prevent further evaluation.
Action:	Preceding errors messages should indicate the problem to solve.
Programs:	qwmixtures.pl
Message:	In reference-data file -r <i>filename</i> , the -n <i>column</i> does not exist.
Action:	The name column specified after the “-n” option does not exist in the reference-data file specified after the “-r” option. Specify a different column name or a different file name.
Programs:	qwmixtures.pl
Message:	In reference-data file -r <i>filename</i> , the -b <i>column</i> does not exist.
Action:	The benchmark column specified after the “-b” option does not exist in the reference-data file specified after the “-r” option. Specify a different column name or a different file name.
Programs:	qwmixtures.pl
Message:	Input file not found: <i>filename</i> .
Action:	Change the filename specified after the “-f” option.
Programs:	qwmixtures.pl
Message:	Input file not readable: <i>filename</i> .
Action:	Change the filename or access permissions to the file specified.
Programs:	qwmixtures.pl

Table 1-1. Error messages printed by the mixture software.—Continued

Message:	Missing required argument: -d directory.
Action:	Add the required option to the command line.
Programs:	qwmixtures.pl, qwmixtrace.pl, qwmixbq.pl
Message:	Missing required argument: -f filename.
Action:	Add the required option to the command line.
Programs:	qwmixtures.pl, qwmixtrace.pl
Message:	Missing required argument: -m mixture.
Action:	Add the “-m” and (or) the “-n” options to the command line.
Programs:	qwmixtrace.pl
Message:	Mixtures were not normalized to benchmarks in xlate.rdb.
Action:	Choose a different (“-d”) directory, or rerun the qwmixtures.pl program and supply a (“-r”) reference-data file and (“-b”) benchmarks option.
Programs:	qwmixbq.pl
Message:	Name for <i>column</i> is missing.
Action:	The row for the specified analytical-data column does not contain a compound name in the specified column in the (“-r”) user-supplied reference-data file. Add a column name to that row, select a different naming column (“-n”), or omit the reference-data file.
Programs:	qwmixtures.pl
Message:	No mixtures occurring $\geq N$ found in <i>filename</i> .
Action:	Choose a lower minimum value for the -m option, or select a different directory corresponding to an earlier run of program qwmixtures.pl. Otherwise, no mixtures are traced.
Programs:	qwmixtrace.pl
Message:	No -n column or -b column for -r <i>filename</i> specified.
Action:	Specify the “-n” and (or) the “-b” options when using the “-r” option. Otherwise, omit the “-r” option.
Programs:	qwmixtures.pl
Message:	No such directory: <i>directory</i> .
Action:	Choose a different directory name which corresponds to an earlier run of the qwmixtures.pl program because the specified directory name does not exist.
Programs:	qwmixtrace.pl, qwmixbq.pl
Message:	<i>N</i> nonpositive thresholds supplied in: <i>thresholds</i> .
Action:	Alter the specifications of the “-t” threshold option.
Programs:	qwmixbq.pl
Message:	Not a directory: <i>directory</i> .
Action:	The name specified after the “-d” option exists but is not a directory. Specify a different name for the directory.
Programs:	qwmixtures.pl, qwmixtrace.pl, qwmixbq.pl
Message:	-p parameter column required with -r reference-data file.
Action:	A reference-data file was supplied (“-r”) without the “-p” option to specify the column name of the analyte in the analytical-data file. Add the “-p” option, or omit the reference-file option.
Programs:	qwmixtures.pl
Message:	-p <i>column_name</i> column not found in reference-data file -r <i>filename</i> .
Action:	The reference-data file supplied with the (“-r”) option does not contain the analyte-column name specified by the “-p” option. Alter the “-p” column name, specify a different reference-file name, or omit the reference-file option.
Programs:	qwmixtures.pl

Table 1-1. Error messages printed by the mixture software.—Continued

Message:	Reference-data file: <i>Filename</i> does not contain benchmarks for compounds: <i>column_a</i> , <i>column_b</i> .
Action:	Columns of analytical-data file included in the contiguous block of columns for mixture analysis (“-b,” “-e”) do not have benchmarks specified in the supplied (“-r”) reference file. Alter the block of analyte columns, supply the missing benchmarks in the reference-data file, use different reference and (or) analytical-data files, or add artificial benchmark rows with relatively high concentrations.
Programs:	qwmixtures.pl
Message:	Start column not found: <i>column</i> .
Action:	The initial column name given after “-b” for the block of analytical columns does not exist in the (“-f”) analytical-data file. Specify a different column name, or a different analytical-data file.
Programs:	qwmixtures.pl
Message:	Too many compounds <i>N</i> for sample <i>identifier</i> .
Action:	The computational limit has been exceeded by a sample in the analytical-data file with <i>identifier</i> in the first column. The number of detections in that sample is <i>N</i> , which exceeds twice the value of the environment variable QWMIXLIMIT. Omit the sample, or alter the included analytes. Reset the variable QWMIXLIMIT only after reviewing the “Installation” section of this report.
Programs:	qwmixtures.pl
Message:	Unable to create directory: <i>directory</i> .
Action:	Alter the access permissions, or choose a different directory.
Programs:	qwmixtures.pl
Message:	Unable to find input file: <i>filename</i> .
Action:	Choose a different directory that corresponds to an earlier run of the qwmixtures.pl program.
Programs:	qwmixtrace.pl, qwmixbq.pl
Message:	Unable to open <i>filename</i> for writing.
Action:	Choose a different directory, alter the access permissions to the directory, or copy the directory to a new location. The program was unable to write to the named output file.
Programs:	qwmixtures.pl, qwmixtrace.pl, qwmixbq.pl
Message:	Unable to open compressed file during enumeration: <i>filename</i> .
Action:	Specify a different TMPDIR, alter system open-file limit, delete unneeded files, or reduce the size of the input dataset.
Programs:	qwmixtures.pl
Message:	Unable to open reference-data file: <i>filename</i> .
Action:	Specify a different filename after the “-r” option, or alter the access permissions.
Programs:	qwmixtures.pl
Message:	Unable to open temp file during enumeration in <i>directory</i> .
Action:	Specify a different TMPDIR, alter system open-file limit, delete unneeded files, or reduce the size of the input dataset.
Programs:	qwmixtures.pl
Message:	Unexpected arguments: <i>arguments</i> .
Action:	Correct or remove the unexpected option specifications (arguments) from the command line.
Programs:	qwmixtures.pl, qwmixtrace.pl, qwmixbq.pl

Table 1–2. Warning messages printed by the mixture software.

Message:	Added “.gz” suffix to file name <i>filename</i> .
Explanation:	The specified (“-f”) file is compressed, and the “.gz” suffix is appended to the name. Decompression software will be needed to read the file.
Programs:	qwmixtrace.pl
Message:	Creating nonexistent directory: <i>directory</i> .
Explanation:	The <i>directory</i> specified with the “-d” option did not exist.
Programs:	qwmixtures.pl
Message:	Environment variable QWMIXLIMIT not properly set, defaults to <i>limit</i> .
Explanation:	A missing or invalid value was detected for the environment variable QWMIXLIMIT.
Programs:	qwmixtures.pl
Message:	M & N remarked results will be censored to non-detections.
Explanation:	Nonquantified detections are treated as nondetections when the “-c” option is used to censor data.
Programs:	qwmixtures.pl
Message:	No reference-data file supplied, arguments: -b, -n, -p are inapplicable.
Explanation:	No “-r” option was supplied.
Programs:	qwmixtures.pl
Message:	Overwriting existing file: <i>filename</i> .
Explanation:	The qwmixtrace.pl program found an existing file matching the file name specified with the “-f” option.
Programs:	qwmixtrace.pl
Message:	Setting minimum mixture occurrence count to 2.
Explanation:	The default value for the “-m” option was assigned, so mixtures occurring only once will be omitted from the mixtures.rdb.gz file.
Programs:	qwmixtures.pl
Message:	Temporary-file directory does not exist: <i>directory</i> .
Explanation:	The directory specified by the environment variable TMPDIR or the command-line “-T” option was not found. The output directory will be used for temporary files.
Programs:	qwmixtures.pl, qwmixbq.pl
Message:	Temporary-file directory not writeable: <i>directory</i> .
Explanation:	The directory specified by the environment variable TMPDIR or the command-line “-T” option was not found. The output directory will be used for temporary files.
Programs:	qwmixtures.pl, qwmixbq.pl
Message:	Using temporary-file directory: <i>directory</i> .
Explanation:	The mixture software warns where potentially large temporary files will be written.
Programs:	qwmixtures.pl, qwmixbq.pl

Appendix 2. Example Installation Procedure for the Chemical-Mixture Analysis Software

The installation procedure for the chemical-mixture analysis software (subsequently, mixture software) will vary for different computer-operating systems and will probably change as operating systems evolve. This appendix presents an installation procedure that was successful for two computer-operating systems at the time of this report. The example installation presented here should be used as a guide to the procedural requirements rather than as a precise step-by-step procedure. Administrator privileges may be needed to complete the installation steps.

1. Determine if the Perl language is installed. From a command prompt, type the following.

```
perl -e 'print "Hello World\n" '
```

2. If “Hello World” is not printed, then Perl needs to be installed. Install a Perl version appropriate for the computer-operating system, such as one available from <http://www.perl.org/get.html>.
3. Determine if any additional Perl packages are needed. At a command prompt, type the following (all one command).

```
perl -e 'use File::Spec::Functions;
use File::MergeSort; use File::Sort;
use File::Temp; use File::Basename;
use Sort::External; use IO::Zlib; use
Math::Combinatorics; use Math::BigInt;
use IO::Pipe; '
```

If no error messages are printed, then no additional Perl packages are needed. Any messages like the following indicate that one or more packages are needed.

```
Can't locate File/MergeSort.pm in @
INC ...
```

For example, the message above indicates that package “File::MergeSort” is needed.

4. Install all Perl packages needed. Repeat the procedure in this step for each needed Perl package identified in the previous step. The package installation procedure varies, depending on the computer-operating system and Perl version. If the Perl Package Manager (ppm) is installed, then a simple command should obtain and install the requisite package (one per command). To install the “File::MergeSort” package by using ppm, the following command would be used:

```
ppm File::MergeSort
```

5. A different procedure should be used for systems without the Perl Package Manager.

- a. Find and download the package from the Comprehensive Perl Archive at <http://search.cpan.org/>.
- b. Extract the files from the downloaded package, for example.

```
gzcat File-MergeSort-1.23.tar.gz | tar
-xf -
```

- c. Test and install the package by using the example commands below (also documented in the package “README” file).

```
cd File-MergeSort-1.23
perl Makefile.PL
make
make test
make install
```

6. Create a directory for the mixture source-code files, and install the files into that directory. Download the “chemical-mixture analysis” software from <http://water.usgs.gov/software/>, and store the files (four programs: qwmixtures.pl, qwmixtrace.pl, qwmixbq.pl, and qwmixsort.pl) in the selected directory. Set permissions that allow the programs to be run, if in a UNIX environment.

```
chmod +x qwmixtures.pl qwmixtrace.pl
qwmixbq.pl qwmixsort.pl
```

7. Set environment variables by using a method appropriate to the computer-operating system.
 - a. For example, a UNIX C-shell user would edit the “.cshrc” file in their home directory and add lines to set the three mixture-software environment variables and modify the system PATH variable to include the mixture-software directory. The maximum number of open files must be increased for some UNIX installations.

```
setenv QWMIXLIMIT 24
setenv QWMIXDIR /usr/local/qwmixtures
setenv TMPDIR /tmp
set path = ($QWMIXDIR $path)
limit descriptors 2048
```

- b. A Microsoft Windows user would use “Control Panel” and “System” options to set these same four variables, ensuring that the PATH variable is appended, not overwritten.

ISBN 978-1-4113-3641-4



9 781411 336414