GEOCHEMICAL SURVEY OF MISSOURI

Methods of Sampling, Laboratory Analysis, and Statistical Reduction of Data

Geochemical Survey of Missouri– Methods of Sampling, Laboratory Analysis, and Statistical Reduction of Data

By A. T. MIESCH
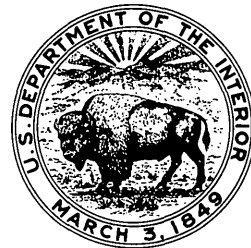
With sections on LABORATORY METHODS

By P. R. BARNETT, A. J. BARTEL, J. I. DINNIN, G. L. FEDER, T. F. HARMS, CLAUDE HUFFMAN, JR., V. J. JANZER, H. T. MILLARD, JR., HARRIET G. NEIMAN, M. W. SKOUGSTAD, *and* J. S. WAHLBERG

GEOCHEMICAL SURVEY OF MISSOURI

GEOLOGICAL SURVEY PROFESSIONAL PAPER 954 - A

Methods used in the collection, analysis, and interpretation of data in a regional geochemical survey

UNITED STATES DEPARTMENT OF THE INTERIOR

THOMAS S. KLEPPE, *Secretary*

GEOLOGICAL SURVEY

V. E. McKelvey, *Director*

CONTENTS

ILLUSTRATIONS

TABLES

GEOCHEMICAL SURVEY OF MISSOURI—METHODS OF SAMPLING, LABORATORY ANALYSIS, AND STATISTICAL REDUCTION OF DATA

By A. T. Miesch

ABSTRACT

A reconnaissance geochemical survey of the State of Missouri was made primarily to develop and explore methods for conducting environmental geochemical surveys of large regions. Another important purpose was to provide information needed by epidemiologists in search of relationships between the geochemistry of the environment and health. The information will also prove useful in studies of environmental pollution and, perhaps, in the search for unknown mineral deposits.

The geochemical survey included bedrock, unconsolidated surficial deposits, agriculture and currently uncultivated soils, selected native vegetation and farm crops, and ground and surface water. Approximately 7,000 samples of these materials were collected and each sample was analyzed for 30-40 chemical constituents. Nearly all of the sampling was according to hierarchical designs and involved complete randomization in the selection of sampling points. The samples from each phase of the survey were analyzed in completely randomized sequences, including duplicate samples unknown to the analysts, principally by methods of emission spectroscopy, X-ray fluorescence, and atomic absorption. Results indicate that the total experimental error in this data is, in general, more a function of sampling than of laboratory analysis.

The laboratory and field data pertaining to the collected samples were stored in the computer-based U.S. Geological Survey Rock Analysis Storage System (RASS) and were selectively retrieved onto magnetic tapes and disks for statistical reduction using programs of the U.S. Geological Survey Statistical Package (STATPAC). The principal statistical methods used were analysis of variance, Duncan's test for significance of differences among means, and Q-mode factor analysis.

The ultimate objective of the statistical analysis was to identify geochemical variations over the State that are unlikely to have resulted from accidents of sampling or laboratory treatment. Most of these variations are described by means of geochemical maps.

Q-mode factor maps serve to summarize the data and to provide bases for geochemical interpretations of the processes that created the variation.

INTRODUCTION

A reconnaissance geochemical survey of the State of Missouri was undertaken during the years 1969–73. The purpose of the survey was to describe the major geochemical variations over the State. Specifically, it has been directed at compositional variations in bedrock, unconsolidated surficial deposits, uncultivated and agricultural soils, selected native vegetation and farm crops, and both ground water and surface water. Because of the large area of concern and the variety of materials examined, and because of the anticipated need for future investigations of this kind in other large areas of the United States, special concern has been given to the problem of efficiency. A concerted effort has been made to obtain the maximum information at the lowest possible cost. This effort required extensive use of formal experimental design procedures in planning and executing sampling programs in order to reduce the fieldwork and the number of samples to be analyzed in the laboratories. Statistical methods were employed to a more than ordinary extent in identifying variation that is unlikely to have resulted from errors in sampling or laboratory analysis.

The geochemical survey of Missouri has been a pilot study designed to explore and develop field and laboratory methods for large-scale geochemical surveys. Missouri was selected for the pilot investigation partly because of the immediate need for geochemical information by the Environmental Health Surveillance and Research Center (EHSRC) of the University of Missouri at Columbia. Research at EHSRC has been aimed at the relation between health and the character of the environment. The geochemical character of the environment has been of special interest owing to a growing awareness among medical scientists of the importance of trace substances in human and animal health. (See, for example, Hemphill, 1967, 1969, 1970, 1971, 1972, and 1973.) The bearing of geochemistry on the problem and the interest of medical scientists in the geochemical environment have been described by Hopps (1971).

The broad goals and general methods being used in the epidemiological research of the EHSRC were reviewed by Marienfeld (1972). One of the principal products of the research is a series of maps showing variations in the rates

of human and animal diseases over the entire State. The search for causes of disease consists, in part, of a search for correlation between disease rates and environmental factors. The principal objective of the geochemical survey of Missouri, therefore, has been to produce statewide maps of the geochemical variations.

The subject matter scope of the survey—that is, the inclusion of rocks, soils, plants, and water—has been made broad because of the present elementary state of knowledge concerning relations between health and the geochemistry of the environment. Ground water and surface water, as well as farm crops and agricultural soils, have been included for obvious reasons. The bedrock and unconsolidated surficial deposits are included because these materials are the natural sources of chemical constituents in water, soil, and vegetation (Miesch, 1972). Selected species of native vegetation are included in order to search for indicators of not only the gross chemical character of the environment, but also the availability of the chemical elements to organisms. Uncultivated soils are included because they are the link between native vegetation and the underlying geology.

Although the development of methods for conducting geochemical surveys and support of epidemiological research have been the primary purposes for the geochemical survey of Missouri, there are a number of other important purposes that this survey, and similar surveys of other large regions, may serve. It has become abundantly clear in recent investigations of environmental pollution that the identification of pollution and the delineation of its extent, presuppose a knowledge of the normal or natural state of the environment. The natural environment is very far from being geochemically uniform, and the presence of an element, or suite of elements, in sufficient abundance to be indicative of environmental pollution in one area may be entirely normal for another. Thus, the same geochemical maps provided to support epidemiological research may be useful in defining the needed baselines in investigations of pollution. They have already proved useful in at least one instance (Ebens and others, 1973).

The results of the survey may also find application in investigations and research unrelated, or only marginally related, to concern for the environment. For example, the baseline data of the environmentalist are the background data of the geochemical explorationist, and the survey results may prove useful in mineral exploration. The geochemical explorationist must presuppose a normal range of concentrations for various elements in the media he is sampling before he can identify the anomalies he is seeking. Aside from this, geochemists engaged in a wide range of research activities have continuing need for information on the occurrences and distribution of elements in natural materials.

This report serves as an introduction to a series of specific reports on various phases of the survey. It describes the field and laboratory methods and the methods of statistical analysis and data reduction that are common to more than one of these phases. These methods have been held consistent wherever possible so that variations in the data, as summarized on geochemical maps and in statistical tables, will reflect actual variations in the environment rather than variations in the field and laboratory methods that were used.

Preliminary results of the geochemical survey of Missouri have been released during the course of the work in a series of semiannual progress reports (U.S. Geological Survey, 1972a–f, 1973). The purpose and methods being used were summarized by Connor, Feder, Erdman, and Tidball (1972), and some preliminary findings were described by Tidball (1971), Connor, Erdman, Sims, and Ebens (1971), and Connor, Shacklette, and Erdman (1971).

Feder (1972) described the sampling methods used in his study of ground and surface waters, and Feder, Ebens, and Connor (1972) described some relations between the composition of ground water and the composition of the host rock.

Tidball and Sauer (1972) have discussed some preliminary findings on the relations between soils and human health on a statewide basis. One small area in central Missouri has been studied intensively by Ebens and others (1973) and by Ebens, Erdman, and Feder (1972). The medical aspects of the investigations were discussed more fully by Case and others (1972, 1973).

ACKNOWLEDGMENTS

and Director, and Larry D. Fellows, Assistant State Geologist.

The success of all phases of the geochemical survey depended heavily on the support and cooperation of U.S. Geological Survey analysts and those in charge of the major chemical laboratories. Special acknowledgment is made of the cooperation extended by Frank S. Grimaldi, Irving May, Leonard B. Riley, Claude Huffman, Jr., Marvin W. Skougstad, and Frederick N. Ward.

SOME FEATURES OF MISSOURI AND ITS LANDSCAPE

Missouri has an area of 180,500 square kilometres and a population of 4.3 million. Nearly two-thirds of the population, however, is concentrated in two large metropolitan areas, St. Louis and Kansas City. The remainder of the population is in smaller cities, rural communities, and farms widely scattered throughout the State.

Northern Missouri is part of the central interior lowland of the United States, and most of southern Missouri occurs within the Ozark Plateaus region of the Interior Highlands (Fenneman, 1938). Six physiographic regions within the State are recognized by Vineyard (1967), and each of them has boundaries which coincide, to varying degrees, with those that have been drawn on the basis of geology, soils, and vegetation (fig. 1). The domal uplift centered around the St. Francois Mountains in the southeastern part of the State is the dominant control on the nature of the landscape. Precambrian igneous rocks, both intrusive and extrusive, are exposed in the core of the uplift, and sediments of Paleozoic age dip gently away from the core in all directions. Much of the southern part of the State is underlain by westward- and northwestward-dipping beds of dolomite, limestone, and sandstone of Cambrian and Early Ordovician age, whereas nearly all of the north half is underlain by northwestward-dipping beds of shale, limestone, and sandstone of Mississippian and Pennsylvanian age. Flood-plain deposits of alluvium occur along the Mississippi and Missouri Rivers and in the extreme southeastern part of the State where they compose part of the extensive Mississippi embayment deposits. Claystone, sandstone, and gravel of Cretaceous and Tertiary age are exposed locally within the southeastern flood-plain area. Deposits of glacial till cover most of the northern half of the State, severely restricting exposures of the sedimentary rocks of Mississippian and Pennsylvanian age. Deposits of windblown silt (loess) are prominent along the banks of the major rivers and they occur elsewhere throughout the State. In the southern part of the State, where glacial till is absent, a residuum of reddish clay and silt as much as several hundred feet thick has accumulated on the carbonate rocks. The glacial deposits in the northern part of the State and the reddish

residuum in the southern part are the two most widespread and prominent types of geologic deposits, and are the major parent materials from which the soils have developed.

The soils of Missouri are highly diverse in specific type; about 275 different soil units (Series) have been mapped. However, more than 90 percent of the State is covered by only three broad types (Orders)—Mollisols, Alfisols, and Ultisols. The Mollisols contain black friable organic material in their upper horizons and occur both along the flood plains of the Mississippi and Missouri Rivers and over the glacial till and loess deposits in the northwestern part of the State. The Alfisols occur in broad areas along the major rivers above the flood plains. The Ultisols are developed principally on the dolomites in the Ozark Plateaus region in the southern part of the State. There are only a few small areas in the State that are believed to never have been under cultivation, and it is likely that the chemical and physical character of all soils—those at present supporting agriculture and those currently uncultivated—have been modified by man to various degrees, depending on various agricultural techniques employed.

The Missouri landscape as mapped by Küchler (1964) according to the potential natural vegetation the land would support if reverted to nature falls into five categories. These categories—vegetation-type areas in this series of reports—are: (1) a mixed bluestem prairie and Oak-Hickory Forest at the higher elevations in the north half of the State; (2) an Oak-Hickory Forest in most of the central part of the State and at lower elevations, along the rivers, in the northwestern part, and (3) an Oak-Hickory-Pine Forest in the southeast, (4) a Floodplain Forest along the Mississippi River in the extreme southeastern part of the State, and (5) Cedar Glades immediately north of the Arkansas border in the southwest (fig. 1). The first category has been subdivided for the geochemical survey into two vegetation-type areas—(1) Glaciated Prairie north of the Missouri River and (2) Unglaciated Prairie south of the river. This subdivision was made because of the fundamentally different types of surficial geologic materials underlying the two parts of the region. Thus, six different vegetation-type areas have been studied and compared. Agriculture is greatly limited in two of these areas, the Cedar Glades and the Oak-Hickory-Pine Forest, but is extensive in parts of the other four. The principal crops are corn, soybeans, and wheat, with some production of cotton and rice in the southeastern flood-plain area. Livestock production, mostly cattle and swine, is widespread except in the southeastern flood-plain area.

Cities and towns along the Mississippi and Missouri Rivers derive their water supplies mostly from these sources. Surface waters are also used by many communities in the northern parts of the State that are covered with glacial deposits. Elsewhere in Missouri,

A. Physiography

1. Dissected Till Plains
2. Osage Plains
3. Springfield Plateau
4. Salem Plateau
5. St. Francois Mountains
6. Southeastern Lowlands

B. Geology

1. Quaternary-Tertiary rocks
2. Cretaceous rocks
3. Pennsylvanian rocks
4. Mississippian rocks
5. Devonian-Silurian rocks
6. Ordovician rocks
7. Cambrian rocks
8. Precambrian rocks

0   100   200   300   400 KILOMETRES

C. Soils (Order/Suborder)

1. Mollisols/Udolls
2. Mollisols/Aquolls
3. Alfisols/Aqualfs
4. Alfisols/Udalfs
5. Ultisols/Udults
6. Inceptisols/Aquepts

D. Vegetation-type

1. Glaciated Prairie
2. Unglaciated Prairie
3. Oak-Hickory Forest
4. Oak-Hickory-Pine Forest
5. Cedar Glades
6. Floodplain Forest

FIGURE 1.—The principal subdivisions of Missouri according to physiography (from Vineyard, 1967), geology (from Missouri Geological Survey and Water Resources, 1969), soils (from Soil Conservation Service, 1970), and vegetation (from Küchler, 1970).

water supplies are derived largely from wells. The surface and ground waters of the State may be classified into six geohydrologic units, each unit consisting of a group of geologic strata in which the waters occur: (1) Quaternary alluvium, (2) Pleistocene glacial drift, (3) strata of Tertiary and Cretaceous age, (4) strata of Pennsylvanian age, (5) strata of Mississippian age, and (6) strata of Ordovician and Cambrian age. The Cambrian and Ordovician strata are further subdivided into a shallow geohydrologic unit which extends over much of southern Missouri, and a deeper geohydrologic unit covered by Mississippian strata in the southwestern part of the State. Thus, seven broad

geohydrologic units are recognized.

The principal metalliferous mining in Missouri has been for lead and zinc, and for small amounts of associated copper, cobalt, nickel, silver, and cadmium. The State has had a long history of lead and zinc production, dating back to the early 1700's. Although small deposits are known to be widespread in the southern part of the State, most of the production has been from three districts: (1) the Southeast district in the vicinity of the St. Francois Mountains, (2) the Southwest district, which includes the Joplin area and the Missouri part of the famous Tri-State district, and (3) the less important Central district with a very limited amount of mining activity in recent years. In each of these districts the ores occur as replacement bodies in limestone and dolomite (Hayes and Guild, 1967).

The ores occur from the surface to depths exceeding 1,000 feet. In the Southeast district the ores occur mainly in dolomites of Cambrian age and consist largely of lead sulfide (galena). The Southeast district includes the Viburnum Trend, an area mainly in northwestern Reynolds and Iron Counties where significant new lead deposits were discovered in 1955 (Weigel, 1965). In the Southwest district the ores occur mainly in strata of Mississippian age and consist largely of zinc sulfide (sphalerite) with minor amounts of lead. The Central district contains many small deposits of barite, lead, and zinc, but most of the mining since 1910 has been for barite (Kiilsgaard, 1967, p. 62).

The only other important metalliferous mining in Missouri has been for iron which occurs in several kinds of sedimentary and vein-type deposits throughout much of the Ozark Plateaus region. At present, two large vein-type deposits in igneous rocks of Precambrian age are being worked at Pea Ridge and Pilot Knob in the St. Francois Mountains region.

METHODS OF SAMPLING

GENERAL DISCUSSION

A thorough discussion of sampling procedures should consist of two main aspects: (1) the sampling design, or plan, including the spacing of sampling sites and the selection of specific sampling points within sites, and (2) the techniques used to obtain the samples after these sites and points have been determined. Only the first aspect is described here; the second aspect is discussed in reports covering specific phases of the geochemical survey. The general features of the sampling designs can be discussed here because they are basically similar, and follow the same general statistical model, from one phase of the survey to another.

Before discussing the sampling designs employed in the geochemical survey of Missouri, two separate and independent properties of geochemical maps—resolution and stability—are described as they have been viewed during the course of the program. A general plan for efficient sampling of large regions to achieve various degrees of detail in the final geochemical maps, as described previously (Connor and others, 1972), is also discussed briefly. Each of the sampling designs used in the Missouri program has been regarded as part of this general plan.

A geochemical map is commonly constructed by plotting geochemical values and illustrating the variation among the values by means of contours, symbols, or other graphical devices. An individual value may represent (1) the analytical determination on an individual sample of rock, soil, plant, or water, or (2) the mean for a number of determinations on samples that were collected from a sampling locality. The first situation may be regarded as a special case of the second, one where the sampling locality is small and the single analytical determination is the mean of one value.

The resolution of a geochemical map is determined by the spacing between the centers of the sampling localities, a locality being defined as some part of the population being sampled that is larger than an individual specimen (Miesch, 1967a, p. A5). Where the localities are closely spaced, small-scale geochemical variations may be detected and described, and so the resolution of the map is high. Where the spacing between localities is greater, small-scale geochemical variations will go undetected and so the resolution is poor. The variability among the sampling locality means and the confidence intervals, the degree to which the means are known, determines the geochemical map stability. If the variability among the locality means is large in relation to the confidence intervals about the means, then the stability of the geochemical map is relatively high. Where the variability among the locality means is small relative to the confidence intervals about the means, the map is more likely to undergo substantial changes with the addition of new data, and so is less stable.

It is rather common in field geochemical studies to use sampling designs of the hierarchical type, and designs of this type were used in the geochemical survey of Missouri. Where hierarchical designs are used, the sampling localities occur at several levels. For example, 15-minute quadrangles may serve as the sampling localities at the uppermost level of the design and, thereby, constitute the "master sampling units" in the terminology of Krumbein and Slack (1956, p. 744). Sampling localities at the next lower level of the design may consist of, say, 7½-minute quadrangles selected from the 15-minute quadrangles, and at the next lower level the sampling localities may consist of 30-second quadrangles selected from the 7½-minute quadrangles. In such situations, the resolution of the geochemical map is determined by the spacing between the values plotted in construction of the geo-

chemical map and the proportion of the total map area that each value is intended to represent. Ordinarily, the plotted values will be averages of sampling units chosen in such a way that the plotted values will be as equally spaced as possible on the map. This will commonly preclude averaging over minor sampling localities at lower levels of the hierarchical design and, more often than not, will lead one to average over the master sampling units. In situations where approximately equal spacing can be achieved by averaging over sampling units at various levels of the design, a choice must be made. In general, if the smaller sampling units are averaged, the map will have higher resolution but poorer stability. If the larger sampling units are averaged, the resolution is lost in the interest of achieving stability. The latter choice is more often the better one because an unstable geochemical map is subject to possible misinterpretation by readers who are not acquainted with the nature of the difficulties faced in geochemical mapping. A generalized map that is reliable is more useful than a detailed map that is, in large part, not reproducible.

Connor, Feder, Erdman, and Tidball (1972) have described a general sampling plan for geochemical surveys of large regions, a plan that has grown out of the geochemical survey of Missouri. The plan is designed to offer maximum efficiency in fieldwork and laboratory analysis; it is composed of two phases of sampling with two sampling stages within each phase. The first phase leads to geochemical maps of poor resolution, but a map of this type may be all that is required or all that available resources allow in some investigations. The second phase is aimed at the production of maps of higher resolution and is conducted only where maps from the first phase are determined to be inadequate for a given purpose. The first stage of sampling within each phase of the general design is for the purpose of estimating the extent of sampling and laboratory analysis required in order to produce geochemical maps and other statistical summaries with the required degree of stability. The second stage is merely the execution of the sampling as determined to be required by the results of stage one.

The first phase of sampling in the general plan is aimed at estimation of the differences among the average geochemical characters of various categories of materials being investigated. In most cases, the categories are mappable units of bedrock, soils, vegetation, and water. Thus, the first phase is aimed at a description of what are probably the major geochemical variations over any large region. Because the master sampling units in the first phase of the general plan are mapped, or mappable, subdivisions of the materials being investigated, the resolution of the geochemical maps resulting from this phase can be no better than that of the geologic, soils, vegetation, or hydrologic map on which the sampling is based.

The second phase of sampling in the general plan is conducted on any one or on any number of the mapped rock, soil, vegetation, or hydrologic units, but each of these would require a separate and perhaps unique sampling design depending on the type of geochemical variability in the unit. If much of the variability is on local scales, a high resolution sampling design involving closely spaced sampling localities will be required if any large proportion of the total variability is to be described in the final geochemical map. If much of the variability is on regional scales, widely spaced sampling localities providing poor resolution in the final map may be adequate to describe a large proportion of the total variability.

The general plan has been summarized as follows (Connor and others, 1972):

Phase 1: Sampling to describe differences among categories.
  Stage 1a: Preliminary sampling designed to determine the extent to which the categories are indeed geochemically distinct, and to provide the basis for planning stage 1b.
  Stage 1b: Final sampling to provide reliable estimates of differences among categories, and the amounts of compositional variability within each category.
Phase 2: Sampling to describe patterns of variation within categories.
  Stage 2a: Preliminary sampling within each category to determine the sampling locality spacing that would be most efficient for describing the geochemical variation patterns within each category, and the number of samples required from each locality.
  Stage 2b: Final sampling to describe the geochemical patterns within each category.

SAMPLING DESIGNS

Most of the sampling designs used in the geochemical survey of Missouri were of the phase 1 type, and were for the purpose of describing differences in geochemical character among mapped rock, soil, vegetation, and hydrologic units. Most of the final geochemical maps, therefore, are of poor resolution and are intended only to show the major geochemical variations across the State. Until it has been determined by medical researchers, or by others, that maps of higher resolution are required, and where they are required, it seems unnecessary to proceed to phase 2 of the general plan. Phase 1 provides a means for showing the gross geochemical variations over a large area at relatively low cost, whereas phase 2 should be entered into only where the need for the more detailed information can justify the much greater cost per unit area that is involved.

Various statistical populations are recognized within the rocks, soils, vegetation, and waters of the State, and

these populations constitute the first breakdown of these materials. Examples of these populations are (1) carbonate rocks of Cambrian and Early Ordovician age, (2) sandstones of Pennsylvanian age, (3) glacial loess, (4) clay residuum developed on carbonate rocks, (5) currently cultivated (agricultural) soils, (6) soils of the prairie in the northern parts of the State, (7) sumac of the State, (8) pasture grasses of the State, (9) ground water of the State, and (10) surface water of the State. However, the sampling in the geochemical survey was not aimed directly at these populations. Rather, it was aimed at units of area because the purpose of the survey has been to describe variations among areas. Consequently, the individuals in the populations of interest have been units of area in all cases, and the rocks, soils, vegetation, and waters have been regarded as attributes of the areas. Accordingly, in the randomization procedures that were used, each unit area wherein the attribute of interest occurred was given an equal chance of being selected. This is quite different from sampling in such a way that each individual in the population is given an equal chance of being selected, and, as stated above, was done because the purpose of the investigation was to produce maps showing chemical variations in the rocks, soils, vegetation, and water across the State.

The populations and groups of populations that were the attributes of interest within the areas sampled are listed in table 1. The table also shows the master sampling units for each sampling design; for most designs the master sampling units are mapped units within the corresponding population. In addition, the table shows how each of the master sampling units was subdivided areally for sampling within it. The hierarchical-type subdivision was used in the stage 1a sampling in order to search for an efficient design for stage 1b. This is discussed further later in this section.

The general statistical model followed in the sampling within each population was:

$$x_{ijkm...} = \mu + \alpha_i + \beta_{ij} + \gamma_{ijk} + \delta_{ijkm} + \ldots \tag{1}$$

where x is a measure of the amount of a given constituent in an individual sample, μ is the mean for all areas of the State where the population occurs, α_i represents a mapped unit, and β, γ, and δ represent subdivisions of decreasing size within the mapped unit. In sampling of the bedrock units of the State, the term "α" was not included in the formal design because the mapped units are vastly different with respect to most chemical constituents and there seemed to be little point in testing for statistically significant differences. The term "α" was also omitted from the sampling models used for the loess deposits of the State and for agricultural soils, because useful mapped units were not recognized in either of these materials.

The definition of μ in equation 1 as the means for all areas of the State where the population occurs is important because it provides a more precise definition of the population of individuals that were sampled within each phase of the survey. For example, the individuals within the sampled population of carbonate rocks of Cambrian and Early Ordovician age do not include all of the potential carbonate samples that exist throughout this sequence of strata; they include only those within areas where the strata have been mapped, and exclude those in the subsurface. The individuals within the sampled population of sumac are not all sumac of the State, but rather all of the equal-size areas (for example, 7½-minute quadrangles) where sumac occurs. These rather subtle, but important, distinctions are in keeping with the objective of the survey which is to describe geochemical variation among areas.

The intent behind the phase 1 sampling in each of the sampling programs, except those programs directed at the loess and agricultural soils, was to produce estimates of the average compositions of the mapped units as designated on table 1, and the symbol "α_i" is used throughout this series of reports, except where noted, to designate a mapped unit of rock, soil, vegetation, or water. Thus, in the model used as a basis for vegetation sampling, α_1 represents the mean concentration of a chemical constituent in a plant species among all areas within the glaciated prairie vegetation-type area; α_2 represents the same for the unglaciated prairie, and so forth. The principal objective of the survey is to identify significant differences among the α_i. α_i is in units of deviation from the grand mean, μ; the absolute mean for any given mapped unit is $\mu + \alpha_i$.

The ability to identify significant differences among the mapped units, with a given number of samples collected and analyzed, depends on the variation among the α_i's and on the homogeneity within each mapped unit. This homogeneity determines the variation among the individual values of β_{ij} and among similar terms on the right side of equation 1. In the sampling of bedrock and surficial deposits, β_{ij} represents the mean concentration of a chemical constituent in a geologic province or some major area of occurrence of the bedrock unit within the population. In sampling vegetation and associated soils, β_{ij} represents the mean for a 7½-minute quadrangle. Like α_i, β_{ij} is in units of deviation; the absolute mean for any subarea within a mapped unit is expressed as $\mu + \alpha_i + \beta_{ij}$. In a similar way, subsequent terms on the right side of equation 1 represent subdivisions of the units represented by the preceding terms. Each of them is in terms of deviation from a mean. A final term, ε, has been added in most phases of the survey to represent errors in laboratory analysis. Where this term is absent, the laboratory errors contribute to variation in the last term actually used. The

TABLE 1.—*Definition of terms in the hierarchical statistical models used in sampling for various phases (populations) of the geochemical survey of Missouri*

Population or group of populations	Mapped units	Subdivisions of mapped units				Analytical error
	α	β	γ	δ	η	ε
Bedrock[1]	Geologic systems or sequences.[2]	Major areas of occurrence.[3]	Stratigraphic sections within major areas.	Portions of stratigraphic sections.	Samples within portions of sections.	Duplicate analyses of samples.
Surficial geologic deposits.	Glacial loess and carbonate residuum.[2]	....do..........	....do..........	Samples within stratigraphic sections.		Do.
Classified soils	Soil suborders.[3]	Soil subgroups	Soil series	Localities	Samples within localities.	Do.
Agricultural soils	None	Counties.[3]	Samples within counties.			Do.
Prairie soils	Parent material types.[3]	7½-minute quadrangles.	Composite samples of soil profiles within quadrangles.			Do.
Selected native vegetation and associated soils.	Vegetation-type areas.[3]	7½-minute quadrangles within areas.	Samples within quadrangles.			Do.
Selected farm crops and associated soils.	....do..........	....do..........	....do..........			Do.
Ground water	Geohydrologic units[3]	Townships within units.	Wells within townships.	Samples from wells.		
Surface water	....do..........	Samples within units.				

[1]Sampling designs varied somewhat for each lithologic type and for each geologic system or sequence.
[2]The variance of α was not estimated.
[3]Master sampling units in the sampling design.

general model has been discussed further in a previous report (Miesch, 1967a).

In all of the sampling based on the model in equation 1, the concentration of a chemical constituent in a sample is regarded as a function of the mean concentration for all areas of the State plus deviations determined by the particular part of the State from which the sample was taken and by random errors arising in laboratory treatment. Accordingly, the variance among the determined concentrations is a function of the variance among mapped units, the variance within units, and the variance arising from analytical procedures. This follows from equation 1 when we subtract μ from both sides and square. Then, summing across $i, j, k, m, \ldots$, the sums of the resulting cross-product terms on the right side tend toward zero if the selection of units represented by $\beta, \gamma,$ and δ has been unbiased. Finally, dividing each side by N, where N represents the number of samples, we obtain:

$$\sigma_x^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_\delta^2 + \ldots , \qquad (2)$$

where σ_x^2 is the variance among all areas of the State, σ_α^2 is the variance among mapped units, and subsequent terms represent variance among areas within mapped units. The final term may represent variance arising from analytical imprecision, or combined effects of analysis and of variation within the smallest areal unit used in the design.

Estimates of the variance of x, and of the variance components on the right side of equation 2, are denoted by:

$$s_x^2 = s_\alpha^2 + s_\beta^2 + s_\gamma^2 + s_\delta^2 + \ldots , \qquad (3)$$

and are derived by means of the analysis of variance procedures outlined by Anderson and Bancroft (1952, p. 325–330). These procedures are applicable to data from both balanced and unbalanced designs; unbalanced designs result when, for some reason, the number of areas sampled varies from one master sampling unit to another, the number of subareas sampled varies from one area to another, or where a similar circumstance occurs at any level of the design.

The estimates of the variance components in equation 3 provide a useful description of the type of areal variation in composition that the sampled population possesses. Where s_α^2 is large and the remaining components are small, the mapped units about which the sampling plan has been designed are compositionally homogeneous and distinct from one another. Where s_α^2 is small and the other components are large, the mapped units are not compositionally distinct. Where s_β^2 is large compared with subsequent terms in equation 3, the variation within the mapped units tends to be on a large scale; where the opposite is true, the variation tends to be dominantly local. If the final term represents variance arising from analytical imprecision, its magnitude compared with preceding terms can indicate the adequacy or inadequacy of the laboratory method for the particular problem at hand.

Extensive use has been made of a quantity termed the variance ratio, v, in examination of the efficiencies of various executed and potential sampling designs. The variance ratio is given as:

$$v = \frac{N_v}{D_v} = \frac{s_\alpha^2}{s_\beta^2 + s_\gamma^2 + s_\delta^2 + \ldots} \qquad (4)$$

The numerator of v, N_v, is the estimated variance among mapped units, and the denominator, D_v, is the estimated total variance within mapped units. Thus, the variance ratio is a relative measure of the compositional heterogeneity among, and the compositional homogeneity within, mapped units of the population being studied. It

is also a relative inverse measure of the degree of the sampling problem. Where the variance ratio is large, relatively little sampling is required to describe the compositional differences among mapped units. Where the ratio is small, more sampling and laboratory work will be required. The variance ratio is related to the conventional and familiar F-statistic as follows:

$$F = 1 + n_r v \qquad (5)$$

where n_r is the effective number of samples collected at random, using a simple random sampling design, from within each mapped unit. The number of such samples required in order to distinguish compositional differences between any two mapped units, at a given level of confidence, is the value of n_r in equation 5 that will produce a value of F that exceeds the critical value for 1 and $2n_r-2$ degrees of freedom. Curves giving the required n_r for various values of v, at the 80- and 95-percent confidence levels, are shown in figure 2A.

Once n_r has been determined from the curves in figure 2A, or from iterative calculation, the maximum acceptable error variance of the means for mapped units is given by:

$$E_r = \frac{s_\beta^2 + s_\gamma^2 + s_\delta^2 + \ldots}{n_r} \qquad (6)$$

Where the sampling is according to a hierarchical design such as given in equation 1, the variance of the mean for a mapped unit is (Baird and others, 1967, p. 16)

$$E_s = \frac{s_\beta^2}{n_\beta} + \frac{s_\gamma^2}{n_\beta n_\gamma} + \frac{s_\delta^2}{n_\beta n_\gamma n_\delta} + \ldots, \qquad (7)$$

where n_β, n_γ, and n_δ are, respectively, the number of randomly selected areas sampled within each mapped unit, the number of randomly selected subareas within each area, and so forth. Where, after an initial sampling program (stage 1a) has been conducted, the computed E_s from the expression in equation 7 exceeds the computed E_r from expression 6, additional sampling (stage 1b) is required in order to achieve sufficient power to describe the gross compositional differences among mapped units.

Adjustment of the subscripted n's in expression 7 to reduce the variance of the means for mapped units sufficiently affords an opportunity to consider not only the number of additional samples it may be necessary to collect, but also the most efficient spacing of the sampling points. In general, the fieldwork will be less costly for an increase in n_δ than for an increase in n_γ, and an increase in n_γ will be less costly than an increase in n_β. The reason, of course, is that intensive sampling in a relatively few small areas does not call for the time and travel required for the collection of an equal number of samples from widely separated points within a large region. However, in situations where s_β^2 is large compared with s_γ^2 and s_δ^2, increasing n_β may be the only way possible to reduce E_s sufficiently.
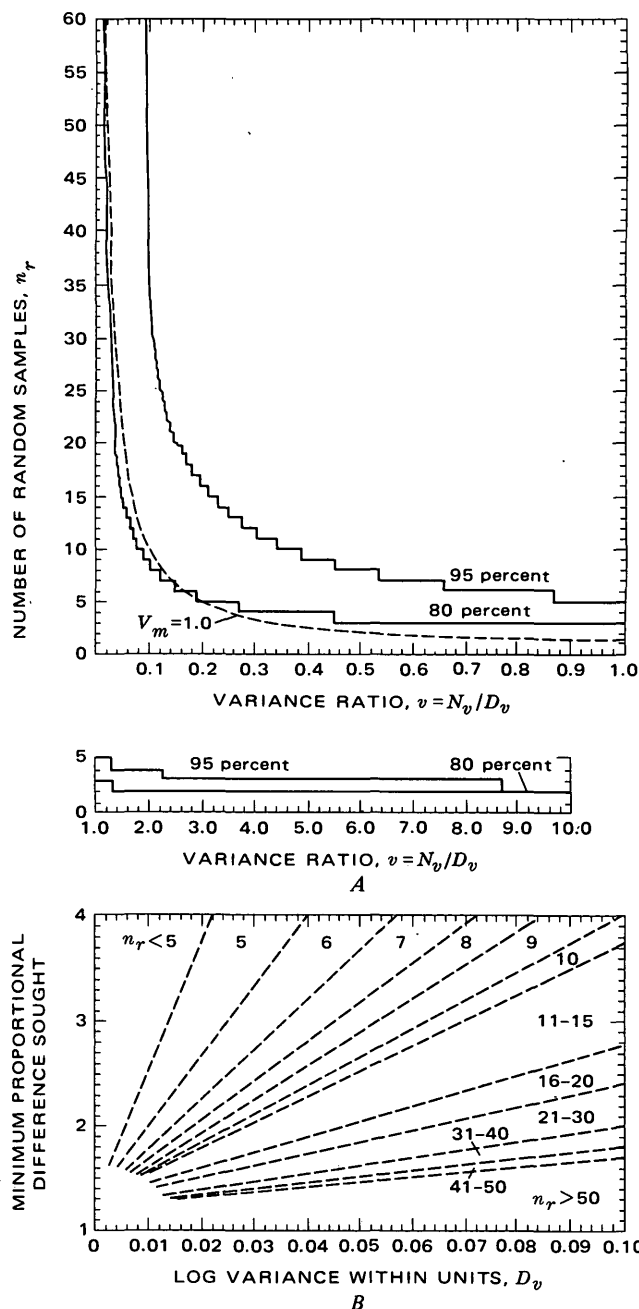
FIGURE 2.—Graphs showing the approximate number of random samples, n_r, required from each mapped unit. A, number required in order to describe the gross differences among a number of units. V_m, variance mean ratio (see text). B, number required in order to search for a minimum proportional difference between two units.

Computer simulation experiments, by R. R. Tidball (unpub.), employing a Monte Carlo technique, have shown that the degree to which the general configuration of a map pattern can be recovered through sampling is proportional to the ratio of the variance among sampling sites to the variance within them. At the very least, the variance among sites must equal the variance within sites in order for the true map pattern to be approximated by

means of sampling. Accordingly, it is useful to define a variance mean ratio, v_m, as:

$$v_m = \frac{N_v}{D_m}, \qquad (7a)$$

where N_v is as defined for equation 4 and D_m is the error variance of the means for mapped units and is equal to either E_r or E_s in equations 6 and 7, respectively, depending on the type of sampling used. In planning stage 1b sampling, then, it is possible to increase the n's in equations 6 and 7 until v_m is 1.0 or greater, as an alternative to using the F-test criterion already discussed. This procedure is approximately equivalent to using the F-test at the 80-percent confidence level, however, as may be seen by comparison of the curve for v_m=1.0 with the 80-percent curve in figure 2A. Whether or not it is to be used in planning stage 1b sampling, v_m is a valid and useful relative measure of the stability of a geochemical map.

A different F-test approach has been previously recommended by Garrett (1969; see also Michie, 1973, and Garrett, 1973), and the v_m ratio approach appears to have been used, in effect, by Baird and others (1967).

These procedures are appropriate where the objective of the sampling is to describe gross compositional variations among a number of mapped units or sampling localities. The collection and analysis of the numbers of samples indicated by the F-tests or the v_m ratio will lead to the production of a geochemical map that at least has a good chance of reflecting the actual geochemical variations in the materials being sampled. However in some instances the objectives may be somewhat more ambitious, and the sampling requirements will be correspondingly greater. For example, the objective may be to identify mapped units or sampling localities that differ with respect to the concentration of some element by a factor of 2 or more. Further, one may wish to make these identifications in such a way that there is small likelihood of erroneously identifying a difference that does not actually exist, or of not recognizing a difference of this magnitude that does exist. These two kinds of errors are identified in statistical terminology as Type I and Type II errors, respectively. Where the objective is of this kind, a method described by Davies (1954, p. 36) is more appropriate than the methods previously discussed. Where the variances are of logarithms, and where the minimum difference being sought is relative rather than absolute, the tables given by Davies can be summarized into diagrams such as given in figure 2B. Figure 2B is based on Davies' tables for the case where the probabilities of Type I and Type II errors are both 0.05. Thus, knowing the log variance within mapped units, $D_{\bar{y}}$, and the minimum proportional difference sought, one can use the diagram in figure 2B to determine the minimum number of independent samples required from each unit.

METHODS OF LABORATORY ANALYSIS

The experimental designs used in the geochemical survey of Missouri included evaluation of the precision of each method of laboratory analysis that was used, as well as a determination of the adequacy or inadequacy of that precision for the problem at hand. In general, the laboratory methods were of the rapid low-cost variety, and the precisions were poorer than might have been realized with more costly methods. In a few instances laboratory methods had to be abandoned because of inadequate precision, but in most cases the precision ranged from satisfactory to excellent. In fact, the results of the survey provide clear evidence that, in general, more costly laboratory procedures resulting in better precision would have been a waste of money. The reasoning that leads to this conclusion is that total error, not simply error resulting from laboratory analysis, is the important factor. Where the material being sampled is highly variable in composition on a local scale, the errors resulting from sampling tend to be large, and where sampling errors are large there is little point in expending resources to reduce error in the laboratory. The resources are better spent reducing the error in sampling. That is, where the sampling error is large, the total experimental error is more effectively reduced by rapid laboratory treatment of a large number of samples than by costly treatment of only a few. In fact, where sampling error composes the dominant part of the total experimental error, even a perfect analytical technique would not be helpful in achieving the required experimental precision.

Throughout the survey all samples from each phase were analyzed for a given constituent by the same method, in the same laboratory employing the same instruments, and, with some unavoidable exceptions, by the same analyst. Consistency in laboratory method also extended across various phases of the survey wherever this was possible. For example, all rock, soil, and plant ash samples were analyzed by the same spectrographic method in the same laboratory. Residues from water samples were analyzed by a nearly identical method, although in a different laboratory. Rock and soil samples were treated by identical laboratory methods for each constituent. Laboratory methods for treatment of plant samples and water samples necessarily differed. Methods of preparation of samples for laboratory analysis also differed for rocks, soils, plants, and water.

Within each phase of the survey (except in the studies of ground and surface water) all samples were collected before any of them were submitted to the laboratories. In general, 10–30 percent of the specimens were then split into two apparently homogeneous parts. All samples, including the duplicate splits, were placed in a randomized sequence—achieved by use of a table of permuted random numbers. The laboratories received and

analyzed the samples in this sequence, without identification of the duplicates. This procedure assured fair and realistic estimates of laboratory precision, estimates not possibly affected by the conscious or unconscious bias of the analyst. This fact, and the fact that laboratory precision as used here includes the effects of sample preparation, should be borne in mind when comparing estimates of laboratory precision in the Missouri survey with the analytical precision claimed in many other studies. The laboratory treatment of samples in a formally and completely randomized sequence is of utmost importance. The procedure insures that drifts in the analytical method, or different biases among individual analysts, will not generate false patterns on geochemical maps. Such drifts and biases are effectively converted to random errors, the effects of which can be overcome through statistical treatment of the data.

In the study of ground and surface water, it was not possible to hold samples for randomization until all were collected because chemical changes in the water might have occurred. Consequently, ground-water samples were collected in a randomized sequence and each sample was submitted to the laboratories immediately after collection. In the study of surface waters, such randomization in collection was not feasible, and the results of the laboratory work must be interpreted with the possibility of analytical drift in mind.

When a randomized sequence of rock samples was received in the laboratory, each sample was crushed in a jaw crusher and a ½-litre portion was split out for reserve storage. The remainder was ground in a vertical Braun pulverizer with ceramic plates set to pass 80 mesh. The ground sample was then mixed and split into seven parts for distribution to various laboratories.

The soil samples were thoroughly dried at 28°–30°C in an air-circulating oven, some taking 3 days to dry. The caked soil aggregates were then gently broken up in a ceramic mortar and passed over a 2-mm (millimetre) stainless steel screen. A Nasco-Asplin soil grinder, equipped with a ceramic mortar, ceramic screw type grinding head, and stainless steel screen, was used for this purpose. The material greater than 2 mm, mostly gravel, was discarded. The finer material was ground in a vertical Braun pulverizer using ceramic plates set to pass 80 mesh. As for the rock samples, the ground soil samples were then mixed and split into seven parts for distribution to various laboratories.

Plant samples were dried at 38°–40°C and ground in a Wiley mill to pass through a 1.3-mm screen. Mercury, selenium, arsenic, fluorine, and iodine were determined on this dried pulverized material. Other constituents were determined on the plant ashes obtained by dry ignition of the specimen in a furnace which is slowly heated from room temperature to 500°C.

Each ground-water sample consisted of about 6 litres collected in seven polyethylene containers ranging in volume from 120 ml (millilitres) to 2 litres. The seven containers were submitted to various laboratories for different chemical and spectrographic determinations. Similarly, each surface water sample consisted of about 6 litres collected in six polyethylene containers that were distributed to various laboratories. One set of surface water samples was filtered at the time of collection and another was not filtered. Certain containers of both ground water and surface water were acidified at the time of collection as was appropriate for the determinations to be made.

Details of the collection and preparation procedures are given in the following discussions of analytical methods and in the individual reports of this series on various parts of the geochemical survey.

ANALYSIS OF ROCKS AND SOILS BY X-RAY FLUORESCENCE
By J. S. WAHLBERG

The suites of samples of rocks and soils, collected for the geochemical survey of Missouri, were analyzed by X-ray fluorescence spectrometry methods for SiO_2, Al_2O_3, total Fe as Fe_2O_3, CaO, K_2O, P_2O_5, and Se.

The basis of analysis by X-ray spectrometry is that when an X-ray of sufficient energy is absorbed by an atom, that atom re-emits an X-ray with a wavelength characteristic of the element. Using a crystal of known lattice spacing to diffract these emitted X-rays, their wavelengths can be identified from their angle of diffraction. The analogy to optical emission spectrometry is nearly perfect, with a crystal in place of a diffraction grating. The intensity of the specific emitted X-ray is measured by a device such as a scintillation counter rather than a photographic plate. Useful texts on this method are those by Jenkins and deVries (1967) and by Bertin (1970).

As in most forms of spectrometry, the concentration of an element may be obtained by comparing the intensity of the emitted rays from the sample with those from known standards. However, as in other spectrometric methods, the amount that the unknowns may differ in composition from the standards is subject to many limitations. The various methods of fluorescence analysis are largely concerned with removing or suppressing interferences so that the concentrations of interest in the unknowns can be obtained by comparison with a limited number of standards. Three different methods were used for the samples of rocks and soils collected for the geochemical survey of Missouri.

FUSION METHOD

A fusion method was used for determining five "major" elements—silicon, aluminum, total iron, calcium, and potassium. These elements were generally present in

major amounts, and by custom and for convenience were expressed as percent oxides (SiO_2, Al_2O_3, Fe_2O_3, CaO, and K_2O). The fusion method is relatively fast, it overcomes most interferences by dilution, and it homogenizes the sample since the fusion is essentially a fluid solution.

In this procedure, 0.3000 g (gram) of ground sample is weighed onto 10.300 g of flux and mixed. This flux consists of 10.000 g $Na_2B_4O_7$ (anhydrous) and 0.300 g $NaNO_3$. The mixture, in a 30-ml platinum crucible, is placed in a muffle furnace and heated at 1050°C for approximately 20 minutes. A 2½-cm i.d. (inside diameter) brass ring is placed on a gold disk that is on a hotplate heated to about 400°C. The fusion is poured into the brass ring and allowed to cool slowly.

The samples are counted in a vacuum X-ray spectrometer. The percent concentrations are obtained by comparing the net counting rate of the sample with the counting rate from standards of known concentration similarly prepared. National Bureau of Standards samples were used for this study, such as NBS No. 1A (limestone), No. 98 (plastic clay) and No. 102 (silica brick). The following table summarizes the X-ray parameters used:

Element	X-ray parameter				
	Target	Detector	Crystal	Path	2θ (in degrees)
Si	Cr	Flow	PET[1]	Vacuum	109.06
Al	Cr	...do	PET	...do	144.67
K	Cr	...do	PET	...do	50.64
Ca	Cr	...do	PET	...do	45.15
Fe	W	...do	LiF(200)	Air	57.52

[1]Pentaerythritol

DIRECT POWDER METHOD

A direct powder method was used to determine phosphorus in the Missouri samples. It is a simple method and not of high accuracy—it may even be considered "semiquantitative."

The direct powder procedure for phosphorus involved half filling a 2½-cm o.d. (outside diameter) Caplug with a sample previously ground between ceramic plates set to pass 80 mesh. The Caplug was then covered with a polypropylene film held in place by two 2½-cm i.d. iron curtain rings. The sample was placed in a vacuum X-ray spectrometer and a 1-minute peak and offpeak count was made. The conditions used for the count were: Cr target X-ray tube, flow proportional detector, germanium analyzing crystal, vacuum path, and 2θ of 141.38° peak and 143.00° off-peak. The phosphorus was reported as percent P_2O_5.

PRECONCENTRATION METHOD

In the determination of trace elements, dilution is obviously undesirable, whereas a preconcentration method may both remove interferences and concentrate the element or elements sought to amounts more amenable to detection and estimation. In some cases a suitable matrix may also be obtained. Such a method was used to determine selenium in the Missouri samples.

The determination of the traces of selenium included a preconcentration. Two grams of sample was fused with 15 g Na_2CO_3 and 6 g Na_2O_2. The cake from the fusion was dissolved in water, about 20 g of NH_4Cl was added, and the mixture filtered. To the filtrate were added 0.1 mg of Na_2TeO_4, 5 g KI, 0.2 g $N_2H_4 \cdot H_2SO_4$(hydrazine sulfate) and 0.1 g Na_2SO_3 to precipitate the selenium. This precipitate was collected on a micropore membrane filter, dried, and placed on a 1-inch Caplug for counting. A peak and offpeak count was made and the amount of selenium was determined by comparing the net count with that obtained from a known amount of selenium similarly processed and counted. The X-ray parameters were: Mo target X-ray tube, scintillation detector LiF(200) crystal, airpath, and 2θ of 31.89° peak and 33.00° offpeak. The method has a lower limit of determination of 0.2 ppm selenium.

ANALYSIS OF ROCKS AND SOILS BY ATOMIC ABSORPTION SPECTROMETRY AND OTHER METHODS

By CLAUDE HUFFMAN, JR., and J. I. DINNIN

This section describes the analytical procedures other than X-ray fluorescence, emission spectroscopy, and neutron activation that were used for the analysis of rocks and soils throughout the duration of the project. The selection of analytical methods was based on such criteria as speed, general applicability to the kinds of materials to be sampled, and the required precision. Analytical procedures for the determinations of Na, Mg, Li, Zn, Cd, C (three types), F, As, and Hg are described briefly.

Na and Mg analyses of rocks and soils were made by V. M. Merritt and J. M. Gardner; Li and Zn analyses by J. A. Thomas; Cd by J. P. Cahill; and the three types of carbon determinations were by I. C. Frost and T. L. Yager; F analyses for the rocks and soils were made by R. Moore, J. W. Budinsky, B. A. McCall, H. Kirschenbaum, S. D. Kevan, and M. J. Joyce; As analyses by F. W. Brown, L. Mei, B. A. McCall, C. L. Burton, H. Kirschenbaum, and M. J. Joyce. Analyses for Hg were made by R. L. Turner.

Atomic absorption spectrophotometric methods were used to determine Na, Mg, Li, Zn, and Cd. The theory and concepts of atomic absorption spectroscopy have been explained in detail by Slavin (1968). The atoms of every element can absorb light radiation at narrow-wavelength emission lines which are different for each element. In order to be in a condition to absorb, the atoms must be chemically unbound and in their minimum energy state (ground state); this condition may be achieved by vaporizing the sample solution in a flame. A common source of radiation is the hollow-cathode lamp, whose cathode is made of the element to be determined, and which emits the wavelength lines of that element. The

light from the lamp is passed through the flame containing the sample vapor. The ground-state atoms of the element to be determined absorb an amount of light proportional to the concentration of that element present in the sample. The light passes through a grating monochromator which isolates the desired wavelength for the element to be determined from all other energy in the light beam and is subsequently measured by a photomultiplier tube in combination with a precise photometer. Na, Mg, Li, and Zn were determined by use of a single sample solution. The method consists of decomposing 1 g of rock or soil sample with nitric, hydrofluoric, and perchloric acids, fuming to dryness, and finally taking the salts into solution in 100 ml of 5-percent volume/volume hydrochloric acid. Portions of this sample solution were aspirated into the air-acetylene flame of an atomic absorption spectrophotometer to determine Na, Mg, Li, and Zn. Appropriate hollow cathode lamps were used for each element. The aliquot taken for the determination of Mg was diluted with a lanthanum chloride solution so that the final volume of solution contained 5 percent HCl and 1 percent lanthanum. Lanthanum acts as a releasing agent in the Mg determination. Standard solutions containing known concentrations of each of the elements to be determined were used for calibration.

Cd was determined by an atomic absorption spectrophotometric method slightly modified from that described by Nakagawa and Harms (1968). In this procedure a 1-g portion of sample is digested with nitric acid in a culture tube and allowed to cool. The volume of the solution is adjusted to 20 ml with demineralized water, mixed thoroughly, and centrifuged. A portion of this solution is then aspirated into the air-hydrogen flame of the atomic absorption spectrophotometer.

The Na, Mg, Li, Zn, and Cd values were obtained with a Perkin-Elmer model 303 atomic absorption spectrophotometer using the following instrument parameters recommended by the manufacturer:

Element	Grating	Wavelength (nm)	Slit	Source[1]	Lamp current (mA)	Burner	Flame	Flame condition	Filter
						Instrument parameter			
Na	Vis	587.0	3	Na	8	5- and 10-cm premix.	air-acetylene.	Oxidizing	Out
Mg	U	285.2	5	Mg	15	10-cm premix.	...do......	Reducing	Out
Li	Vis	670.8	5	Li	15	...do......	...do......	Oxidizing	In
Zn	U	213.8	4	Zn	15	...do......	...do......	...do......	Out
Cd	U	228.8	1	Cd	8	...do......	air-hydrogen.	...do......	Out

[1]Hollow cathode lamp.

Three types of carbon determinations were made: for carbonate carbon (that is, the carbon present in carbonate minerals), total carbon, and organic carbon. The gasometric method described by Rader and Grimaldi (1961) was used to determine carbonate carbon. In this method, carbon dioxide is liberated by the action of 1+1 hydrochloric acid on the sample. The sample size is varied from 0.5 to 2.0 g. The volume of the liberated carbon dioxide plus the air present in the reaction flask is measured at a known temperature and known atmospheric pressure. The combined gases are then scrubbed free of carbon dioxide by passage through a potassium hydroxide solution. The volume of the residual gases is then measured at the same temperature and pressure. The difference in the observed volumes, due to the carbon dioxide absorbed, is calculated to standard conditions of temperature and pressure, using the ideal gas relationships. The percentage of carbon dioxide is then calculated from its volume, and reported as C.

Total carbon was determined by the Leco induction-furnace gasometric method, a method which has been used in the steel indistry for many years. I. C. Frost (written commun., 1968) has shown how the method may be applied to the analysis of rocks, soils, and other geologic materials. A sample of known weight (0.25–0.40 g) is placed in a Leco ceramic crucible to which are added one or more combustion accelerators (tin metal, iron chips, and tin-coated copper). The crucible containing the sample is then placed in a high-frequency induction furnace within an enclosed combustion tube through which pure oxygen is passed. The sample is heated to over 1300°C where it oxidizes or burns readily. Evolved gases, including carbon dioxide, oxides of sulfur, and possibly carbon monoxide, are passed through a trap containing specially prepared manganese dioxide to remove the sulfur oxides, then through a heated catalyst tube where any CO formed is converted to CO_2. The carbon dioxide gas volume is measured with the Leco semiautomatic gasometric determinator. Carbon dioxide is measured and the percentage carbon is calculated in the same manner as used in the determination of carbonate carbon.

Organic carbon is a computed value based on the difference between the separate determinations of total carbon and carbonate carbon.

Fluorine was determined by the procedure described by Ingram (1970). In this procedure, fluorine is determined potentiometrically using a fluoride specific-ion electrode. A 100-mg sample is fused with a mixture of sodium carbonate and zinc oxide and the melt leached with water. The leachate is filtered, acidified with hydrochloric acid, and diluted to a specific volume. An aliquot is taken, sodium citrate and potassium nitrate added, and the potential of the solution is compared to the potential of solutions containing known concentrations of fluoride.

Arsenic was determined by an arsine evolution-spectrophotometric-isotope dilution method (F. W. Brown, L. P. Greenland, and F. O. Simon, written commun., 1973). One-half gram of sample plus added [76]As tracer is fused

with a mixture of sodium hydroxide and sodium peroxide, leached with water, and acidified with hydrochloric acid. Zinc shot is added and the evolved arsine is absorbed in an iodine-potassium iodide-sodium bicarbonate solution. the [76]As activity of the solution is counted with a single-channel analyzer to determine the percentage removed of the arsenic originally present. The arsenic in solution is then determined by a molybdenum blue spectrophotometric procedure and corrected for loss.

Mercury was determined by the mercury vapor detector described by Vaughn (1967). In this flameless atomic absorption technique, mercury is vaporized from 100 mg of powdered sample in an induction furnace. The mercury vapor is collected by amalgamation onto silver to eliminate interferences. The silver amalgamator is subsequently heated by an induction furnace and the released mercury vapor is swept through an absorption chamber where its concentration is measured.

The mercury vapor detector is calibrated with mercury-saturated air, which is obtained by placing a few grams of mercury in a 60-ml serum bottle having a sleeve-type rubber-membrane stopper. The air is withdrawn from the bottle with a hypodermic syringe by puncturing the membrane to insert the needle. The membrane seals itself as the needle is removed. The temperature-saturation curve for mercury vapor in air is shown by Vaughn (1967, fig. 6).

The lower limits of determination for the elements analyzed by the preceding methods are listed in table 2. The lower limits are based on the sample weights normally used for the determination.

TABLE 2.—*Lower limits of determination for some of the methods used for the analysis of rocks and soils*

Element	Method	Sample weight (g)	Lower limit (ppm)
Na₂O	Atomic absorption	1.0	100
MgO	....do	1.0	300
Li	....do	1.0	5
Zn	....do	1.0	10
Cd	....do	1.0	1
C (total)	Induction furnace-gasometric	.25-.40	500
C (in CO₃)	Acid liberation-gasometric	.5-2.0	100
C (organic)	By difference	.....	1,000
F	F specific-ion electrode	.1	40
As	Arsine evolution-spectrophotometric-isotope dilution.	.5	1
Hg	Flameless atomic absorption....	.1	.01

ANALYSIS OF ROCKS, SOILS, AND PLANT ASHES BY EMISSION SPECTROSCOPY

By HARRIET G. NEIMAN

The optical emission spectroscopic analyses of rock and soil samples were performed by methods similar to those described by Myers, Havens, and Dunton (1961). However, the results were reported in six geometric brackets per order of magnitude, rather than three, as explained below.

The rock and soil samples were received in the spectrographic laboratory after having been ground with alumina-ceramic plates set to pass 80 mesh and thoroughly mixed. A 10-mg sample was mixed with 20 mg of a sodium carbonate-graphite mixture (1 percent Na) and then packed into a shallow crater electrode and burned for 2 minutes in a direct-current arc. The sodium carbonate-graphite mixture serves as a spectrographic buffer and ensures even burning of the sample. The resulting spectra were recorded on a photographic plate which was developed under suitable conditions and visually compared to reference standards.

The reference standards were prepared as follows: mixtures of the elements to be determined, generally as oxides, were added to a powdered matrix of pegmatite-like composition consisting of 60 parts quartz, 40 parts perthite, and 1 part Fe_2O_3. The element mixture was diluted, stepwise, with matrix material so that the element concentration decreased geometrically by a factor equal to the reciprocal of the cube root of 10 (about 0.464), usually from 10 percent to 0.0001 percent (1 ppm). The decreasing series of concentrations of elements in these standards, therefore, is: 10., 4.6, 2.15, 1., 0.46, ..., 0.001, 0.00046, 0.000215, 0.0001 percent.

The mixtures of the elements and matrix material were mixed with graphite powder in the amounts of 10 mg of standard mixture to 20 mg of graphite powder and packed into a shallow-crater electrode. Using the same procedure as for samples, spectra of the various standards were recorded on photographic plates. These "standard plates" are made for each batch (about 90 dozen plates) with the same emulsion number.

The amounts of the various elements in each sample were determined by visually comparing the sample's spectrum to the spectra on the standard plates. The analyst determines whether the density (blackness) of a specific element line in the sample matches that of the same line of known concentration in the standard, or whether the density appears to be about midway between those of two adjacent spectra of the standards. If the density of the element line for the sample is determined to approximately match that of the line for one of the standards, the concentration of the element is reported as one of the values in the series 10, 5, 2, 1, 0.5, ..., 0.001, 0.0005, 0.0002, 0.0001 percent (values of the standard samples rounded to one significant figure). If the density of the element line for the sample is determined to be about midway between those for the standards, the concentration is reported as one of the values in the series 7, 3, 1.5, 0.7, ..., 0.0015, 0.0007, 0.0003, 0.00015 percent (values which are approximate geometric midpoints between the corresponding values of the standards). The procedure just described is analogous to one that has been used

previously in the same laboratory which consisted of preparing standards in the following series of concentrations: 12, 8.3, 5.6, 3.8, 2.6, 1.8, 1.2, 0.83, 0.56, ..., 0.0012, 0.00083, 0.00056, 0.00038, 0.00026, 0.00018, 0.00012, 0.000083 percent. This series was generated by using a factor equal to the reciprocal of the sixth, rather than cube, root of 10 (about 0.681). Element concentrations were determined to occur between two of the values in the series (no matching was done) and reported as one of the values in the series of geometric midpoints; 10, 7, 5, 3, 2, 1.5, 1, 0.7, ..., 0.0010, 0.0007, 0.0005, 0.0003, 0.0002, 0.00015, 0.0001 percent. This is the same series of values reported with the method presently used, but the smaller number of standards now used greatly reduces the time required for analysis, and extensive testing has shown that little precision is sacrificed.

The lowest concentration that can be measured (the limit of determination) varies for each element, and this limit can also vary for a given element depending on the other elements present in the material being analyzed. Table 3 gives the elements looked for in all rock, soil, and plant ash samples and the approximate lower limits of determination. Where the element is not discussed or referred to in subsequent reports of this series, the concentrations were found to be generally less than these approximate lower determination limits.

The reported concentration values are reproducible within a factor of 1.5 about 68 percent of the time, and within a factor of 2 about 95 percent of the time. However, these percentages vary somewhat from one element to another and for various types of materials being analyzed. More appropriate estimates of reproducibility are given in the individual reports of this series.

TABLE 3.—*Approximate lower limits for the elements determined by emission spectroscopy*

	Lower limits of determination, in percent				
Si	0.002	Mg	0.002	K	0.7
Al	[1].01	Ca	.002	P	.2
Fe	.001	Na	.05	Ti	.0002

	Lower limits of determination, in parts per million				
Ag	0.5	Hf	100	Sb	200
As	1,000	In	10	Sc	5
Au	20	Ir	50	Sm	100
B	20	La	50	Sn	10
Ba	2	Li	100	Sr	5
Be	1.5	Lu	30	Ta	500
Bi	10	Mn	1	Tb	300
Cd	50	Mo	3	Te	2,000
Ce	200	Nb	10	Th	200
Co	5	Nd	70	Tl	50
Cr	1	Ni	5	Tm	20
Cu	1	Os	50	U	500
Dy	50	Pb	10	V	7
Er	50	Pd	2	W	100
Eu	100	Pr	100	Y	10
Ga	5	Pt	50	Yb	1
Gd	50	Re	50	Zn	300
Ge	10	Rh	2	Zr	10
Ho	20	Ru	10		

[1]Concentrations of Al below 0.1 percent may result from grinding with high-alumina ceramics plates.

Plant samples are received in the spectrographic laboratory after having been ashed and thoroguhly mixed, in accordance with procedures used in the plant laboratory of the U.S. Geological Survey. (The ashing procedure is described in the section on analysis of plants and plant ashes by T. F. Harms.)

In general, the technical procedure for the spectrographic analysis of plant ash is similar to that used for the analysis of rock and soil samples but there are some important differences.

As previously described, the mixtures of elements used in the preparation of the standards are contained in a matrix that approximates that of silicate rocks. In order for the plant ash samples to be in a similar milieu, one part plant ash is diluted with 1.15 parts of a mixture of 90-percent pure quartz and 10-percent sodium carbonate. Then, 10 mg of this diluted plant ash is mixed with 20 mg of graphite powder and the procedure is continued exactly as described for the analysis of rocks and soils.

Reporting the plant-ash results differs from that of rocks and soils in that the dilution factor must be considered. Therefore, the observed reading from the photographic plate is increased by one-third order of magnitude (two "steps" in the series used for reporting results); for example, a reading of 0.7 percent is reported as 1.5 percent, and 5 percent as 10 percent. Because the diluting mixture contains sodium and silica, these elements cannot be determined by this procedure.

Owing to the dilution of the plant-ash samples, the limits of determination of the various elements are generally one-third order of magnitude higher than those for rock and soil samples.

ANALYSIS OF ROCKS AND SOILS FOR IODINE BY RADIOISOTOPE DILUTION-NEUTRON ACTIVATION

By A. J. BARTEL and H. T. MILLARD, JR.

The reliable determination of iodine in most geologic materials has been a problem to the analyst. This problem results from the low concentrations encountered (for example, 0.0X to 0.X ppm in igneous rocks and 0.X to X. ppm in sedimentary rocks and soils), the volatility of the elemental form, and interferences from other, more abundant halogens (especially bromine). Most chemical methods lack sufficient sensitivity to determine iodine in rocks and soils.

Neutron activation analysis (NAA) has been used by a number of researchers to determine iodine at the parts-per-billion concentration level in meteorites and lunar samples (Goles and Anders, 1962; Reed and Allen, 1966; Goles and others, 1967; Clark and others, 1967; and Reed and Jovanovic, 1970), and at the parts-per-million concentration levels in deep-sea sediments (Bennett and Manuel, 1968) and in soils (Lag and Steinnes, 1971). The 25-min

(minute) [128]I produced can be measured by either beta or gamma counting. Although beta counting is more sensitive, the lack of discrimination from the interferences of fission-produced iodine isotopes decreases this advantage to some extent.

The radioisotope dilution-neutron activation analysis (RD-NAA) technique (A. J. Bartel and H. T. Millard, Jr., unpub. data, 1975) used in the present study employs preconcentration of the iodine followed by determination by neutron activation and gamma counting. This approach utilizes the high sensitivity of neutron activation analysis but avoids the necessity of rapid radiochemical separation following activation. However, because chemical separations are performed prior to irradiation, the reagent blank must be evaluated and its magnitude determines the lower limit of detection.

PROCEDURE

The preconcentration steps involve fusion of 1 g of rock powder or soil with sodium peroxide, dissolution of the fusion cake with water, and acidification with sulfuric acid. The resulting iodate is then reduced to iodide with stannous sulfate, oxidized to iodine with sodium nitrite, and the iodine extracted into MIBK (methyl isobutyl ketone). The chemical yield for the preconcentration steps is determined by adding 50.0μ l (microlitre) of a carrier-free ^{125}I tracer solution ($\sim$0.02 mCi(millicurie)/ml) prior to fusion of the sample and then comparing the ^{125}I activity in the preconcentrate to a reference standard.

The iodine-MIBK solutions are poured into 2-dram polyethylene snap-cap vials, which are then heat-sealed and irradiated. Monitoring presents special difficulties. Monitors prepared by dispersing 50.0μ l of a standard iodide solution in MIBK and heat-sealing in polyethylene vials experience loss of iodine during irradiation to the extent of about 4 percent. In addition, the heat-seal occasionally fails and the top blows off. The first problem is treated by adding ^{125}I tracer prior to irradiation and performing a chemical yield determination after irradiation. The second problem is solved by using short lengths of aluminum wire loaded with cobalt as secondary monitors and counting the 10.5-min [60m]Co which is produced. These wires are attached to all samples and iodine monitors and are calibrated against iodine monitors (now the primary monitors) in a series of irradiations.

Two samples at a time are irradiated for 30 min in the rotating specimen rack of the U.S. Geological Survey TRIGA Reactor at a nominal thermal neutron flux of 2×10^{12} neutrons cm^{-2}S^{-1}. Immediately following irradiation, the Al-Co flux monitors are rinsed with water, placed on Al planchets, and the 59-keV (kiloelectronvolt) gamma peak of 10.5-min [60m]Co is counted for 6.7 min on a 30-cm^3 coaxial Ge(Li) detector. During these counts, the polyethylene vials containing the iodine-MIBK solutions are opened, the solutions transferred to new polyethylene

vials, and nitrogen gas bubbled through the solutions to remove [41]Ar, which results from irradiation of trapped air. The polyethylene vials are heat-sealed and the 443-keV gamma peak of 25-min [128]I is counted on the 30-cm^3 Ge(Li) detector for 16.7 min. All solutions are allowed to decay for 16 hrs after counting [128]I. Then the 35-keV gamma peak of 60-day ^{125}I in the samples, iodine monitors, and ^{125}I reference standards are counted on a high-resolution planar GE(Li) detector (LEPS).

CALCULATIONS

The blank is assumed to have two parts. One part of the blank is due to contamination from any iodine in the ^{125}I tracer solution, the sodium peroxide, the zirconium crucible, and the sulfuric acid. The fraction of iodine in this blank which reaches the preconcentrate is evaluated from the chemical yield. Another part of the blank is due to any iodine in the MIBK, the stannous sulfate, and the sodium nitrite. This part is assumed to reach the preconcentrate as a constant amount and is not corrected for chemical yield. One set of reagents contained 0.007±0.003 μg(microgram)I and 0.008±0.001 μg I respectively for these two portions and another set contained 0.022±0.009 μg I and 0.011±0.004 μg I respectively for the same two portions.

The net total areas, A, in counts per minute (CPM), under the 35-keV gamma-peak of ^{125}I, the 443-keV gamma-peak of [128]I, and the 59-keV gamma-peak of [60m]Co are found by integrating under the peak and substracting the background below the baseline. The concentrations of iodine in the sample, I_{sample} in parts per million, are then computed using the equation:

$$I_{sample} = \frac{1}{W_{sample}}\left[\frac{A^{128}_{sample}}{R_{wire}}\frac{W_{wire}}{A_{wire}} - I_{blank}\right]$$

where W_{sample} is the weight of sample in grams, W_{wire} is the weight of the Al-Co flux monitor wires in micrograms, A^{128}_{sample} is the activity of [128]I in the sample in CPM (corrected for chemical yield and radioactive decay), A_{wire} is the activity of [60m]Co in the wire in CPM (corrected for decay), and I_{blank} represents the amount of iodine in the blank in micrograms. R_{wire} is the calibration ratio for the Al-Co wires,

$$\frac{CPM^{128}I}{\mu g\ I}\Bigg/\frac{CPM\ ^{60m}Co}{\mu g\ wire}\ .$$

RESULTS

The accuracy of the RD-NAA technique was evaluated by means of multiple determinations of iodine, over a 6-month period, in a soil sample in which iodine had been determined by conventional neutron activation analysis (NAA). Two values of iodine were determined in each of

three NAA runs. The means and standard errors of the means are 3.11±0.081 ppm I (chemical yield found using ^{125}I tracer) and 3.31±0.046 ppm I (chemical yield found by re-irradiation). The mean and standard error of the mean of 23 RD-NAA runs are 3.03±0.044 ppm I. Thus the maximum bias indicated for the RD-NAA technique relative to the NAA technique is -8.5 percent.

The precision of the RD-NAA technique was estimated from the data for the 23 runs on the soil sample. The coefficient of variation is 6.9 percent. The three-sigma detection limit for the technique, as computed from the standard deviation of the blank, is 0.02 μg iodine.

ANALYSIS OF PLANTS AND PLANT ASHES BY METHODS OTHER THAN EMISSION SPECTROSCOPY

By T. F. HARMS

Prior to any analyses, the plant samples are dried at a low temperature (38°–40°C) and ground in a Wiley mill to form a homogeneous sample. After grinding, the particle size of the vegetation is such that it will pass through a screen with apertures of 1.3 mm.

Analyses for mercury, selenium, arsenic, fluorine, and iodine are made on the dried vegetation. Cadmium, cobalt, zinc, sodium, lithium, calcium, potassium, and phosphorus are determined on the ash obtained by dry ignition of the vegetation in a furnace which is slowly heated from room temperature to 500°C.

Selenium is determined fluorometrically; fluorine is measured by a selective ion electrode; and iodine, arsenic, and phosphorus are determined by colorimetric procedures. All the other elements determined in the dried vegetation and plant ash are measured by atomic absorption spectrometry.

A brief description of the chemical methods which were used follows.

Mercury.—Dried ground vegetation is digested with nitric, sulfuric, and perchloric acids under reflux to prevent loss of mercury during the early stages of digestion. A small amount of sodium molybdate solution is added to serve as a catalyst for the digestion (Munns and Holland, 1971). After sample dissolution, the digestate is heated to fumes of sulfur trioxide, cooled, and diluted with metal-free water. Stannous chloride solution is added to reduce the mercury to its elemental state.

The mercury is then swept from solution by a stream of air into an absorption cell attached to an atomic absorption spectrophotometer. The absorbance is measured and compared to the absorbances of known concentrations of mercury that have been treated similarly.

Selenium.—Air-dried ground vegetation is partially digested by heating with nitric acid. The solution is cooled slightly and small amounts of hydrogen peroxide are added to help destroy the remaining organic material. The digestion is completed by adding perchloric acid to the sample solution and boiling off the nitric acid. The perchloric acid digestate is warmed with dilute hydrochloric acid to reduce the selenate in the solution of selenite ion. The pH is adjusted to 2.0 with ammonium hydroxide and the selenium is reacted with 2,3-diaminonaphthalene to form a fluorescent complex which is extracted into hexane. Known concentrations of selenium are reacted with 2,3-diaminonaphthalene and the fluorescence is plotted against concentration to form a calibration curve. The fluorescence of the sample is measured and the selenium content is established by reference to the calibration curve.

Arsenic.—Air-dried ground vegetation is mixed with a slurry of magnesium oxide and magnesium nitrate, dried in an oven, and ashed in a porcelain dish at 500°C for 3 hours. The ash is acidified with hydrochloric acid, filtered to remove insoluble material, and treated with stannous chloride to reduce the arsenic to the trivalent arsenious ion. Zinc shot is added, and the evolved hydrogen and arsine are passed into an absorbing tube containing silver diethyldithiocarbamate dissolved in pyridine. The arsine reacts with silver diethyldithiocarbamate to form a soluble red complex which has a maximum absorbance at 540 nm (nanometres). The absorbance is measured spectrophotometrically and is compared with the absorbances measured from known concentrations of arsenic to determine the arsenic content of the sample.

Fluorine.—Air-dried ground vegetation is mixed with 50-percent sodium hydroxide solution, dried in an oven, and completely ashed at 500°C in an open nickel crucible. The ash is leached several times with hot distilled water and the leachate is placed in a plastic petri dish which serves as a diffusion cell. Perchloric acid is added to acidify the solution and the sample is diffused overnight using a small container of dilute sodium hydroxide to trap the hydrogen fluoride. A small amount of dilute perchloric acid is saturated with hexamethyldisiloxane and is added to the solution in the petri dish to aid in the diffusion of the fluoride ion (Taves, 1968). The sodium hydroxide solution is mixed with TISAB (total ionic strength adjustment buffer) and the fluoride ion is measured using a fluoride selective ion electrode. Known concentrations of fluoride which have been diffused and diluted with TISAB before measurement are used for preparing calibration curves.

Iodine.—The iodine content was determined using the method of Cuthbert and Ward (1964). One hundred milligrams of finely divided plant material is burned in an oxygen combustion flask, and the combustion products are collected in dilute sodium hydroxide. The alkali is acidified with sulfuric acid and the iodide ion is determined indirectly by its catalytic effect on the reduction of

ceric ion by arsenious ion in a dilute acid solution. For a given time and at a given temperature, the extent of the reduction of ceric ion is directly proportional to the amount of iodide present. The time allowed for the reaction is controlled by the addition of silver ion, and the extent of the reduction of ceric ion is measured spectrophotometrically. A standard curve is prepared by placing iodide solution on ashless filter paper, burning the paper in an oxygen flask, and carrying the combustion products through the procedure. Concentration is plotted against the extent of reduction of ceric ion as determined spectrophotometrically to establish a calibration curve.

Percent ash.—Four grams of dried, ground vegetation is weighed and placed in a crucible. The crucibles are placed in a cold muffle furnace and the temperature of the furnace is slowly raised over a period of approximately 10 hours to 480°–500°C. The furnace is maintained at this temperature for 12 hours and then is allowed to cool. The ash is weighed to three significant figures and the percent ash is calculated.

After the percent ash for each sample has been determined, sufficient quantities of vegetation are ashed to obtain the amount of ash required for those elements which are determined in the ash and for spectrographic analysis. A small plastic ball is added to each container of ash and the contents are shaken thoroughly to pulverize and mix the ash.

Cadmium and cobalt.—Air-dried ground vegetation is ashed at 500°C. The ash is heated with nitric and dilute hydrochloric acids. The samples are filtered to remove the insoluble materials, and the pH of the filtrate is adjusted to 2.2 with ammonium hydroxide. The solubilized cadmium and cobalt are reacted with diethylammonium diethyldithiocarbamate to form complex species which are then extracted into methyl isobutyl ketone. The ketone is atomized into an air-acetylene flame of an atomic absorption spectrophotometer, and the absorption due to cadmium or cobalt is measured. Known concentrations of cadmium and cobalt are extracted into ketone and atomized as for the samples, and a calibration curve is established by plotting concentration of the element against absorbance.

Zinc, lithium, sodium, calcium, and potassium.—These elements are determined by atomic absorption spectrometry. Air-dried ground vegetation is ashed at 500°C. For the determination of zinc, lithium, and sodium the ash is leached with hot 4 N nitric acid. The samples are centrifuged and the supernatant solution is atomized directly into an acetylene-air flame. For calcium and potassium, the ash is leached with 6 N HCl on a steam bath. The sample solutions are diluted and 10-percent lanthanum solution is added to the aliquot for calcium so that the final solution contains 1-percent lanthanum ion to mask interferences caused by aluminum, phosphate,

and sulfate ions. The solution is atomized directly into an air-acetylene flame. Standard solutions containing known concentrations of the element to be determined are used for calibration.

Phosphorus.—Air-dried ground vegetation is ashed at 500°C, and the ash is heated with 4 N HNO$_3$. The samples are diluted and centrifuged, and an aliquot of the supernatant liquid is reacted with ammonium molybdate-ammonium metavanadate reagent to form a yellow phosphorus complex. The yellow color is compared visually with the color of standard solutions prepared in a similar manner to determine the phosphorus concentration.

The lower limits of determination for the elements analyzed by the preceding methods are listed in table 4. The lower limits are based on the sample weight normally used for the determination and may vary because of limitations on the sample size.

TABLE 4.—*Summary of methods used for analysis of plants and plant ashes, and lower limits of determination*

Element	Method	Sample weight (g)	Lower limit (ppm)
Dry weight of sample			
Hg	Flameless AA	2	0.025
Se	2, 3-diaminonaphthalene	2	.01
As	Silver diethyldithiocarbamate	2	.25
F	Selective-ion electrode	2	.5
I	Catalytic..	.1	1
Ash of sample			
Cd	Atomic absorption............................	0.5	0.2
Co	....do..	.5	1
Zn	....do..	.05	20
Na	....do..	.05	25
Li	....do..	.05	4
Ca	....do..	.05	100
K	....do..	.05	100
P	Colorimetric	.05	100

ANALYSIS OF WATER BY CHEMICAL METHODS

By M. W. Skougstad and G. L. Feder

LABORATORY DETERMINATIONS

Instrumentation in the water-analysis laboratory permits rapid, accurate determination of a wide variety of constituents occurring as dissolved material in natural waters. Many of these constituents commonly occur at very low concentrations, but their determination is nevertheless important from the standpoint of uses of the water for various purposes as well as investigations of water as a geologic agent in weathering and decomposition of rocks, minerals, and soils. Thus, the present-day laboratory is ordinarily well equipped to determine not only the common, major constituents in water, but also a number of minor metallic and other constituents. For the most part, methods for the determination of the various substances are fairly well standardized and widely used by

analysts engaged in this sort of work. The methods used by the U.S. Geological Survey to analyze water samples collected as a part of the geochemical survey of Missouri, reported here, have been described by Brown, Skougstad, and Fishman (1970). The reader is referred to this publication for procedural details. However, a brief description of the method used for the determination of each constituent is given here.

All samples, with the exception of those collected for the determination of nitrogen and phosphorus species, were analyzed in the U.S. Geological Survey laboratories in Denver, by M. J. Fishman and O. J. Feist, Jr., or in Salt Lake City, under the direction of R. L. McAvoy and A. H. Handy. Determinations of nitrogen and phosphorus species were made in the U.S. Geological Survey laboratory in Little Rock, Ark., under the general supervision of R. T. Sniegocki and C. T. Bryant, by J. W. Gibbs, R. L. James, B. F. Lambert, R. D. McKibban, Marie K. Moore, and E. E. Morris.

Alkalinity.—The alkalinity determination involves an electrometric titration with standard sulfuric acid solution, and reflects the overall concentration of substances reacting with hydrogen ions under the conditions of the titration. The titration will show two endpoints if the sample pH exceeds 8.3, indicating the presence of CO_3^{-2} or CO_3^{-2} and OH^{-1} ions. In such cases the titration is carried to PH=8.3, and then continued to pH 4.5, enabling the calculation of OH^{-1}, CO_3^{-2}, and HCO_3^{-1} concentrations, when all are present. Total alkalinity is reported in terms of equivalent concentration of $CaCO_3$.

While water alkalinity is commonly due to bicarbonate or carbonate ions, or both, the presence of other titratable substances should not be precluded. Several acid radicals such as borate, phosphate, and silica, if present, contribute to the measured alkalinity.

Ammonia-nitrogen.—To determine ammonia (or ammonium ion) the sample is buffered to a pH of 9.5 to prevent hydrolysis of organic nitrogen compounds and yet ensure the subsequent complete recovery of ammonia upon distillation. An aliquot of the collected distillate is nesslerized or reacted with a reagent consisting of mercuric iodide in a potassium iodide solution. The reaction yields an intensely red compound, mercuric amidoiodide ($Hg(NH_2)I$), whose color intensity is directly proportional to the ammonia concentration of the sample.

Arsenic.—Arsenic is determined spectrophotometrically by reduction to arsine (AsH_3) gas by zinc metal in acid solution, distillation of the arsine through a lead acetate scrubber to remove sulfides, and collection of the arsine in a solution of silver diethyldithiocarbamate, which reacts to form a deep-red compound. The intensity of the resultant red color is proportional to the arsenic content of the sample. The method is exceptionally free of interferences if carefully performed; only antimony inter-

feres seriously, and its occurrence in water is generally quite rare.

Bromide ion.—Trace concentrations of bromide ion, as most commonly occur in fresh waters, are determined by a spectrophotometric method based on the catalytic effect of bromide ion on the oxidation of iodine to iodate by permanganate in acid solution. Two reactions occur consecutively: (1) iodide oxidizes rapidly to iodine, and (2) the resulting iodine oxidizes slowly to iodate. It is the latter reaction that is catalyzed by traces of bromide ion, and at any given temperature the oxidation rate is directly proportional to the amount of bromide present. In the determination, the reaction is stopped after a suitable time by extracting the unreacted iodine with carbon tetrachloride and measuring the absorbance of the extract. The ratio of the absorbance of an unknown sample to that of a reference sample containing no bromide is inversely proportional to the bromide ion concentration.

Calcium and magnesium.—Both calcium and magnesium are determined directly by AAS (atomic absorption spectrophotometry). Alkaline samples (pH>7.0) must first be acidified, and all samples must be treated with a lanthanum chloride solution in order to minimize interferences due to phosphate, sulfate, and aluminum. An air-acetylene flame is used in the spectrophotometer.

Cadmium.—Cadmium is determined by AAS by either of two methods: (1) directly without pretreatment of the sample if its cadmium concentration exceeds about 20 $\mu g/l$, or (2) after chelation of the cadmium with ammonium pyrrolidine dithiocarbamate (APDC), extraction of the cadmium complex into methyl isobutyl ketone (MIBK), and aspiration of the extract into the air-acetylene flame of the atomic absorption spectrophotometer. The second method is very sensitive and is preferred when concentrations of less than 10 $\mu g/l$ are to be determined.

Chloride.—Chloride is a common constituent and its determination is either by the well-known Mohr titration method, or, preferably, especially for samples containing less than 10 mg/l, by a mercurimetric titration method, using a solution of mercuric nitrate as the titrant and diphenyl carbazone as the indicator. Bromide and iodide titrate along with the chloride, but their concentrations are usually insignificant by comparison.

Iodide ion.—The very low concentrations of iodide ion commonly found in fresh waters are best determined spectrophotometrically by a method based on the catalytic effect of iodide on the oxidation of arsenic (III) to arsenic (IV) by cerium (IV) in acid solution. At a given temperature, the oxidation rate is directly proportional to the amount of iodide present. The reaction may be stopped at any suitable time by adding silver nitrate solution. The absorbance of the solution at 450 nm, when compared with the absorbance of a sample containing no iodide,

provides a ratio whose value varies inversely with the iodide concentration of the sample.

Iron.—Two forms of iron may be present in water samples: Fe(II) and Fe(III). Both may be determined and reported separately, or the total dissolved iron in both forms may be determined without distinguishing the relative amounts of each. Fe(II) is determined spectrophotometrically on the basis of its reaction with 2, 2'-bipyridine to yield a red complex. Total dissolved iron may be determined by the same reaction after first reducing all Fe(III) to Fe(II) with hydroxylamine hydrochloride; Fe(III) is then determined by difference. Alternatively, the total dissolved iron concentration is determined by AAS, the sample being aspirated directly into the air-acetylene flame of the spectrophotometer without pretreatment other than dilution or filtration as may be required. Samples containing iron in concentrations of at least 0.10 mg/l may be analyzed by this latter method; those containing less than 0.10 mg/l iron must be analyzed by the spectrophotometric-bipyridine method.

Magnesium.—See "Calcium and Magnesium."

Manganese.—Manganese is determined by AAS, either by direct aspiration of the sample without pretreatment (for concentrations of at least 10 $\mu g/l$), or after chelation with APDC and extraction of the complex into MIBK (for concentrations of 1–10 $\mu g/l$).

Mercury.—A flameless AAS (cold-vapor AAS) method permits determination of as little as 0.5 $\mu g/l$ mercury with a precision approaching 0.1 $\mu g/l$. The method is based on the preliminary digestion and reduction of all mercury compounds, including organic-mercury compounds, to metallic mercury, flushing the mercury vapor from solution by aeration, and then measuring the absorbance of the vapor at 253.7 nm in the conventional atomic absorption spectrophotometer. The percent absorption is plotted on a recorder and mercury is determined by reference to an analytical curve prepared from standards. Some samples may contain volatile organic compounds which absorb radiation at 253.7 nm and which may be swept from the solution along with the mercury vapor. These constitute a positive interference and the possibility of their presence must not be overlooked.

Nitrite- and nitrate-nitrogen.—Nitrite-nitrogen is determined spectrophotometrically by direct diazotization with sulfanilamide and subsequent coupling of the resulting diazo compound with 1-naphthylethylenediamine to form an intensely colored red dye. Nitrate-nitrogen is determined by measuring spectrophotometrically the intensity of the yellow color that is produced by the reaction between the alkaloid, brucine, and nitrate ion in an acid medium.

Organic-nitrogen.—A high-temperature digestion (Kjeldahl) of a water sample decomposes organic-nitrogen-containing compounds including amino acids, polypeptides, and proteins. The digestion is carried out in concentrated sulfuric acid solution in the presence of copper sulfate which catalyzes the decomposition and leads to the ultimate formation of ammonium ion. The digested mixture is subsequently made alkaline with sodium hydroxide, and the free amonia is distilled off and is nesslerized. When high concentrations of organic-nitrogen are present, the ammonia may be distilled into boric acid (H_3BO_3) and titrated with standard sulfuric acid solution.

Phosphorus, dissolved orthophosphate.—Orthophosphate is the most common ionized form of phosphorus in water. The method of analysis is highly specific for orthophosphate, but does not distinguish between the three ionization products of phosphoric acid, $H_2PO_4^{-1}$, HPO_4^{-2}, and PO_4^{-3}. The relative concentrations of these ionic forms in any given sample are a function of pH. Orthophosphoric acid reacts with molybdate ion in acid solution to form the yellow phosphomolybdate ion which can be reduced to form an intensely colored (blue) complex. Ascorbic acid, in the presence of antimony ion as a catalyst, is used as the reducing agent.

Phosphorus, total dissolved.—If a sample is digested with a mixture of sulfuric acid and potassium persulfate, complex polyphosphates are hydrolyzed to orthophosphate, and complex organic phosphorus-containing compounds are decomposed and ultimately form the orthophosphate ion. The resulting orthophosphate may then be determined by the spectrophotometric method outlined in the preceding paragraph. This method permits determination of total dissolved phosphorus: Organic phosphorus content can be computed by difference between total phosphorus and orthophosphate, although it must be understood that such a value for organic phosphorus includes any hydrolyzable polyphosphates that may be present.

Potassium.—See "Sodium and potassium."

Selenium.—The spectrophotometric-diaminobenzidine method is used to determine selenium. Selenium is quantitatively separated from most other elements by distillation of the volatile tetrabromide from an acid solution containing bromine, the bromine being generated in situ by a hydrogen peroxide–bromide reaction. Selenium tetrabromide is volatilized and absorbed in water, the excess bromine precipitated as tribromophenol, and the selenium then determined by reacting the resulting selenous acid with 3, 3'-diaminobenzidine to form a piazselenol of intense yellow color. The yellow color of the solution is compared with that of standards prepared in a similar manner.

Silica.—Silica may occur both as dissolved silicate or silicic acid, and as colloidal or so-called unreactive silica, SiO_2. Dissolved silicate or silicic acid reacts with molybdate ion in a strongly acid medium to form a yellow

silicomolybdate compound. This substance reacts further with sulfite ion to form a reduced compound of intense blue color whose intensity is proportional to the silica content of the sample. The silicomolybdate complex exists in two different polymorphic forms, α and β, of which the latter has the deeper color. Formation of the β-polymorph is favored if the solution is kept strongly acid. Silica may also be determined directly by AAS provided a high-temperature burner, such as the ntirous oxide-acetylene burner, is used.

Sodium and potassium.—The alkali metals sodium and potassium, found in all natural waters, are determined directly by AAS.

Solids, dissolved.—Evaporation of a known volume of a filtered sample followed by heating for an hour at 180°C provides a dry residue whose weight represents the dissolved-solids content of the sample. A platinum dish is commonly used for the evaporation. A final drying temperature of 180°C is a compromise between a temperature high enough to ensure complete removal of all water of hydration from the hydrated salts and one low enough to avoid either volatilization or thermal decomposition of salts and organic and labile materials.

Sulfate ion.—Sulfate-ion concentrations are determined titrimetrically with standard barium chloride solution using thorin as the indicator. Thorin and barium react to form a deep-red complex. The color intensity is greater in organic solvents, so the titration is carried out in a mixture of about 66 percent dioxane by volume and the mixture adjusted to a pH of between 2.2 and 5.0. Several multivalent cations besides barium form intensely colored complexes with thorin. To eliminate interference from such cations, the sample aliquot is treated with a cation-exchange resin prior to titration.

Zinc.—Zinc, like most other minor, transition elements, is best determined by AAS. Because of the good sensitivity of the AAS method for this element, it may be determined by direct aspiration of an untreated sample into the air-acetylene flame of the spectrophotometer. Concentrations as low as 10 μg/l may readily be determined.

FIELD DETERMINATIONS

Soon after a water sample is collected, its physical and chemical properties may change. Among the changes are: (a) temperature, (b) loss of gases, (c) reactions with suspended sediment in the sample, (d) hydrolysis, (e) oxidation, and (f) precipitation of compounds, especially calcium carbonate. In order to minimize the possible inaccuracies to chemical and physical determinations caused by these factors, certain determinations are made in the field. These determinations were: alkalinity, specific conductance, dissolved oxygen, pH, and temperature. Other changes in chemical properties of samples may be greatly minimized by pretreatment of samples before shipping. The methods used were described by Brown, Skougstad, and Fishman (1970) and Barnett and Mallory (1971).

Alkalinity.—A field determination of alkalinity is made of each sample at the time it is collected. The determination is made by electrometric titration with standard sulfuric acid solution. A Beckman Electromate solid-state portable pH meter is used to measure the pH during titration. If the sample pH exceeds 8.3, the titration is carried out to two end points. The first titration to pH 8.3 indicates the presence of CO_3^{-2} or CO_3^{-2} and OH^{-1} ions. The second titration which is continued to pH 4.5 indicates the presence of HCO_3^{-1}.

The main interferences in this determination are caused by salts of weak organic and inorganic acids, such as silicates, borates, and phosphates, which if present, contribute to the measured alkalinity. A more detailed discussion of field procedures for determining alkalinity is given in Barnes (1964).

Conductance, specific.—Specific conductance is measured in the field using a Beckman model RB-3 portable specific conductance meter, with manual temperature compensator. After measuring the temperature of the water sample and adjusting the manual temperature compensator, the conductivity is read by inserting a dip-type conductivity cell into a 1,000-ml beaker filled with the sample.

Oxygen, dissolved.—Dissolved oxygen is measured at each surface water sampling site during low flow. The sample is collected in a 300-ml BOD (biochemical oxygen demand) bottle, with a minimum of aeration. The dissolved oxygen content of the sample is determined using a modification of the Winkler process (Brown and others, 1970, p. 126–129). The method depends on the formation of a precipitate of manganous hydroxide. The oxygen dissolved in the water is rapidly absorbed by manganous hydroxide, forming a higher oxide. Upon acidification in the presence of iodide, iodine is released in a quantity equivalent to the dissolved oxygen present. The liberated iodine is then titrated with standard sodium thiosulfate solution using starch indicator. The method is susceptible to interferences from heavily polluted waters, free chlorine, more than 200 mg/l ferric iron, and readily oxidizable or reducible organic substances.

pH.—The pH is measured at the sampling site using a Beckman Electromate solid-state portable pH meter. The instrument is first standardized with pH 4.00 and 7.00 standard buffer solutions. The buffer solutions are kept at the same temperature as the sample. Most potable waters have few interferences that may affect the accuracy of the pH determination. When not being used, the glass electrode is kept immersed in distilled water.

Temperature.—Temperature is measured at the time of sampling using a mercury thermometer. The temperature is read to the nearest 0.2° Celsius.

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac															

	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lw

FIGURE 3.—Minor elements (outlined) commonly determined in water by emission spectroscopy.

ANALYSIS OF WATER RESIDUE BY EMMISSION SPECTROSCOPY

By P. R. BARNETT

The emission spectrograph provides a means of reasonably rapid analysis of water samples for many elements simultaneously. The elements determined or specifically sought in the water samples of the Missouri survey are enclosed with solid lines in the periodic table of figure 3. Some of these elements are ubiquitous in natural waters, and each of them has been found in various waters.

The spectrographic detection limits vary for different elements. The detection limits for arsenic, cadmium, and zinc, for example, are relatively poor; whereas, the limits for silver and copper are very good. As a result of poor spectral sensitivity and (or) low concentrations, a few of the elements are frequently not detected and must be reported as less than some computed value. Where the combination of these two factors made the lack of success of finding a given element a reasonable certainty, it was not sought. Notable among the elements of probable environmental significance that fall into this category are arsenic and selenium.

The detection limit for each of the elements, in terms of the concentration in the water sample, varies inversely with the quantity of total solids in the sample, whenever the dried residue is used for analysis. This is true because the minimum detectable concentration of an element in the residue is a relatively high concentration with respect to water containing abundant solids and a lower concentration with respect to water low in solids. In spite of this limitation, the residue technique used, involving greater preconcentration, has considerably better limits of detection than so-called direct methods such as the rotating disk, porous cup, vacuum cup, and plasma jet.

PROCEDURE

Before any aliquot of a sample was taken, the sample was homogenized by placing the polyethylene bottle in an ultrasonic bath for 30 minutes or in an especially modified paint shaker for 3 minutes.

Because the elements were determined as a percentage of the analyzed solids but reported in terms of micrograms per litre of water, it was necessary to determine the total solids per litre. A carefully measured 50-ml aliquot was evaporated to dryness in a weighed 2- to 2½-inch clean glass bulb prepared from a Christmas ornament. A heat lamp placed above the bulb served as the heat source. When the water was completely evaporated, the bulb was dried at 180°C for 1 hour, cooled to room temperature in a desiccator, and weighed. From the increase in weight, the total solids per litre was calculated. This determination was repeated on a different day and the average of the two values used if the duplicates differed by not more than 3 percent from the average. Otherwise, the determination was repeated.

The residue used for analysis was prepared by evaporating to dryness, in a dish lined with plastic film, sufficient water to yield about 100 to 150 mg of solids. The residue was transferred as nearly quantitatively as possible from the plastic film to an agate mortar for homogenizing with an agate pestle. The residue was then transferred to a glass vial, dried in an oven at 180°C for 1 hour and stored in a desiccator.

The arcing mixture was prepared by weighing 3 parts high-purity graphite, 1 part sample residue, and 2 parts buffer-matrix[1] into a small polystyrene vial, in which a plastic ball had been placed, and mixing for 3 minutes in a small, high-speed, impact shaker. Duplicate charges of 24 mg of this mixture were weighed into high-purity graphite electrodes having a cup of inner diameter of 3.66 mm, a wall thickness of 0.38 mm, and a depth of 6.10 mm. The sample end of the electrode was preheated to red heat in a flame to drive off any remaining water or gases that might cause expulsion of the sample during arcing.

[1] The matrix material has an element composition approximating the average major-element composition of North American waters (Clarke, 1924) and is made by thoroughly mixing the following high-purity compounds in the proportions indicated: $CaCO_3$, 46.6 percent; $MgSO_4$, 23.2 percent; NaCl, 18.3 percent; KCl, 3.6 percent; and SiO_2, 8.3 percent. The addition of 2 parts matrix to 1 part sample helps to minimize differences among samples and between the sample and the standards used for quantitative comparison, and minimizes the variation in the intensity of minor-element lines due to differing matrices. Occasionally, a particular element may be diluted below its detection limit but is sacrificed in the interest of increased accuracy for the others.

TABLE 5.—*Analytical lines for emission spectrographic analysis of water residues*

[The two-step method of emulsion calibration described in Method E 116-67 (American Society for Testing and Materials, 1968) was used]

Element	Wavelength (nm)	Lower limit of determination (percentage of arcing mixture)
Aluminum (Al)	308.216	0.0003
	266.039	.005
	265.249	.01
Barium (Ba)	455.404	.0001
Beryllium (Be)	313.042	.0002
Bismuth (Bi)	306.772	.0005
Boron (B)	249.773	.0005
	249.678	.001
Cadmium (Cd)	326.106	.002
Chromium (Cr)	302.156	.0005
Cobalt (Co)	345.350	.0005
Copper (Cu)	324.754	.0001
	327.396	.0002
Gallium (Ga)	294.364	.0002
Germanium (Ge)	303.906	.0005
Iron (Fe)	271.902	.0005
	272.358	.002
Lead (Pb)	283.307	.0005
Manganese (Mn)	257.610	.0005
	294.920	.001
	293.306	.002
	257.276	.02
Molybdenum (Mo)	317.035	.0001
	320.883	.001
Nickel (Ni)	341.476	.0005
	305.082	.001
Silver (Ag)	328.068	.00005
Strontium (Sr)	460.733	.0001
Tin (Sn)	317.502	.001
Titanium (Ti)	323.452	.0005
	324.199	.002
Vanadium (V)	318.341	.0005
Zinc (Zn)	334.502	.02
	334.557	.1
Zirconium (Zr)	327.926	.001
	327.305	.002

A series of comparison standards were prepared by incorporating the elements to be determined in the buffer-matrix. The concentration factor between each successive standard was 0.464 giving three standards per order of magnitude. The lower end of the series was at the detection limit of the most sensitive analytical line.

Internal standards were not used. Instead, the complete series of external standards were recorded on each plate, along with the spectra of the duplicated samples. This technique was found to be more precise than internal standardization with standards and samples on separate plates.

The excitation and exposure conditions were as follows:

Voltage .... 220 (open circuit) volts
Amperage .... Gradual increase from 0 to 8 A during first 7 sec, then increase to 12 A
Spectral region .... 220–350 nm
Slit width .... 25 μm
Slit length .... 1.4 mm
Neutral filter .... 24 percent transmission
Length of burn .... To completion

The percentage transmittances of the analytical lines were measured with a microphotometer and a background reading made near each line. The lines most frequently used, together with the lower limit of determination for each line, are given in table 5. With the line-width measuring device described by Barnett (1967), the upper limit of the concentration range for any line is that of the standard.

A computer program was used to calculate the concentrations of the elements in the residue and the concentrations in the water from the densitometric data.

On each plate of samples and primary standards were also recorded the duplicate spectra of one of four so-called Standard Reference Water Samples (SRWS). These are water samples prepared in one of the laboratories of the U.S. Geological Survey and analyzed by a number of laboratories of the U.S. Geological Survey and analyzed by a number of laboratories throughout the Nation. The determinations on these secondary standards were used as a check on the quality of the results on a particular plate. The long-term precision of analysis is given for each SRWS in table 6. A more detailed account of the method of analysis is given by Barnett and Mallory (1971).

TABLE 6.—*Precision[1] of analysis of water residues from four standard reference water samples*

[GM, geometric mean concentration as micrograms per litre in the water sample; GD geometric deviation, or antilog of standard deviation of the logarithms of the concentrations; number of determinations on which estimates are based is approximately 50 for SRWS-25, 25 for SRWS-28 and SRWS-32, and 17 for SRWS-38]

Element	SRWS-25		SRWS-28		SRWS-32		SRWS-38	
	GM	GD	GM	GD	GM	GD	GM	GD
Al	554	1.070	620	1.077	211	1.097	401	1.081
3a	36.8	1.110	578	1.100	51	1.103	28	1.067
B	.....	.....	5.4	1.156	15.5	1.132	9.9	1.163
Cr	13.3	1.104	21.8	1.092	9.9	1.119	52.8	1.092
Cu	76.6	1.131	274	1.135	620	1.115	125	1.132
Fe	390	1.075	140	1.084	902	1.046	1200	1.078
Pb	14.6	1.140	77.9	1.088	26.6	1.106	15.9	1.110
Mn	312	1.100	75.0	1.121	126	1.152	39.5	1.214
Mo	.....	.....	3.2	1.117	18.1	1.072	9.3	1.112
Ni	14.0	1.261	4.8	1.170	29.6	1.153	4.2	1.372
Ag	4.2	1.322	14.9	1.269	.....	.....	.....	.....
Sr	252	1.149	56.0	1.095	251	1.116	158	1.128
V	4.8	1.126	.....	.....	4.6	1.155	7.6	1.118
Zn	.....	.....	580	1.092	603	1.085	187	1.076

[1]Precision is expressed as a geometric deviation, GD. For example, about 68 percent of the aluminum determinations on sample SRS-25 are in the range from 518 (554/1.070) to 593 (554×1.070) and about 95 percent are in the range from 483 (554/(1.070)²) to 634 (554×(1.070)²).

ANALYSIS OF WATER FOR RADIOACTIVITY

By V. J. JANZER

As water passes through the hydrologic cycle, it gradually dissolves traces of the rocks, soils, and other materials with which it comes in contact. All of these materials contain trace or greater amounts of radioactivity if due only to the presence of potassium. Light-colored sedimentary rocks such as salt beds, gypsum, limestone, and dolomite, with the exception of potash deposits, usually contain the lowest concentrations of the radioelements normally found in rocks. Thorium occurs

primarily in deposits of monazite sands and in some relatively rare minerals. Black marine shales and some phosphate rocks contain the highest concentrations of uranium. For example, at one time the black marine Gassaway Member of the Chattanooga Shale of Devonian age, with a uranium content locally greater than 50 ppm, was considered to be a potential source of uranium (Pollaro and others, 1958).

Uranium ore deposits consisting of uraninite and pitchblende occur in igneous and metamorphic rocks, whereas secondary deposits such as carnotite and coffinite occur in sedimentary rocks. Generally, the effect of these deposits on water quality is very localized and relatively minimal, although uranium concentrations in water near ore bodies may exceed several milligrams per litre.

Either natural or manmade radioactivity, or both, are found in varying concentrations in all precipitation, surface, and ground waters. The natural radioactivity which is present in these waters is due to traces of (1) the long-lived parent members of the three natural radioactive decay series (uranium-238, uranium-235, and thorium-232) and their daughters, (2) single long-lived radioactive elements, with potassium-40 being the most abundant, and (3) relatively short lived cosmic-ray-produced radionuclides such as tritium and carbon-14. Tritium and carbon-14 are also produced by nuclear devices.

Several hundred manmade radionuclides are present worldwide as a result of the use of nuclear reactors and the detonation of many fission and fusion devices. The relatively long lived radionuclides such as cesium-137, strontium-90, and iodine-131 can be detected in many environmental samples, and under some conditions they may constitute serious health hazards.

Radioactive decay of all these nuclides results in the emission of alpha, beta, or gamma radiation either singly or in combination. These emissions can be counted using the appropriate detectors and electronic instruments.

Determination of the gross alpha and gross beta-gamma radioactivity of a water sample is a relatively simple and rapid analysis which is useful as an indication of total sample radioactivity or as a screening technique. If radioactivity is detected that is above the normal range for the type of sample analyzed, more definitive analyses for specific elements may be required. However, identification and measurement of the specific alpha and beta emitters in a sample is often extremely difficult and time consuming. In most cases, such detailed knowledge is unnecessary and a gross alpha and beta value is adequate.

A gross value merely indicates that the radioactivity of unknown origin in a sample is equivalent to a known amount of a specified calibration isotope.

The counting rates of a series of calibration planchets, each containing a known amount of alpha or beta radio-activity, and a range of solid weights obtained by evaporation of simulated natural waters, are used to prepare a graph of counting efficiency versus weight of residue for each radiation counting system that is used. The counting efficiency factors from these curves are used to calculate and express the radioactivity of each sample in terms of equivalent fractional curies per litre of the respective calibration isotopes.

A curie (Ci) is that amount of radioactivity which yields 3.7×10^{10} disintegrations per second. A picocurie (pCi) or 1×10^{-12} curie yields 2.22 disintegrations per minute (dpm) and is one of the more common terms used to express the low level of radioactivity usually associated with environmental samples.

Natural uranium is commonly used as an alpha-calibration isotope. One microgram (μg) of natural uranium contains $\frac{1}{3}$ picocurie of uranium-238, an equilibrium amount of uranium-234, or an additional $\frac{1}{3}$ pCi, and a trace of uranium-235. The total activity associated with 1 μg of natural uranium is thus approximately 0.68 pCi.

Two of the most common beta calibration isotopes are cesium-137 and strontium-90. Cesium-137, with half-life ($T_{1/2}$) of approximately 30 years, decays by beta emission to the 3-minute half-life metastable barium-137 which in turn decays to barium-137 by emission of a 0.662 MeV gamma ray. One beta and one gamma event are thus associated with each cesium-137 atom which decays.

The proportional counter used for counting alpha and beta particles can also detect gamma rays but with a greatly reduced efficiency.

Strontium-90 ($T_{1/2} \cong 28$ days) decays by beta emission to yttrium-90 ($T_{1/2} \cong 3$ days) which decays in turn by another beta emission to stable zirconium-90. Two beta events are thus associated with each strontium-90 atom which decays.

Gross beta values for a sample will differ slightly depending upon which of the calibration isotopes is used as a reference.

If the detection system used "saw" 100 percent of the beta emissions for every calibration isotope, it would register two counts for each atom of strontium-90 which decayed. (This assumes that sufficient time has elapsed for a complete ingrowth of the yttrium-90 daughter.) The system would register only one count for each atom of cesium-137 decaying by beta emission. Because of the lower detection efficiency of the counting system for gamma rays associated with the cesium atom (metastable barium-137) decay, it might only "see" one-tenth of the gamma events. The total counts registered for each cesium atom decay would thus average about 1.1 cpm.

Consequently, each count registered and considered as a strontium beta represents less activity than one count considered as a cesium beta. Thus, the gross beta radio-

activity of a sample expressed as strontium-90 will be approximately 20 percent lower than if the same amount of radioactivity is expressed as cesium-137. No direct conversion can be made from one value to the other, however. Similar problems exist in attempting to convert (or compare) gross values expressed in terms of other calibration isotopes.

Analysis of a water sample for gross alpha and gross beta-gamma radioactivity generally requires no more than 1 litre or a sample aliquot selected to yield, on evaporation, a residue weighing approximately 50–150 mg. This residue is transferred to a stainless steel counting planchet in the form of a thin deposit of uniform size and thickness. The planchet is then counted for alpha and beta radioactivity using a low background proportional counter calibrated for counting efficiency (Barker and Robinson, 1963). The sample count rates, corrected for background, are then used to calculate the gross alpha and beta radioactivity for each sample in terms of micrograms or picocuries, respectively, for the calibration isotope used, per litre of water.

All gross alpha and gross beta radioactivity analyses of water samples collected for the geochemical survey of Missouri were made by Robert S. Dewar and P. K. Roscio.

METHODS OF DATA MANAGEMENT AND AUTOMATIC DATA PROCESSING

Extensive use was made of the U.S. Geological Survey RASS-STATPAC system of computer programs developed specifically for data from studies in field geochemistry. The programs are in FORTRAN IV and have been used on an IBM 360/65 in Washington, D.C., accessed from Denver, Colo. through a 360/20 terminal.

The Rock Analysis Storage System (RASS) programs are used for storage and selective retrieval of laboratory data on rock, soil, plant, or water samples. Coded descriptions of the samples, and identification of the localities from which the samples were collected, were provided by the investigator at the time the samples were submitted to the laboratories. The coded descriptions and the laboratory results were received by a permanent data processing group which entered both into the RASS file. The RASS file is maintained on a magnetic disk; each record consists of an alphanumeric portion containing identification of the sample and where it was collected, the collector, the coded sample description, and some administrative data, followed by a numeric portion containing the laboratory determinations. Each record also contains free format comments about the sample as provided by the investigator. These comments are the only part of the record that cannot be queried on retrieval.

Selected records can be retrieved from the RASS file according to sample laboratory number, or by specifying such criteria as name of collector, locality of collection,

type of material, date of submittal to the laboratory, or by any combination of these. They may also be retrieved by specifying ranges or upper and lower limits for any of the analytical determinations. Thus, it is possible to retrieve, say, all records on granites collected by R. J. Ebens from St. Francois County that contain more than 80 percent SiO_2 and 10–15 percent Al_2O_3. The alphanumeric and numeric portions of each record are printed in either one of two formats, and the numeric portion may also be punched on cards or written on a magnetic tape or disk for further computer processing. The formats of the cards and the magnetic devices are directly acceptable to the STATPAC programs for automatic plotting and statistical data reduction.

A generalized diagram of the RASS-STATPAC system is given in figure 4. It shows that numeric data retrieved from the RASS file may be entered onto a magnetic device in the STATPAC format and, from there, passed into three other types of computer programs. Output from the programs for data preparation are generally printed, but are also written on another magnetic device for further processing. Data preparation commonly consists of data transformations, such as conversion to logarithms, or changes in units, such as conversion from weight percent of the oxide as may have been reported by the analyst, to parts per million of the element as may be required for some further computation. It consists also of computation of a correlation or some similar matrix for further examination by methods of factor analysis, or may consist simply of selection of some part of the total data from the STATPAC device.

The statistical programs in the STATPAC system include most of the univariate, bivariate, and multivariate techniques that have proved to be useful in geochemical studies.
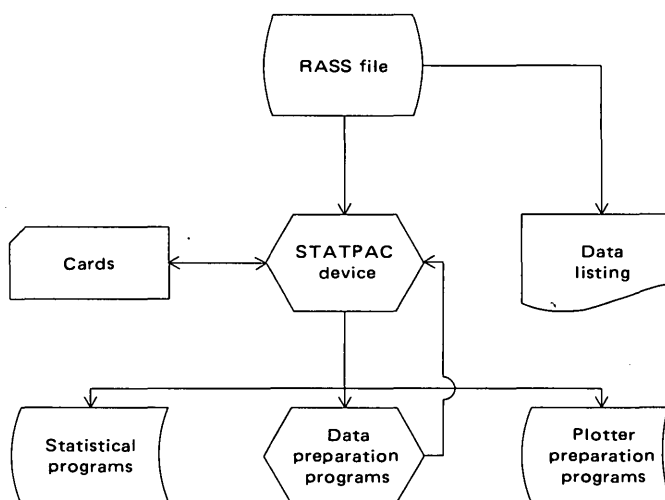

FIGURE 4.—Generalized diagram of the RASS-STATPAC system of computer programs for the management and statistical reduction of geochemical data.

The programs for the preparation of plotter tapes allow preparation of the data for plotting on symbol maps, contour-type maps, triangular diagrams, or simple x-y graphs.

Because the RASS-STATPAC system has been designed specifically for geochemical data, each numeric quantity entered into the system may be qualified, as are many of the geochemical values produced by the laboratories. Five types of qualifying codes are used: N (not detected), L (less than), T (trace,) G (greater than), and H (analytical interference). (The code "H" is used here rather than "I" because of possible confusion of the latter with the number "1.") In addition, a sixth code of B is used to indicate a blank, or the fact that no value is available for some unspecified reason. No unqualified values of zero are entered into the RASS data files because zero is regarded as an unrealistic geochemical value. An analyst's report of 0.00 percent, for example, would be recorded in RASS as less than 0.01 percent, or actually as L 0.01. Some further examples of the manner in which data values are stored in RASS, and their meanings, are as follows:

N 0.002   Constituent was not detected using an analytical method having a lower limit of determination of 0.002 percent.

L 0.01   Constituent may be present, but in a concentration of less than 0.01 percent.

T 0.0001   Constituent appears to be present, but in very minor concentrations too low to be measured by the analytical method used. The method has a lower limit of determination of 0.0001 percent.

G 10.0   Constituent is present in a concentration too high to be measured by the analytical method employed—greater than 10.0 percent.

H 0.0005   Determination of the constituent is interferred with by the presence of another constituent. Maximum concentration is 0.0005 percent.

B 0.0   No data available.

0.03   Unqualified estimate of the concentration present. Most constituents are recorded in units of either percent or parts per million.

N 0.0
L 0.0
T 0.0   Same as above, but lower or upper limits of
G 0.0   concentration are not specified.
H 0.0

Most of the univariate statistical computer programs in the STATPAC system, and the plotter tape preparation programs, accept these qualified analytical values directly and treat them as is appropriate for the calculation being performed. Some of the programs for bivariate statistical analysis also accept the qualified analytical values. The programs for analysis of variance and for multivariate statistical analysis, however, require data matrices consisting entirely of unqualified values. Accordingly, one of the data preparation programs allows the user to replace a qualified geochemical value with any unqualified value he feels is appropriate before proceeding with analysis of variance or multivariate methods. These replacement values may be means representing similar kinds of specimens, or they may be more or less arbitrary values immediately outside the range of determination for the analytical method used. In general, qualified values are replaced with unqualified values only where they compose fewer than 20 percent of the values for that constituent in the data set being examined. In most instances, values qualified with N or L were replaced with a value equal to seven-tenths of the lower limit of determination of the analytical methods used. For example, where 80 percent of the values form a continuous frequency distribution censored at the lower limit of analytical determination, say 0.01 percent, it seems more reasonable to assume that the remaining values cluster near 0.01 than near zero (an infinitely low value on a logarithmic scale). For purposes of computation the remaining values would be treated as 0.007. However, if the final geochemical interpretations depend on whether they are treated as, say, 0.007, 0.005, or 0.001, there is no justification for the procedure at all. The only possible justification for the replacement of qualified analytical values with values that are unqualified is that the final geochemical interpretation is largely insensitive to the choice of unqualified values used; that is, any reasonable assumed value would lead to about the same final geochemical interpretation.

FREQUENCY DISTRIBUTIONS AND DATA TRANSFORMATIONS

More than one-half of the data accumulated in the geochemical survey of Missouri consist of semiquantitative spectrographic analyses (Myers, and others, 1961) which are reported in a series of six geometric classes for each order of magnitude. That is, the data as received are grouped in statistical terminology, and, therefore, can be analyzed using grouped data methods. The theoretical boundaries of the groups, to two figures, are as follows: 0.000083, 0.00012, 0.00018, 0.00026, 0.00038, 0.00056, 0.00083, 0.0012,..., 5.6, 8.3, and 12 percent, and the corresponding geometric midpoints are 0.0001, 0.00015, 0.0002, 0.0003, 0.0005, 0.0007, 0.0010,..., 7, and 10 percent. A geometric series is used because the line density on the spectrographic plate is approximately proportional to the log of the concentration of the element, and the expected error in reading line densities is, consequently, also logarithmically related to concentration. The semiquantitative spectrographic method, thus, is like most other methods used for the determination of minor constituents in natural materials; the error variance on an

arithmetic scale is generally proportional, in at least an approximate way, to the amount of the constituent present. For this reason alone, a logarithmic transformation of all minor element data was appropriate. Log transformation of the spectrographic reporting classes listed above yields a series of values equally spaced on an arithmetic scale, and thus provides a convenient and appropriate means for expressing the data in a form suitable for conventional statistical procedures.

The proportional nature of laboratory error is paralleled in sampling. Even when a highly precise laboratory method is used, it is found that the geochemical variation in a rock or soil is at least approximately proportional to the average amount of the constituent present. On log transformation of the data, the variance either is homogeneous across all parts of the rock or soil unit, or it varies independently of the average. Thus logarithmic transformation of the data tends to homogenize both the variance arising from laboratory treatment and that actually present in the material being sampled. Homogeneous variance is assumed in attempts to extract variance components in the analysis of variance and in most statistical tests for significant variation between two or more populations or strata within populations.

There are a number of helpful consequences of the log transformation other than the tendency to equalize the variances. One of these for minor element data is that the frequency distributions of the logarithms are almost always closer to the normal (gaussian) form than are the original data. Consequently, the requirements of the data for probability tests are nearly always more closely satisfied. Statistical estimates in general are more stable (that is, with smaller error variances) where the data are symmetrically distributed about their means. Moreover, statistical methods for treating censored normal distributions, such as the one described by Cohen (1959, 1961), are valid and applicable.

Another helpful consequence of the log transformation in treating geochemical data is that the variance estimates are independent of the manner in which the original data are expressed. The log variance for titanium, for example, is the same whether the original data are as percent TiO_2, parts per million TiO_2, percent Ti, or parts per million Ti.

Finally, the appropriateness of the log transformation of geochemical data prior to statistical analysis might be obvious from the widespread practices of experienced geochemists who intuitively judge the significance of geochemical differences on a proportional basis. Thus, a geochemical anomaly is spoken of as so many times above background, rather than as being of so many parts per million. Examination of logarithmic variation by means of analysis of variance is a perfectly analogous procedure; proportional rather than absolute variation is being examined and appraised.

Most analytical data were transformed to logarithms whether reported by the analysts in geometric classes or not. There were, however, a few exceptions. Silica in highly siliceous materials (granites, sandstones, and some soils, for example) nearly always displays a frequency distribution with large negative skewness, and log transformation appeared to be wholly inappropriate because it magnifies this skewness. Similarly, a few other major constituents in some rocks and soils, measured largely by the relatively precise X-ray fluorescence method, were almost symmetrically distributed and also did not appear to warrant log transformation. Log transformation here would have imposed a negative skewness on the data.

Log transformations were not used in preparation of the data for Q-mode factor analysis because this method, as used in the Missouri study, requires that all determinations for a given specimen sum to a constant, generally 100 percent or 10^6 ppm. As the determinations included all major and most commonly determined minor constituents, each elemental determination was converted to percent or parts per million as the common oxide (except for ionic constituents in water and a few other constituents that tend to occur in elemental form in various materials), and the analysis was adjusted to sum to 100 or 10^6. The values for each constituent were then scaled to range from zero to unity, in the manner suggested by Manson and Imbrie (1964) for giving each variable approximately equal weight in the factor analysis outcome. The factor analysis method as used in these studies does not presuppose any given form for the parent frequency distributions.

DETECTION RATIOS

Every analytical method is applicable over a limited range of concentrations. The semiquantitative spectrographic method used for the analysis of rocks, soils, and plant ashes from Missouri, for example, is used to measure silver concentrations in the range 0.0001 to 10 percent; the X-ray fluorescence method is used to measure most major oxides in the range 0.01 to 100 percent. Concentrations determined to occur outside these ranges are reported as qualified values using N (not detected), L (less than), or G (greater than). Therefore, in statistical analysis of the resulting data the situation is commonly encountered where some of the data for a particular constituent in a particular suite of samples are straightforward estimates of concentration, and some are qualified—that is, maximum or minimum possible concentrations. The frequency distribution of the data is censored, generally at either the lower or upper end. Doubly censored distributions have not been encountered in the Missouri survey.

The estimation of means and variances from censored data can be accomplished by use of rigorous and efficient formal procedures, and approximate estimates of variance

components, bivariate statistical properties, and multivariate models can be made, under certain conditions, by assuming some reasonable approximate values for the missing data. In addition, it is commonly useful to estimate statistical properties for the unqualified values only, but the final estimates must be clearly labeled as applicable to only part of the total range of the population. Regardless of the course chosen to treat the censored data, it is imperative that reports of the final statistical estimates be accompanied by some indication of the degree to which the raw data were censored. For this purpose, we have used the *detection ratio*. The detection ratio is given in the form, for example, of 295:300, where the numerator gives the number of data values that are unqualified and the denominator gives the total number of samples that were analyzed. Detection ratios are given for each constituent included in the geochemical survey of Missouri—generally in tables of estimated mean concentrations, but the ratios are also applicable to the data as used subsequently in bivariate and multivariate statistical analyses.

ESTIMATES OF MEANS AND TOTAL VARIANCE

As previously discussed in the section on sampling designs, the purpose of the geochemical survey of Missouri has been to estimate and identify geochemical variation over the State, and not to estimate mean compositions of the various natural materials. Therefore, the sampling schemes were designed for this purpose and in a strict sense are not valid for the latter. Nevertheless, estimates of mean concentration are of common interest to geochemists and so are included in the final survey results even though they are admittedly biased to varying degrees. Because the sampling schemes have given equal weight to all areas of the State where the population of interest occurs, the bias will be maximum where the population is highly concentrated, and of unusual composition, in restricted areas. The bias will be minimum where large-scale geographic variation in the composition of the population is small or absent.

Because most of the frequency distributions are, by far, more nearly symmetrical on a logarithmic scale than on an arithmetic scale, most averages are estimated as geometric means. The geometric mean is the antilogarithm of the arithmetic mean of the logs:

$$x = \log_{10} y, \tag{8}$$

where y is the analytical value in percent or parts per million,

$$\bar{x} = \frac{\sum x}{n}, \tag{9}$$

where n is the number of values of x, and $\bar{x}$ is the arithmetic mean of the logarithms, and

$$GM = 10^{\bar{x}}, \tag{10}$$

where GM is the geometric mean of y. If the detection ratio is less than one (that is, some qualified values are present),

the mean logarithm is first estimated for the unqualified values only:

$$\bar{x}' = \frac{\sum x}{n'}, \tag{11}$$

where n' is the numerator of the detection ratio. The estimate of the mean logarithm for the entire population is then given by a relation from Cohen (1959):

$$\bar{x} = \bar{x}' - \lambda(\bar{x}' - x_0), \tag{12}$$

where x_0 is the logarithm of the point at which the frequency distribution is censored (that is, the upper or lower limit of the range for the analytical method), and λ is from graphs by Cohen (1959, p. 231; reproduced in Miesch, 1967b, p. B7). The geometric mean, as before, is the antilogarithm of $\bar{x}$.

Where a log transformation of the data has not been made, averages are estimated as the conventional arithmetic mean:

$$AM = \frac{\sum y}{n}, \tag{13}$$

where AM is the arithmetic mean.

In cases where the data warrant logarithmic transformation, and estimates of the arithmetic mean are desired, they are derived by the method of Sichel (1952, 1966; see also Miesch, 1967b):

$$AM = \tau GM, \tag{14}$$

where τ is from values tabulated by Sichel (1952, 1966) or from graphs by Miesch (1967b, p. B8).

Even though the geometric mean is a more appropriate measure of central tendency than the arithmetic mean where the frequency distribution of the data is nearly symmetrical on a log scale, arithmetic means are desirable in some instances as estimates of geochemical abundance (Miesch, 1967b), or for direct comparison with arithmetic means from the literature.

Where the data, y, are logarithmically transformed, the variance of the logarithms is estimated by:

$$s_x^2 = \frac{\sum (x - \bar{x})^2}{n - 1}. \tag{15}$$

If the detection ratio is less than one (that is, some qualified values are present), the variance of the logarithms is first estimated for the unqualified values only:

$$s_x'^2 = \frac{\sum (x - \bar{x})^2}{n' - 1}, \tag{16}$$

where n', as before, is the numerator of the detection ratio. The estimate of the logarithmic variance for the entire population is then derived from another equation of Cohen (1959):

$$s_x^2 = s_x'^2 + \lambda(\bar{x}' - x_0)^2, \tag{17}$$

where λ, as in equation 12, is from graphs by Cohen (1959) reproduced in Miesch (1967b). For convenience in interpretation, the variance of the logarithms is converted to the geometric deviation, GD, where:

$$GD = 10^{s_x}. \tag{18}$$

Where the data are derived from a lognormally distributed population, then 68 percent of the population values are estimated to occur in the range GM/GD to $GM \times GD$, and 95 percent of the population is estimated to occur in the range from GM/GD^2 to $GM \times GD^2$. The range in which 95 percent of the population is estimated to occur is referred to as the *expected range* (Ebens and others, 1973) in subsequent reports of the Missouri study.

Where the analytical data were not logarithmically transformed, the conventional estimate of the standard deviation, rather than the geometric deviation, is used:

$$SD = \left(\frac{\Sigma (y - \bar{y})^2}{n-1} \right)^{1/2},   (19)$$

and the expected range is estimated to be from $AM - 2SD$ to $AM + 2SD$.

Because stratified sampling designs were used, and the various strata commonly vary greatly in size, it is commonly useful to estimate a weighted mean. The weighted mean logarithm is estimated by:

$$\bar{x}_w = \frac{\Sigma W_i \bar{x}_i}{\Sigma W_i},   (20)$$

where W_i and $\bar{x}_i$ are, respectively, the relative size and the mean logarithm for the ith strata.

The weighted geometric mean, GM_w, is the antilogarithm of $\bar{x}_w$. Where the data have not been transformed logarithmically, the weighted arithmetic mean, AM_w, is estimated in the conventional manner.

Some of the statistical estimates discussed in this section are evaluated for two typical groups of data in table 7.

TABLE 7.—*Some univariate statistical estimates used in the geochemical survey of Missouri*

	Barium[1]		Cobalt[1]	
	y_i (ppm)	$x_i = \log_{10} y_i$	y_i (ppm)	$x_i = \log_{10} y_i$
Data	30	1.4771	< 0.83	.....
	30	1.4771	< .83	.....
	50	1.6990	1.5	0.1761
	100	2.0000	3	.4771
	30	1.4771	1	.0000
	70	1.8451	< .83	.....
	10	1.0000	< .83	.....
	50	1.6990	5	.6990
	30	1.4771	1	.0000
	30	1.4771	2	.3010
	70	1.8451	2	.3010
	30	1.4771	3	.4771
Detection ratio	12:12		8:12	
Mean logarithm	$\bar{x} = 1.5792$ (equation 9)		$\bar{x}' = 0.3039$ (equation 11) $\bar{x} = 0.0961$ (equation 12)[2]	
Geometric mean	$GM = 38$ ppm (equation 10)		$GM = 1.25$ ppm (equation 10)	
Log variance	$s^2 = 0.0677$ (equation 15)		$s'^2 = 0.0596$ (equation 16) $s^2 = 0.1396$ (equation 17)[2]	
Log standard deviation.	$s = 0.2603$		$s = 0.3736$	
Geometric deviation.	$GD = 1.82$ (equation 18)		$GD = 2.36$ (equation 18)	
Expected range	11–126 ppm (see text)		0.2–7 ppm (see text)	
Arithmetic mean	$AM = 45$ ppm (equation 14)[3]		$AM = 1.8$ ppm (equation 14)[4]	

[1]Data pertain to the sandstone in the Ordovician Roubidoux Formation of Missouri (from R. J. Ebens and J. J. Connor, written commun., 1973).
[2]The value of x_0, the logarithm of the lower limit of analytical determination, for equations 12 and 17 is -0.0809. The value of λ for equations 12 and 17 is 0.54 and was determined from a graph of Cohen (1959, p. 231) which gives λ as a function of h, computed from the detection ratio ($h = (12-8)/12 = 0.33$), and the quantity $s'^2/(\bar{x}' = x_0)^2$.
[3]The value of τ for equation 14 is 1.19. (See table A of Sichel, 1966.)
[4]The value of τ for equation 14 is 1.41. (See table A of Sichel, 1966.)

MEASUREMENT OF CORRELATION AMONG COMPOSITIONAL VARIABLES

Estimates of correlation among chemical constituents are helpful in understanding the geochemical behavior of the constituents, and have been used in interpretation of some of the Missouri data. The simple correlation coefficient is computed as the covariance divided by the geometric mean of the corresponding variances. Simple correlations among logarithms measure the degree to which the log data fit a straight line on a common x-y plot. Correlations of +1.0 indicate that the data, or the log data, exactly fit a straight line of positive slope; correlations of $^-1.0$ indicate that the slope is negative. Correlations of zero indicate the total absence of a linear relation between the variables.

Where either of the two frequency distributions is censored (that is, where the detection ratios for either of the variables is less than one), the method of estimation of the correlation depends on the way in which it is to be used. If the correlations are to be analyzed by means of R-mode factor analysis methods, the censored (qualified) analytical values are replaced by some reasonable, although arbitrary, estimates of the actual concentration. This assures that the correlation matrix possesses the Gramian properties required for factor analysis. If the correlation matrix is not to be further analyzed by factor analysis methods, the qualified analytical values are ignored, and the correlation is estimated for the uncensored part of the bivariate distribution only. In this case, the correlation coefficient is regarded as an *index of geochemical association* only, and because its frequency distribution is complex and unknown, no attempt is made to judge its statistical significance.

Chayes (1960) showed that correlations among compositional variables do not necessarily reflect petrologic, or geochemical, relationships because these variables must sum to a constant for each specimen. His mathematical argument is directed at untransformed compositional data, but would appear to hold in principle, if not in mathematical detail, for logarithmic data as well. As suggested by Miesch, Chao, and Cuttitta (1966), however, the problem is probably not important where both compositional variables are present in minor concentration. If either variable is present in major concentration the covariance, and consequently the correlation coefficient, does not necessarily reflect a genetic, or geochemical, relation. The test of the correlation recommended by Chayes and Kruskal (1966) was not used, because of arguments presented elsewhere (Miesch, 1969).

ESTIMATION OF VARIANCE COMPONENTS

Estimation of the mean squares at various levels of the hierarchical sampling designs, and, therefore, the variance

components represented in equation 3, was by the method described by Anderson and Bancroft (1952, p. 325–330), as mentioned previously. Where the sampling designs were complete (that is, equal numbers of items across all categories at each level of the design), the method of Anderson and Bancroft is perfectly equivalent to that described in detail by Krumbein and Slack (1956) and by numerous authors in various statistical texts. The primary advantage of the method of Anderson and Bancroft is that sampling designs are commonly intentionally or inadvertently incomplete. They are intentionally incomplete when there is a wish to redistribute the degrees of freedom from that ordinarily encountered in a complete hierarchical design. For example, if only a portion of the specimens are analyzed in duplicate for estimation of the error variance term at the lowest level, then the degrees of freedom can be increased at some higher levels with no increase in the number of specimens to be analyzed. Inadvertently incomplete designs occur when, for example, certain samples were not available in the field, or could not be obtained.

In order to facilitate comparison of estimated variance components for different variables, each component can be expressed as a percentage of the total variance for that variable. Thus, if the estimated variance components for a particular variable are, say, $s_\alpha^2 = 0.3$, $s_\beta^2 = 0.0$, and $s_\gamma^2 = 0.2$, then the estimated total variance is 0.5, and the percentage components are, respectively, 60, 0, and 40. That is, 60 percent of the total variance is estimated to be at the upper level, 0 percent at the intermediate level, and 40 percent at the lowest level of the sampling hierarchy.

ESTIMATION OF SAMPLING AND ANALYTICAL ERROR VARIANCE

As pointed out previously in this report, and elsewhere (Miesch, 1967a), errors in sampling and laboratory analysis are of two fundamental types—bias and imprecision. Bias in sampling can occur when the areas, subareas, or points where samples are to be taken are not selected by some appropriate randomization procedure, and bias in laboratory analysis can occur where the analytical method is inherently incorrect. Attempts have been made to avoid serious bias in sampling by employing formal randomization procedures throughout, with only a few exceptions. Bias was avoided in the laboratory by careful selection and design of the analytical procedures and by repeated analysis of standard reference samples throughout the course of the analytical work.

[2]The variance of the laboratory analysis, as used here, refers to variance resulting from treatment of the samples from the time they are submitted to the laboratory to the time the laboratory results are received. Thus, analytical variance includes the effects of errors arising from sample preparation (crushing and grinding of rock and soil samples, ashing of plant samples, and so forth). This usage follows from a previous discussion of the subject of errors (Miesch, 1967a, p. A11). Because this notion of analytical variance is all inclusive, some writers prefer to use *procedural variance*.

The variances of the errors due to sampling and laboratory analysis[2] were estimated in the analysis of variance, either separately or combined. In some cases, randomly selected samples were split into duplicates and the duplicates were randomly interspersed with the remaining samples before submitting them to the laboratory. The analytical variance, then, was estimated as the lowest mean square, or variance component, in the analysis of variance. Where this was not done, the lowest mean square, or variance component, is properly interpreted as the analytical variance plus sampling error variance at the lowest level of the sampling design.

An estimated analytical variance of, say, 0.09 indicates an analytical standard error of 0.3 (square root of 0.09). If the variance had been estimated for the log data, the value of 0.3 is the log standard error, and the geometric error, GE, is 2.0 (antilog of 0.3). Therefore, the analytical method is reproducible within a factor of 2.0 at the 68-percent confidence level and within a factor of 4 (2.0 squared) at the 95-percent level. The interpretation of the geometric error, GE, therefore, is similar to that of the geometric deviation, GD, described previously. The geometric error of the mean for, say, two independent analyses of the same sample would be the antilog of the square root of the estimated error variance divided by 2, and in the example used previously:

$$GE = 10^{(s_\epsilon^2/2)^{1/2}} = 10^{(0.09/2)^{1/2}} = 1.6. \qquad (21)$$

Duplicate independent analyses of each sample, therefore, would reduce the geometric error from 2.0 to 1.6, and assuming that the laboratory method is unbiased, the mean of the two analyses would be within a factor of 1.6 of the correct value at the 68-percent confidence level, or within a factor of 2.56 (1.6 squared) at the 95-percent level.

Estimates of the variance arising from the imprecision of sampling also follow from the analysis of variance. If the sampling was done according to a hierarchical design, by sampling, say, points within subareas, subareas within areas, and areas within mapped units, sampling imprecision may occur at each of these three levels. Let s_β^2 represent the estimated variance among areas within mapped units, s_γ^2 the variance among subareas within areas, and s_δ^2 the variance among points within subareas. Assume that the last term, s_δ^2, does not include the analytical error variance. Then the variance of the sampling error due to the selection of points within subareas is given by:

$$\frac{s_\delta^2}{n_\delta}, \qquad (22)$$

where n_δ is the number of points sampled within each subarea. The variance of the sampling error due to both the selection of points within subareas and the selection of subareas within areas is:

$$\frac{s_\gamma^2}{n_\gamma} + \frac{s_\delta^2}{n_\gamma n_\delta}, \qquad (23)$$

where n_γ is the number of subareas sampled within each area. The total sampling error variance for a mapped unit is estimated by:

$$\frac{s_\beta^2}{n_\beta}+\frac{s_\gamma^2}{n_\beta n_\gamma}+\frac{s_\delta^2}{n_\beta n_\gamma n_\delta}, \qquad (24)$$

where n_β is the number of areas sampled within each mapped unit. The total error variance for the estimated mean of a mapped unit, due to both sampling and laboratory analysis, is:

$$\frac{s_\beta^2}{n_\beta}+\frac{s_\gamma^2}{n_\beta n_\gamma}+\frac{s_\delta^2}{n_\beta n_\gamma n_\delta}+\frac{s_\varepsilon^2}{n_\beta n_\gamma n_\delta n_\varepsilon}, \qquad (25)$$

where n_ε is the number of independent analyses made of each sample.

Expression 25 is equivalent to the right side of equation 7, and to equation 10.27 of Cochran (1963, p. 286) for the case where all sampling fractions approach zero.

Thus, the four terms in expression 25 represent, respectively, the error variances due to (1) sampling areas within the mapped unit, (2) sampling subareas within areas, (3) sampling points within subareas, and (4) analytical procedures. The numerators of these terms, the estimated variance components, indicate the importance of each source of error; their effects can always be reduced by increasing the appropriate values of n.

SIGNIFICANCE OF DIFFERENCES AMONG MEANS

Because of the general approach used in most phases of the geochemical survey of Missouri, the estimation of mean compositional properties of mapped units, the comparison of means is a crucial part of the statistical analysis of the data. Without statistical testing, we have no way to judge whether observed differences among means could have, or are unlikely to have, resulted from random errors in sampling or laboratory analysis. The conventional statistic for the comparison of two means that were selected before actually being computed is:

$$t=\frac{\bar{x}_1-\bar{x}_2}{s_x(\frac{1}{n_1}+\frac{1}{n_2})^{1/2}}, \qquad (25a)$$

where $\bar{x}_1$ and $\bar{x}_2$ are estimates of the two means from populations with equal variance, estimated by s_x^2, and are based on n_1 and n_2 independent measurements of x_1 and x_2, respectively. For a test at the 95-percent confidence level ($\alpha=0.05$), the computed t is compared with Student's t $(1-\frac{\alpha}{2}, v)$, where v is the degrees of freedom on which the estimate of s_x is based (equal to n_1+n_2-2).

If $n_1=n_2=n$, and the critical value of $t(1-\frac{\alpha}{2}, v)$ is used for t, equation 25a becomes:

$$\bar{x}_1-\bar{x}_2=R_2=\sqrt{2}\,t\,\frac{s}{\sqrt{n}}=1.41\,t\,s_M . \qquad (25b)$$

R_2 is the shortest significant range for the two means, and s_M is the standard error common to $\bar{x}_1$ and $\bar{x}_2$. The quantity $1.41t$ is referred to as the significant studentized range by Duncan (1955, p. 3). For $\alpha=0.05$, and, say, $v=12$,

the significant studentized range is 3.08. Thus, differences of at least 3.08 times the standard error of the means are required in order to declare the means as significantly different. If R_2 is in terms of logarithms, its antilog is the factor by which the means must differ in order for them to differ significantly at the 95-percent level of confidence.

Duncan (1955) has extended this method for the comparison of more than two means. The means are first ordered by magnitude. Then, if two adjacent means are being compared, p=2; if the two means being compared are separated by one other mean, p=3, and so forth. The significant studentized range increases some, but not much, with an increase in p. Where p=100, for example, again with v=12, the value is 3.48 as compared with 3.08 where p=2 ($\alpha=0.05$).

Therefore, tests of means by Duncan's method (the Multiple Range Test) involve α, v, p, and s_M. The selections of α and p are straightforward, and s_M is estimated as $\sqrt{E_s}$ in equation 7. The only uncertain value in the Missouri studies has been v. The uncertainty stems from the estimation of s_M from data collected according to the hierarchical designs previously described.

For example, consider 60 samples collected according to a simple two-level design as follows:

Number of mapped units 6
Number of samples from each unit (collected and analyzed independently) 10
    Total number of samples 60

Degrees of freedom for estimating variance among units 5
Degrees of freedom for estimating error variance 54
    Total degrees of freedom 59

The error variance is clearly estimated with 54 degrees of freedom because 10 independent measurements were obtained from each of 6 units. However, where more than two, say three, levels are involved, the situation may be as follows:

Number of mapped units 6
Number of areas sampled within each unit (selected at random) 5
Number of samples from each area (collected and analyzed independently) 2
    Total number of samples 60

Degrees of freedom for estimating variance among units 5
Degrees of freedom for estimating variance among areas within units 24
Degrees of freedom for estimating variance among samples within areas 30
    Total degrees of freedom 59

Thus, the error variance estimated by equation 7 is based on at least 24 degrees of freedom, but fewer than 54. There are only 5 completely independent values from each unit now, instead of 10, but each of these 5 is based on 2 independent observations. The number of actual degrees of freedom depends on the relative magnitudes of the variances among areas and among samples from within areas.

The uncertainty of v, however, is of little consequence. The significant studentized range for two means is 2.92 where v=24 and 2.84 where v=54. For v equal to infinity, the significant studentized range is reduced only to 2.77. For the comparison across seven means, the significant studentized range varies only from 3.47 for v=10 to 3.19 for v equal to infinity. Whether or not two or more means can be declared significantly different depends almost entirely on the differences among them and on their standard error.

Q-MODE FACTOR ANALYSIS

Each of the rock, soil, plant, and water samples collected in the geochemical survey of Missouri was analyzed for about 30 to 40 chemical constituents. Obviously, not all of these constituents are independent of each of the others. For example, most of the calcium and inorganic carbon, as well as much of the magnesium, occur together in rocks and soils in the carbonate minerals. The alkali metals occur together with aluminum and silicon in feldspar minerals and clay, and mercury is known to occur largely with carbon in organic matter. The result of this geochemical coherence is that variation in these 30 to 40 chemical constituents may be effectively described in terms of fewer than 30 to 40 factors. In fact, most of the variation in rocks, soils, and water can be described in terms of three or four factors. Factor analysis, therefore, offers a means for summarizing and reducing the problem of describing geochemical variation. To be sure, geochemical maps showing the geographic variation of each of the 30 to 40 chemical constituents over the State are required, but factor maps of three or four theoretical variables (factors) can serve to summarize the data and to show variation in general geochemical character.

The method of Q-mode factor analysis was introduced into the geological sciences by Imbrie and Purdy (1962) and Imbrie (1963), and its application to large data sets was made practical by the work of Klovan and Imbrie (1971). Klovan and Imbrie also showed a means for derivation of the normalized form of the factor scores. Extensions of the method of Q-mode analysis that may be applied to compositional data were given by Miesch (1975, 1976), and a computer program for the extended method was given by Klovan and Miesch (1975). The extended method requires that the variables in the matrix of data sum to a constant for each sample, but compositional data generally have, or can be adjusted to have, this property. In most situations the variables sum to 100 percent or 10^6 ppm.

The extended method will be illustrated here by applying it to a hypothetical problem. Consider that 20 samples of rock, soil, vegetation, or water have been collected according to the scheme represented in figure 5 and that the analyses of the samples for 6 constituents yield results as shown in table 8. The data in table 8 are
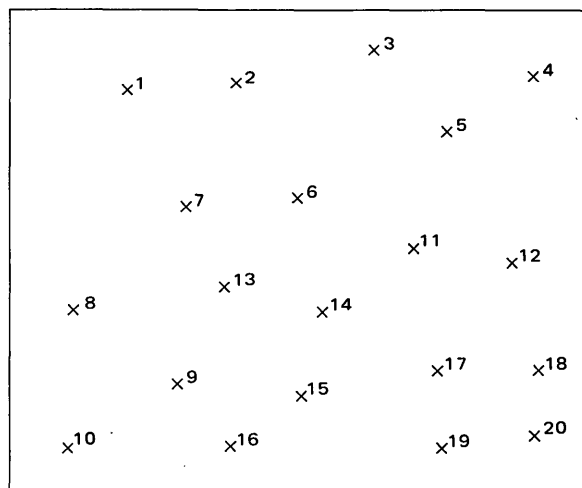

FIGURE 5.—Distribution of sampling localities for the Q-mode experiments. (See hypothetical data in table 8.)

TABLE 8.—*Hypothetical data for 6 constituents in 20 samples (as percentages)*

Sample No.	Constituent						
	A	B	C	D	E	F	Sum
1	82	8	6	3	1	0	100
2	77	11	6	4	1	1	100
3	85	13	0	1	0	1	100
4	72	10	9	6	2	1	100
5	80	16	0	2	1	1	100
6	78	7	9	5	1	0	100
7	67	13	9	7	2	2	100
8	74	6	12	6	1	1	100
9	81	12	3	2	1	1	100
10	86	9	3	2	0	0	100
11	62	20	6	7	3	2	100
12	67	21	3	5	2	2	100
13	57	23	6	8	3	3	100
14	61	24	3	6	3	3	100
15	41	36	3	10	5	5	100
16	46	33	3	9	4	5	100
17	53	1	27	14	3	2	100
18	57	6	21	11	3	2	100
19	53	5	24	13	3	2	100
20	58	2	24	12	2	2	100

constructed to have an underlying structure that is not readily apparent on inspection, just as most real data possess an underlying structure imposed by natural geochemical processes.

The first step in the factor analysis procedure is to transform each column of the data matrix (table 8) to range from zero to one so that each value in the column is expressed as a proportion of the column range. The purpose of the transformation is to put each constituent on an equal basis, with the result that each constituent is given approximately equal weight in determining the factor analysis outcome. The transformed data matrix is given in table 9. The transformed data matrix is then normalized by dividing each row through by the square root of the row sum of squares. The normalized data matrix is given in table 10.

The transformed data matrix in table 10 is then post-multiplied by its transpose to form a 20 by 20 matrix of

TABLE 9.—*Hypothetical data for 6 constituents in 20 samples transformed to proportions of the ranges for each variable*

Sample No.	Constituent					
	A	B	C	D	E	F
1	0.911	0.200	0.222	0.154	0.200	0
2	.800	.286	.222	.231	.200	0.200
3	.978	.343	0	0	0	.200
4	.689	.257	.333	.385	.400	.200
5	.867	.439	0	.077	.200	.200
6	.822	.171	.333	.308	.200	0
7	.578	.343	.333	.462	.400	.400
8	.733	.143	.444	.385	.200	.200
9	.889	.314	.111	.077	.200	.200
10	1.000	.229	.111	.077	0	0
11	.467	.543	.222	.462	.600	.400
12	.578	.571	.111	.308	.400	.400
13	.356	.629	.222	.538	.600	.600
14	.444	.657	.111	.385	.600	.600
15	0	1.000	.111	.692	1.000	1.000
16	.111	.914	.111	.615	.800	1.000
17	.267	0	1.000	1.000	.600	.400
18	.356	.143	.778	.769	.600	.400
19	.267	.114	.889	.923	.600	.400
20	.378	.029	.889	.846	.400	.400

TABLE 10.—*Hypothetical data for 6 constituents in 20 samples transformed and normalized*

Sample No.	Constituent					
	A	B	C	D	E	F
1	0.919	0.202	0.224	0.155	0.202	0
2	.841	.300	.234	.243	.210	0.210
3	.926	.325	0	0	0	.190
4	.689	.257	.333	.385	.400	.200
5	.858	.424	0	.076	.198	.198
6	.843	.176	.342	.315	.205	0
7	.552	.328	.318	.441	.382	.382
8	.739	.144	.448	.388	.202	.202
9	.895	.316	.111	.077	.201	.201
10	.967	.221	.107	.074	0	0
11	.410	.477	.195	.406	.527	.352
12	.554	.548	.107	.295	.384	.384
13	.283	.500	.177	.428	.477	.477
14	.361	.535	.090	.313	.488	.488
15	0	.535	.059	.370	.535	.535
16	.065	.539	.065	.363	.471	.589
17	.166	0	.621	.621	.373	.248
18	.260	.105	.570	.563	.440	.293
19	.178	.076	.593	.616	.400	.267
20	.269	.020	.633	.603	.285	.285

coefficients of proportional similarity among the samples.[3] The coefficients range from zero to plus one and measure the compositional similarity for each sample pair. Two samples that are similar in composition have a coefficient near one; two samples almost totally dissimilar have a coefficient near zero. The coefficient of proportional similarity was designed by Imbrie and Purdy (1962), and the matrix of these coefficients (referred to as a cosine theta matrix) is the basis for the factor analysis. The cosine theta matrix is given in table 11; it shows, for example, that samples 1 and 2 are very similar (cosine theta=0.97) and samples 10 and 15 are almost totally dissimilar (cosine theta=0.15). The first 6 of the 20 eigenvalues of the cosine theta matrix are given in table 12; the fact that the fourth value is considerably smaller than the third indicates that the normalized data matrix in table 10

[3]Where the number of samples is large, it is desirable to avoid actual computation of the cosine theta matrix by use of the method of Klovan and Imbrie (1971).

may be resolved into three factors. That is, most of the variability in the normalized data may be described in terms of three factors rather than six variables. The value of CPN associated with the third eigenvalue indicates that 99.5 percent of the variability (expressed as sum of squares) in the normalized data can be accounted for by a three-factor model. In practical terms, this implies that only about 0.5 percent of the variability in the normalized data will be lost by reduction of the six variables to only three.

The degree to which each sample will conform to the factor model is indicated by the derived sample communality. The original communality of each sample is one, as is shown by the fact that the sum of squares of each row in the normalized data matrix (table 10) is unity. When the factor model is used to reproduce the normalized data matrix the row sums of squares are found to be something less than unity. That is, the derived communalities are less than unity and where greatly less than unity they indicate that the corresponding sample does not conform well with the factor model. Where most of the derived communalities are near one, but a few of them are small, the small values may indicate that the corresponding samples are anomalous with respect to the others. The value of CPN associated with the third eigenvalue in table 12 indicates that the average derived communality for the 20 samples in our hypothetical data, using a three-factor model, will be 0.995.

The degree to which Q-mode models containing up to six factors will account for the variance in each constituent represented in the original data (table 8), rather than the normalized data (table 10), is given by the factor-variance diagram of figure 6. The diagram was constructed by computing the coefficients of determination, r_j^2, between the original data and the data reproduced for each variable, j, by models containing two to six factors (Miesch, 1975, 1976). All values of r_j^2 are zero for the trivial case where the model contains only one factor, inasmuch as one-factor models lead to reproduced data with no intracolumn variance. The factor-variance diagram shows that the data of table 8, like that of table 10, may be reproduced in large part by Q-mode models containing three factors.

Each of the three factors (new theoretical variables) is a combination of the original six variables, and is represented by a set of factor composition scores. There is an unlimited number of sets of factors and associated scores from which to choose, but for the purpose of merely summarizing the data it is appropriate to use the Varimax axes. The scores for the Varimax axes have the property of being totally dissimilar by the measure of cosine theta referred to previously. The Varimax composition scores are given in table 13 and the composition loadings of the 20 samples on these scores are given in table 14. When the scores are combined in the proportions indicated by the

TABLE 11.—*Matrix of coefficients of proportional similarity (cosine theta) for 20 hypothetical samples*

[The complete matrix is symmetrical with unities in the diagonal; only the upper triangle is shown]
Sample No.

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.97	0.92	0.90	0.93	0.98	0.79	0.91	0.96	0.97	0.69	0.77	0.56	0.61	0.29	0.33	0.46	0.56	0.49	0.54
2		.92	.95	.95	.96	.91	.95	.98	.92	.82	.89	.73	.77	.49	.54	.57	.67	.60	.65
3			.76	.97	.84	.69	.77	.97	.97	.60	.76	.51	.60	.28	.35	.20	.33	.24	.31
4				.85	.94	.97	.97	.89	.79	.91	.90	.83	.83	.62	.65	.76	.85	.79	.80
5					.86	.80	.80	.99	.93	.76	.88	.68	.75	.47	.52	.31	.46	.36	.40
6						.85	.97	.91	.91	.73	.77	.62	.63	.34	.38	.62	.70	.64	.70
7							.92	.82	.67	.96	.94	.93	.91	.77	.80	.80	.89	.84	.84
8								.87	.82	.79	.81	.72	.70	.46	.51	.77	.83	.78	.83
9									.95	.75	.86	.66	.72	.42	.48	.39	.52	.43	.48
10										.55	.69	.43	.50	.15	.22	.27	.38	.30	.38
11											.97	.98	.98	.89	.90	.73	.83	.78	.74
12												.94	.97	.82	.85	.58	.71	.64	.62
13													.99	.95	.96	.72	.82	.78	.73
14														.93	.95	.61	.74	.68	.63
15															.99	.60	.69	.66	.58
16																.60	.70	.66	.59
17																	.98	1.00	.99
18																		.99	.98
19																			.99

Sample No.

TABLE 12.—*First six eigenvalues of the matrix of cosine theta, and cumulative proportions of N (CPN)*

Order of eigenvalues	Eigenvalues	CPN
1	14.951	0.748
2	3.277	.911
3	1.673	.995
4	.080	.999
5	.017	1.000
6	.002	1.000

TABLE 13.—*Composition scores for the three-factor Varimax model*

Constituent	Factor		
	1	2	3
A	29.97	90.33	48.76
B	60.77	9.12	-17.84
C	-12.74	.61	45.91
D	7.92	.24	19.42
E	6.46	-.11	2.90
F	7.62	-.19	.86
Sum	100.00	100.00	100.01

TABLE 14.—*Composition loadings for the three-factor Varimax model*

Sample No.	Factor			Sample No.	Factor		
	1	2	3		1	2	3
1	0.05	0.82	0.13	11	.31	.47	.22
2	.12	.73	.15	12	.28	.59	.13
3	.09	.91	.00	13	.38	.38	.24
4	.15	.61	.24	14	.38	.46	.16
5	.15	.82	.03	15	.64	.11	.25
6	.05	.74	.21	16	.59	.19	.22
7	.22	.53	.25	17	.18	.18	.64
8	.08	.65	.27	18	.21	.29	.50
9	.11	.81	.08	19	.21	.21	.58
10	.02	.92	.06	20	.15	.29	.56

loadings for any sample, the result is an approximation of the composition of the sample. For example, the composition loadings for sample 1 are 0.05, 0.82, and 0.13

(table 14). When the three sets of composition scores of table 13 are combined, respectively, in these proportions, the result is an approximation of the composition of sample 1 as given in table 8. The combination of scores referred to an the composition of hypothetical sample 1 are given for comparison in table 15.

Inspection of the factor scores in table 13 reveals that individual samples with high composition loadings (table 14) on factor 1 tend to be relatively rich in constituents *B*, *E*, and *F* and deficient in constituent *C*; samples with high loadings on factor 2 tend to be rich in constituent *A* and contain little else; samples with high loadings on factor 3 tend to be rich in constituents *C* and *D* and are especially deficient in constituent *B*. Hence, maps of the three
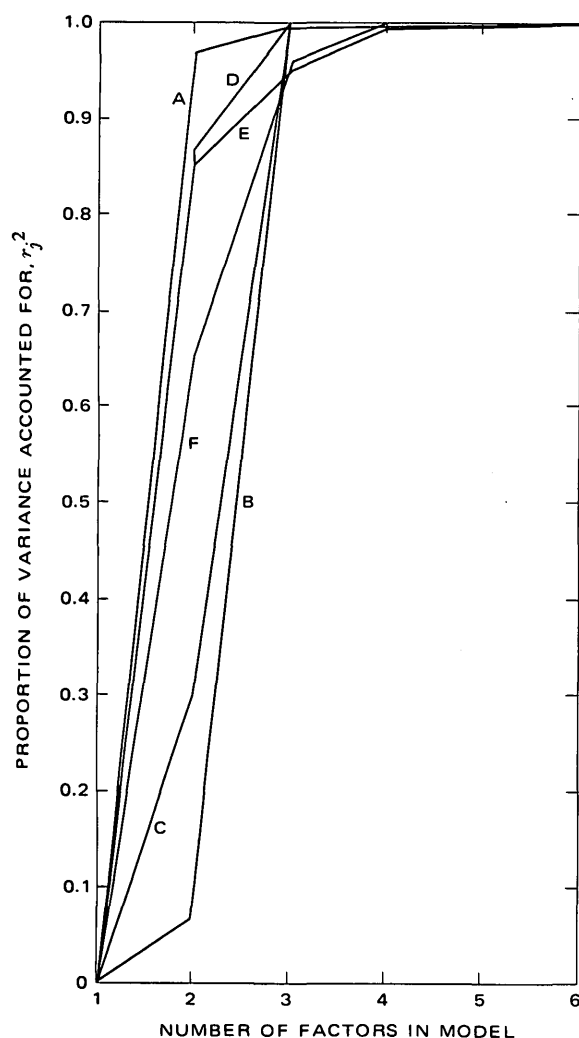
FIGURE 6.—Factor-variance diagram for the hypothetical data in table 8. The diagram shows the proportion of the total variance in each constituent that can be explained by Q-mode models containing one to six factors (end members).

TABLE 15.—*Original hypothetical data for sample 1 (from table 8) and a combination of the three sets of composition scores (table 13) in the proportions indicated by the composition loadings (0.05, 0.82, and 0.13; table 14)*

Constituent	Original data	Combination of scores
A	82	81.9
B	8	8.2
C	6	5.8
D	3	3.1
E	1	.6
F	0	.3
Sum	100	99.9

columns of factor loadings in table 14 tend to show the same compositional variations as do maps of the six individual constituents represented in table 8. This is illustrated by comparison of figures 7 and 8. Figure 7 shows six maps of the individual constituents, and figure 8

shows maps of the three columns of composition loadings. The patterns of variation for the six constituents (fig. 7) display several prominent features that are repeated, to some extent, among the six maps. One of these features is a preponderance of high values in the upper-left portion of the map (constituent A); another is a cluster of high values in the lower right (constituents C and D); another is a cluster of high values in the lower-central part of the maps (constituents B, E, and F). The repetition of these features among the maps occurs because of linear correlations among the six constituents, and it results in partial redundancy. Each of the features also occurs in one of the factor maps of figure 8. Thus, the factor maps tend to show the same patterns of geochemical variations, but redundancy is avoided. The map of factor 1, for example, shows that samples from the lower/central part of the map area tend to be rich in constituents B, E, and F and deficient in constituent C. The other two factor maps can be interpreted in a similar manner.

Although the principal purpose of Q-mode factor analysis, as used in the geochemical survey of Missouri, has been to summarize data and to show patterns of variation in general geochemical character across the State, the results can, in some cases, be interpreted in terms of geochemical processes. In the hypothetical problem treated here, for example, one may infer that factor 1 represents some process that acts to enrich the samples in constituents B, E, and F and removes constituent C. The samples that have been subjected to this process to the greatest extent are those with high loadings on factor 1 (table 14), and the degrees to which they have been subjected to it are proportional to the magnitudes of the loadings. Whether or not interpretations such as this are plausible in actual field studies, however, does not bear on the usefulness of the factor maps for summarizing the data in a convenient map form.

In conventional factor analysis the factor loadings and scores are in normalized form and the signs of these values are somewhat arbitrary. That is, the signs of the entire kth column of loadings may be changed if one also changes the signs of the entire kth row of scores. With the extended method, however, it is required that each set of scores (that is, the scores for each factor) sum to the same constant as do the variables for each sample. This prohibits changing signs of the scores and, therefore, prohibits changing signs of the loadings also (Miesch, 1975). As a result, in the interpretation of loadings in terms of geochemical processes a negative loading must imply the depletion of constituents with positive scores and enrichment of constituents with negative scores.

Most of the data matrices examined by Q-mode factor analysis in the Missouri study contain variables representing all of the major constituents in the materials (rocks, soils, and water) except oxygen and H_2O. Because
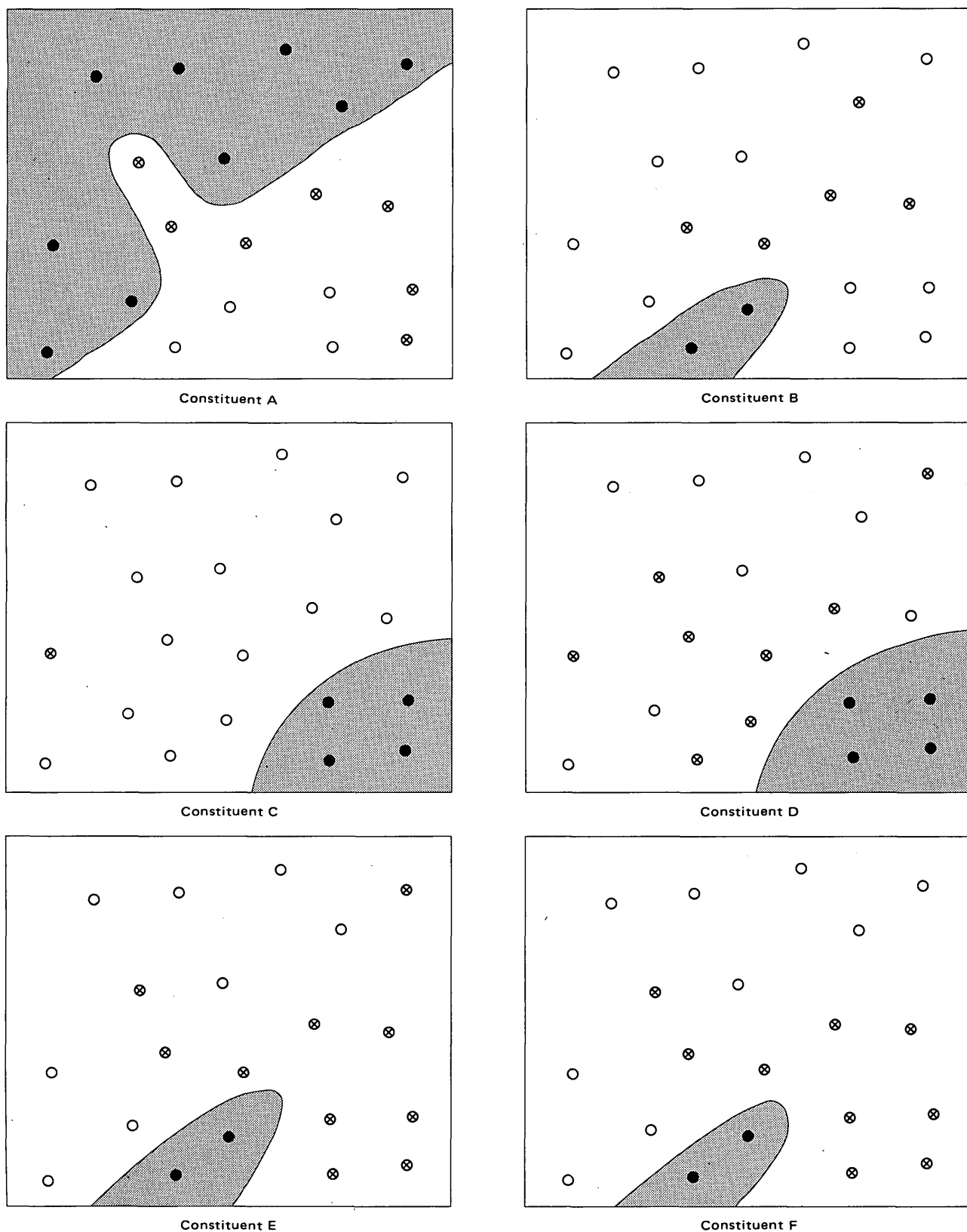
GEOCHEMICAL SURVEY OF MISSOURI

FIGURE 7.—Distribution patterns for 6 constituents in 20 samples used in the Q-mode experiment (see hypothetical data in table 8). A solid dot indicates that the value is in the upper third of the range for the constituent, a circled × indicates that the value is in the middle third, and an open circle indicates it is in the lower third of the range. The regions of high values are shaded.
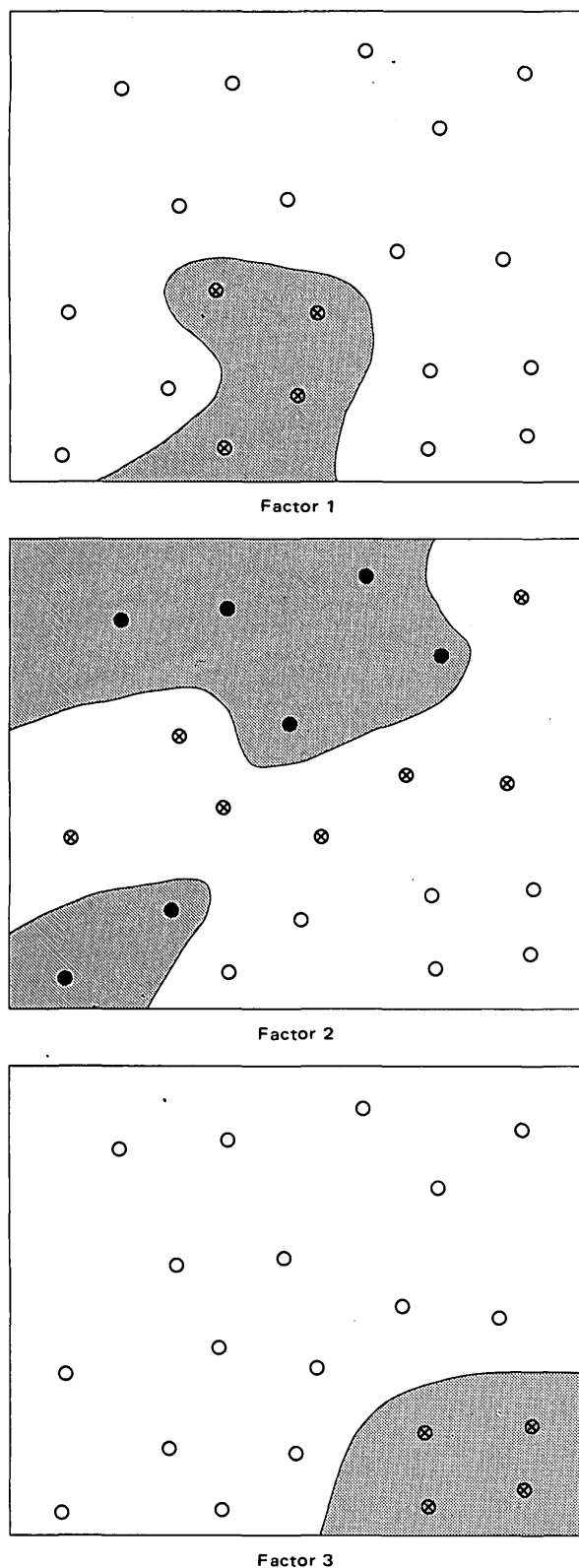
Factor 1

Factor 2

Factor 3

FIGURE 8.—Distribution patterns for the composition loadings from the Q-mode experiment (table 14) on three sets of composition scores (table 13). Solid dot indicates a score from 0.67 to 1.0, circled × indicates a score from 0.33 to 0.67, and open circle indicates a score from 0 to 0.33. The regions of high values are shaded.

these two constituents are not represented, the variables for each sample do not sum to a constant as is required for use of the extended Q-mode method, and some adjustment was necessary. For constituents in rocks and soils reported by the analysts in elemental form, the reported values were converted to oxide forms if the element occurs in oxygen-bearing minerals. Thus, reported values for Ba were converted to BaO and values of Cr were converted to Cr_2O_3, for example. The particular oxides used depended on the common valence states for the element in the surficial environment. Reports for elements such as organic carbon and Hg, which occur in elemental form in the rocks and soils, were not converted to the oxide form. All data on rocks and soils reported as parts per million were then converted to percentages, and the final adjustment consisted of dividing the variables for each sample through by their sum and multiplying through by 100.

The chemical data for each sample of water were adjusted to sum to 10^6 by converting all values to parts per million and summing. The sum was then subtracted from 10^6, and the difference was taken as an estimate of H_2O in parts per million. Although the values for H_2O are overwhelmingly greater than the values for any other constituent in the samples, they all ranged from zero to one, just as for the other constituents, after the data were transformed to proportions of the variable ranges. The procedure led to parts per million data with a constant sum for each sample as is required for the extended Q-mode method.

REFERENCES CITED

American Society for Testing and Materials, 1968, Methods for emission spectrochemical analysis [5th ed.]: Am. Soc. Testing and Materials, 926 p.

Anderson, R. L., and Bancroft, T. A., 1952, Statistical theory in research: New York, McGraw-Hill Book Co., 399 p.

Baird, A. K., McIntyre, D. S., Welday, E. E., and Morton, D. M., 1967, A test of chemical variability and field sampling methods, Lakeview Mountain tonalite, Lakeview Mountains, southern California batholith, in Short contributions to California geology: California Div. Mines and Geology Spec. Rept. 92, p. 11-19.

Barker, F. B., and Robinson, B. P., 1963, Determination of beta activity in water: U.S. Geol. Survey Water-Supply Paper 1696-A, 32 p.

Barnes, Ivan, 1964, Field measurement of alkalinity and pH: U.S. Geol. Survey Water Supply Paper 1535-H, 17 p.

Barnett, P. R., 1967, Device for measuring spectral linewidths: Appl. Spectroscopy, v. 21, no. 2, p. 129-130.

Barnett, P. R., and Mallory, E. C., Jr., 1971, Determination of minor elements in water by emission spectroscopy: U.S. Geol. Survey Techniques Water-Resources Inv. TWI 5-A2, 31 p.

Bennett, J. H., and Manuel, O. K., 1968, On iodine abundances in deep-sea sediments: Jour. Geophys. Research, v. 73, no. 6, p. 2302-2303.

Bertin, E. P., 1970, Principles and practice of X-ray spectrometric analysis: New York and London, Plenum Press, 679 p.

Brown, Eugene, Skougstad, M. W., and Fishman, M. J., 1970, Methods for collection and analysis of water samples for dissolved minerals and gases: U.S. Geol. Survey Techniques Water-Resources Inv. TWI 5-A1, 160 p.

Case, A. A., Selby, L. A., Hutcheson, D. T., Ebens, R. J., Erdman, J. A., and Feder, G. L., 1972, Metabolic disorders in young beef cattle associated with abnormalities in their geochemical environment [abs.], in Trace substances in environmental health: Missouri Univ. Ann. Conf., 6th.

———1973, Infertility and growth suppression in beef cattle associated with abnormalities in their geochemical environment, in Hemphill, D. D., ed., Trace substances in environmental health: Missouri Univ. Ann. Conf., 6th, Proc., p. 15-21.

Chayes, Felix, 1960, On correlation between variables of constant sum: Jour. Geophys. Research, v. 65, no. 12, p. 4185-4193.

Chayes, Felix, and Kruskal, William, 1966, An approximate statistical test for correlations between proportions: Jour. Geology, v. 74, no. 5, pt. 2, p. 692-702.

Clark, R. S., Rowe, M. W., Ganapathy, R., and Kuroda, P. K., 1967, Iodine, uranium and tellurium contents in meteorites, in Meteorites and tektites: Geochim. et Cosmochim. Acta, v. 31, no. 10, p. 1605-1613.

Clarke, F. W., 1924, The composition of the river and lake waters of the United States: U.S. Geol. Survey Prof. Paper 135, 199 p.

Cochran, W. G., 1963, Sampling techniques [2d ed.]: New York, John Wiley and Sons, Inc., 413 p.

Cohen, A. C., Jr., 1959, Simplified estimators for the normal distribution when samples are singly censored or truncated: Technometrics, v. 1, no. 3, p. 217-237.

———1961, Tables for maximum likelihood estimates; singly truncated and singly censored samples: Technometrics, v. 3, no. 4, p. 535-541.

Connor, J. J., Erdman, J. A., Sims, J. D., and Ebens, R. J., 1971, Roadside effects on trace element content of some rocks, soils, and plants of Missouri, in Hemphill, D. D., ed., Trace substances in environmental health: Missouri Univ. Ann. Conf., 4th, Proc. p. 26-34.

Connor, J. J., Feder, G. L., Erdman, J. A., and Tidball, R. R., 1972, Environmental geochemistry in Missouri—A multidisciplinary study, in Earth sciences and the quality of life—Symposium 1: Internat. Geol. Cong., 24th, Proc., p. 7-14.

Connor, J. J., Shacklette, H. T., and Erdman, J. A., 1971, Extraordinary trace-element accumulations in roadside cedars near Centerville, Missouri, in Geological Survey research 1971: U.S. Geol. Survey Prof. Paper 750-B, p. B151-B156.

Cuthbert, Margaret, and Ward, F. N., 1964, Determination of iodine in vegetation: U.S. Geol. Survey Prof. Paper 501-C, p. C154-C156.

Davies, O. L., ed., 1954, The design and analysis of industrial experiments: Edinburgh, Oliver and Boyd, Ltd., and New York, Hafner Pub. Co., 636 p.

Duncan, D. B., 1955, Multiple range and multiple F tests: Biometrics, v. 11, no. 1, p. 1-42.

Ebens, R. J., Erdman, J. A., and Feder, G. L., 1972, Geochemical anomalies in a claypit area, Callaway County, Missouri, where metabolic disorders in cattle were reported: Geol. Soc. America Abs. with Programs, v. 4, no. 5, p. 319.

Ebens, R. J., Erdman, J. A., Feder, G. L., Case, A. A., and Selby, L. A., 1973, Geochemical anomalies of a claypit area, Callaway County, Missouri, and related metabolic imbalance in beef cattle: U.S. Geol. Survey Prof. Paper 807, 24 p.

Feder, G. L., 1972, Problems of sampling in trace element investigations, in Geochemical environment in relation to health and disease: New York Acad. Sci. Annals, v. 199, p. 118-123.

Feder, G. L., Ebens, R. J., and Connor, J. J., 1972, The relationship between lithology and trace element content of ground water: Am. Water Resources Assoc. Ann. Water Resources Conf., 9th, St. Louis, Mo., Short Papers, p. 104.

Fenneman, N. M., 1938, Physiography of eastern United States: New York, McGraw-Hill Book Co., 714 p.

Garrett, R. G., 1969, The determination of sampling and analytical errors in exploration geochemistry: Econ. Geology, v. 64, no. 5, p. 568-569.

———1973, The determination of sampling and analytical errors in exploration geochemistry—a reply: Econ. Geology, v. 68, no. 2, p. 282-283.

Goles, G. G., and Anders, E., 1962, Abundances of iodine, tellurium and uranium in meteorites: Geochim. et Cosmochim. Acta, v. 26, p. 723-737.

Goles, G. G., Greenland, L. P., and Jerome, D. Y., 1967, Abundances of chlorine, bromine and iodine in meteorites, in Meteorites and tektites: Geochim. et Cosmochim. Acta, v. 31, p. 1771-1787.

Hayes, W. C., and Guild, P. W., 1967, Iron, in Mineral and water resources of Missouri: U.S. Cong., 90th, 1st sess., Senate Doc. 19 (U.S. Geol. Survey and Missouri Div. Geol. Survey and Water Resources [Rept.]), p. 74-88.

Hemphill, D. D., ed., 1967, Trace substances in environmental health: Missouri Univ. Ann. Conf., 1st Proc., 236 p.

———1969, Trace substances in environmental health: Missouri Univ. Ann. Conf., 2d, Proc., 318 p.

———1970, Trace substances in environmental health: Missouri Univ. Ann. Conf., 3d, Proc., 319 p.

———1971, Trace substances, in environmental health: Missouri Univ. Ann. Conf., 4th, Proc., 456 p.

———1972, Trace substances in environmental health: Missouri Univ. Ann. Conf., 5th, Proc., 559 p.

———1973, Trace substances in environmental health: Missouri Univ. Ann. Conf., 6th, Proc., 399 p.

Hopps, H. C., 1971, Geographic pathology and the medical implications of environmental geochemistry, in Cannon, H. L., and Hopps, H. C., eds., Environmental geochemistry in health and disease: Geol. Soc. America Mem. 123, p. 1-11.

Imbrie, John, 1963, Factor and vector analysis programs for analyzing geologic data: U.S. Office Naval Research, Geog. Branch Tech. Rept. 6, ONR Task No. 389-135, 83 p.

Imbrie, John, and Purdy, E. G., 1962, Classification of modern Bahamian carbonate sediments, in Classification of carbonate rocks—A symposium: Am. Assoc. Petroleum Geologists Mem. 1, p. 253-272.

Ingram, B. L., 1970, Determination of fluoride in silicate in silicate rocks without separation of aluminum using a specific ion electrode: Anal. Chemistry, v. 42, no. 14, p. 1825-1827.

Jenkins, R., and deVries, J. L., 1967, Practical X-ray spectrometry: New York, Springer-Verlag, 182 p.

Kiilsgaard, T. H., 1967, Lead and zinc, in Mineral and water resources of Missouri: U.S. Cong., 90th, 1st sess., Senate Doc. 19 (U.S. Geol. Survey and Missouri Div. Geol. Survey and Water Resources [Rept.]), p. 41-63.

Klovan, J. E., and Imbrie, John, 1971, An algorithm and FORTRAN-IV program for large-scale Q-mode factor analysis and calculation of factor scores: Jour. Internat. Assoc. Mathematical Geology, v. 3, no. 1, p. 61-77.

Klovan, J. E., and Miesch, A. T., 1975, Extended CABFAC and QMODEL computer programs for Q-mode factor analysis of compositional data: Computers and Geosciences, V. 1 (in press).

Krumbein, W. C., and Slack, H. A., 1956, Statistical analysis of low-level radioactivity of Pennsylvanian black fissile shale in Illinois: Geol. Soc. America Bull., v. 67, no. 6, p. 739-762.

Küchler, A. W., 1964, Potential natural vegetation of the conterminous United States: Am. Geog. Soc. Spec. Pub. 36, 116 p.

——— 1970, Potential natural vegetation, in The national atlas of the United States of America: U.S. Geol. Survey, p. 90-91.

Lag, J., and Steinnes, E., 1971, Study of mercury and iodine in Norwegian forest soils by neutron activation analysis, in The use of nuclear techniques in the measurement and control of environ-

mental pollution—Symposium, Salzburg, 1970: Vienna, Internat. Atomic Energy Agency, p. 429-438.

Manson, Vincent, and Imbrie, John, 1964, FORTRAN. program for factor and vector analysis of geologic data using an IBM 7090 or 7094/1401 computer system: Kansas Geol. Survey Spec. Distrib. Pub. 13, 46 p.

Marienfeld, C. J., 1972, A field study relating geochemical environment to health and disease, in Hopps, H. C., and Cannon, H. L., eds., Geochemical environment in relation to health and disease: New York Acad. Sci. Annals, v. 199, p. 335-348.

Michie, U. McL., 1973, The determination of sampling and analytical errors in exploration geochemistry [discussion]: Econ. Geology, v. 68, no. 2, p. 281-282.

Miesch, A. T., 1967a, Theory of error in geochemical data: U.S. Geol. Survey Prof. Paper 574-A, 17 p.

———1967b, Methods of computation for estimating geochemical abundance: U.S. Geol. Survey Prof. Paper 574-B, 15 p.

———1969, The constant sum problem in geochemistry, in Merriam, D. F., ed., Computer applications in the earth sciences: New York, Plenum Press, p. 161-176.

———1972, Sampling problems in trace element investigations of rocks: Annals of the New York Acad. Sciences, v. 199, p. 95-104.

———1976, Q-mode factor analysis of geochemical and petrologic data matrices with constant row-sums: U.S. Geol. Survey Prof. Paper 574-G (in press).

———1975, Q-mode factor analysis of compositional data: Computers and Geosciences, v. 1 (in press).

Miesch, A. T., Chao, E. C. T., and Cuttitta, Frank, 1966, Multivariate analysis of geochemical data on tektites: Jour. Geology, v. 74, no. 5, pt. 2, p. 673-691.

Missouri Geological Survey and Water Resources, 1969, Generalized geologic map of Missouri: 1:500,000, 1 sheet.

Munns, R. K., and Holland, D. C., 1971, Determination of mercury in fish by flameless atomic absorption—A collaborative study: Assoc. Official Anal. Chemists Jour., v. 54, no. 1, p. 202-205.

Myers, A. T., Havens, R. G., and Dunton, P. J., 1961, A spectrochemical method for the semiquantitative analysis of rocks, minerals, and ores: U.S. Geol. Survey Bull. 1084-I, p. 207-229.

Nakagawa, H. M., and Harms, T. F., 1968, Atomic absorption determination of cadmium in geologic materials, in Geological Survey research 1968: U.S. Geol. Survey Prof. Paper 600-D, p. D207-D209.

Pollaro, F. Z., Levins, N., Killelea, J. R., Musca, R. C., and Hassialis, M. D., 1958, Recovery of uranium from Chattanooga shale: Industrial and Engineering Chemistry, v. 50, no. 12, p. 1750-1753.

Rader, L. F., and Grimaldi, F. S., 1961, Chemical analyses for selected minor elements in Pierre Shale: U.S. Geol. Survey Prof. Paper 391-A, p. A1-A45.

Reed, G. W., Jr., and Allen, R. O., Jr., 1966, Halogens in chrondites: Geochim. et Cosmochim. Acta, v. 30, no. 8, p. 779-800.

Reed, G. W., Jr., and Jovanovic, S., 1970, Halogens, mercury, lithium and osmium in Apollo 11 samples, in Apollo 11 Lunar Sci. Conf., Houston, Tex., 1970, Proc., v. 2, Chemical and isotope analyses: New York, Pergamon Press (Geochim. et Cosmochim. Acta, Supp. 1), p. 1487-1492.

Sichel, H. S., 1952, New methods in the statistical evaluation of mine sampling data: Inst. Mining and Metallurgy Trans., v. 61, p. 261-288.

———1966, The estimation of means and associated confidence limits for small samples from lognormal populations: So. African Inst. Mining and Metall. Jour., preprint no. 4, 17 p.

Slavin, Walter, 1968, Atomic absorption spectroscopy: New York, Interscience Publishers, 307 p.

Soil Conservation Service, 1970, Distribution of principal kinds of soils—Orders, Suborders and Great Groups, in The national atlas of the United States of America: U.S. Geol. Survey, p. 86-87.

Taves, D. R., 1968, Separation of fluoride by rapid diffusion using hexamethyldisiloxane: Talanta, v. 15, p. 969-974.

Tidball, R. R., 1971, Geochemical variation in Missouri soils, in Hemphill, D. D., ed., Trace substances in environmental health: Missouri Univ. Ann. Conf., 4th, Proc. p. 15-25.

Tidball, R. R., and Sauer, H. I., 1972, Soil composition and human death rates in Missouri: Geol. Soc. America Abs. with Programs, v. 4, no. 7, p. 690-691.

U.S. Geological Survey, 1972a-f, 1973, Environmental geochemistry—Geochemical survey of Missouri: U.S. Geol. Survey open-file reports [plans and progress reports for six-month periods as follows]

    1972a, First six-month period (July-Dec., 1969), 49 p.
      b, Second six-month period (Jan.-June, 1970), 60 p.
      c, Third six-month period (July-Dec., 1970), 33 p.
      d, Fourth six-month period (Jan.-June, 1971), 63 p.
      e, Fifth six-month period (July-Dec., 1971), 145 p.
      f, Sixth six-month period (Jan.-June, 1972), 86 p.
    1973, Seventh six-month period (July-Dec., 1972), 59 p.

Vaughn, W. W., 1967, A simple mercury vapor detector for geochemical prospecting: U.S. Geol. Survey Circ. 540, 8 p.

Vineyard, J. D., 1967, Geology—Physiography, in Mineral and water resources of Missouri: U.S. Cong., 90th, 1st sess., Senate Doc. 19 (U.S. Geol. Survey and Missouri Div. Geol. Survey and Water Resources [Rept.]), p. 13-15.

Weigel, W. W., 1965, The inside story of Missouri's exploration boom— Pts. 1 and 2: Eng. and Mining Jour., v. 166, no. 11, p. 77-86, 170, 172; no. 12, p. 81-83, p. 161-162.