

55(481)
N/285

NGU



Norges geologiske
undersøkelse

Nr. 285

Skrifter 3

Bjørn Bølviken: Statistisk beskrivelse
av geokjemiske data

STATENS TEKNISKE INSTITUTT
BIBLIOTEK

Universitetsforlaget 1973
Trondheim · Oslo · Bergen · Tromsø



NGU

Norges geologiske undersøkelse

Geological Survey of Norway

Norges geologiske undersøkelse, Leiv Eirikssons vei 39, Trondheim. Telefon (075) 20166.
Postadresse: Postboks 3006, 7001 Trondheim.

Administrerende direktør: *Karl Ingvaldsen*

Geologisk avdeling: Direktør dr. philos. *Peter Pädget*

Geofysisk avdeling: Direktør *Inge Aalstad*

Kjemisk avdeling: Direktør *Aslak Kvalheim*

Publikasjoner fra *Norges geologiske undersøkelse* utgis som bind med fortløpende hovednummerering, og deles inn i to serier, *Bulletins* og *Skrifter*.

Bulletins omfatter vitenskapelige arbeider over regionale, generelle eller spesialiserte emner av faglig interesse.

Skrifter omfatter beskrivende artikler og rapporter over regionale, tekniske, økonomiske, naturfaglige og andre geologiske emner av spesialisert eller allmen interesse. Skrifter utgis på norsk, med resymé på engelsk (Abstract).

REDAKTØR

Statsgeolog *Tore Torske*, Norges geologiske undersøkelse, postboks 3006, 7001, Trondheim.

UTGIVER

Universitetsforlaget, Postboks 307, Blindern, Oslo 3

MANUSKRIPTER — SÆRTRYKK

Alle henvendelser vedrørende godttatte manuskripter, bestilling av særtrykk, osv. sendes til Universitetsforlaget, Kontoret: Trondheim, Munkegt. 8, 7000 Trondheim.

FORRETNINGSADRESSE

Henvendelser angående abonnement sendes til Universitetsforlaget, Postboks 307, Blindern, Oslo 3.

Statistisk beskrivelse av geokjemiske data

BJØRN BØLVIKEN

Bølviken, B. 1973: Statistical description of geochemical data. *Norges geol. Unders.* 285, 1—10.

A manual procedure for the organization of geochemical observations into frequency distributions is described. When frequency distributions of geochemical exploration data are plotted on logarithmic probability paper, more or less straight lines are often obtained, because trace element concentrations are in many cases approximately lognormally distributed. Values for selected parameters for distribution mean and deviation can be estimated very simply from such plots. If the distribution plot is nonlinear, a mixture of two or more distributions may sometimes be indicated. If the mixture is bimodal, the empirical frequency distribution can be divided graphically into two theoretical distributions: two backgrounds, or background and anomaly. Often the background can simply be represented by a straight line through the lower and middle part of the distribution plot. When grouping the metal contents for map presentation, the geometric deviation of the background can conveniently be used as a unit for concentration intervals. In some cases a universal grouping is preferable. A convenient grouping can then be obtained by dividing each tenth power of concentration into equal intervals on a logarithmic scale, taking into account the analytical sensitivity when setting the lowest interval limit, the total data variability, and the sampling and analytical error when deciding into how many intervals each tenth power should be divided.

A geochemical map should always be accompanied by a frequency distribution plot from which the significance of any high concentration should be judged. The ordinary use of threshold as a fixed border between background and anomaly is not consistent with concept of background and anomaly being statistical distributions, the observation of which may take any value between zero and a common theoretical maximum. According to this concept, a geochemical anomaly does not only imply high concentrations, but it also includes peculiar geochemical patterns at any concentration level.

B. Bølviken, *Norges geologiske undersøkelse*, Box 3006, N-7001 Trondheim, Norway

Det er neppe mulig å tolke tallmaterier som består av et stort antall enkelt-observasjoner ved bare å lese gjennom en liste over måleresultatene. For å kunne ekstrahere maksimal informasjon må dataene organiseres og grupperes på en oversiktlig måte, og beskrives ved hjelp av egnede parametre. Dette gjelder ikke minst for analysedata fra geokjemisk malmleting eller andre praktisk-geokjemiske undersøkelser, fordi tallmaterialet i slike undersøkelser ofte kan være meget stort — ikke sjelden dreier det seg om tusener av analyseresultater. I tillegg vil det normalt være ønskelig å fremstille resultatene på kart, noe som hittil vanligvis har forutsatt en viss gruppering av analysetallene, dersom kartene skal kunne gi brukeren et noenlunde lett tolkbart og visuelt inntrykk av de geokjemiske fordelingsmønstre.

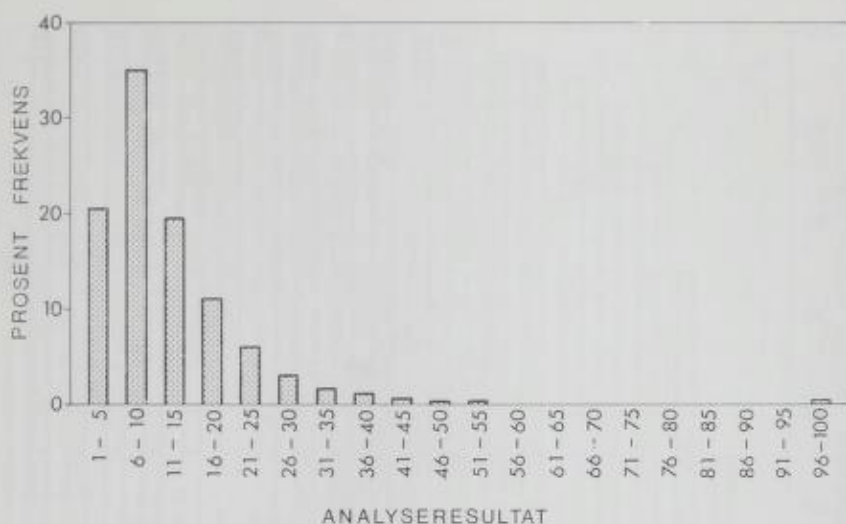


Fig. 2. Histogram av data tatt fra fig. 1.
Histogram of data taken from Fig. 1.

sentrasjonen avsettes langs abscissen og frekvensen langs ordinaten. Søylen plasseres midt på hver konsentrasjonsgruppe (fig. 2).

Hvis man suksessivt summerer alle frekvenser og avsetter disse summene langs ordinaten, fåes en kumulativ frekvensfordeling (fig. 3). Kumuleres *nedenfra*, bør hvert punkt plottes ved *ovre* grense av hver gruppe.

En kumulativ frekvensfordeling lar seg hensiktsmessig representere på sannsynlighetspapir (beskrevet av f.eks. Koch & Link (1970)). Bruk av slikt papir i geokjemisk prospektering er introdusert av Tennant & White (1959) og senere diskutert av Williams (1967), Lepeltier (1969), Bølviken (1971) og Duval et al. (1972). Sannsynlighetspapir har lineær eller logaritmisk skala langs abscissen, og ordinaten er inndelt etter Gauss intergralet for normalfordeling. Når en kumulert normalfordeling fremstilles på sannsynlighetspapir med lineær abscisse, fremkommer en rett linje.

En *lognormal* fordeling er en fordeling hvor logaritmen til den observerte verdi er normalfordelt. Folgelig vil en kumulert lognormal fordeling beskrive en rett linje når den fremstilles på sannsynlighetspapir med logaritmisk abscisse. (Se figur 4. Kurve A er tegnet etter data i figur 1. Legg merke til at ved så lave prøvetall som 250 blir beliggenheten av de ekstreme deler av kurven meget usikker).

Det følgende gjelder for lognormale fordelinger. Ved alle beregninger er brukt logaritmer med grunntall 10. Når fordelingen er tegnet på sannsynlighetspapir, kan uttrykk for middelvei og spredning finnes ved grafisk løsning (fig. 4, kurve A og tabell 1).

Median (M) er den abscisseverdi som tilsvarer ordinaten 50 % (50 prosentilen). For lognormale fordelinger vil dette punkt falle sammen med geo-

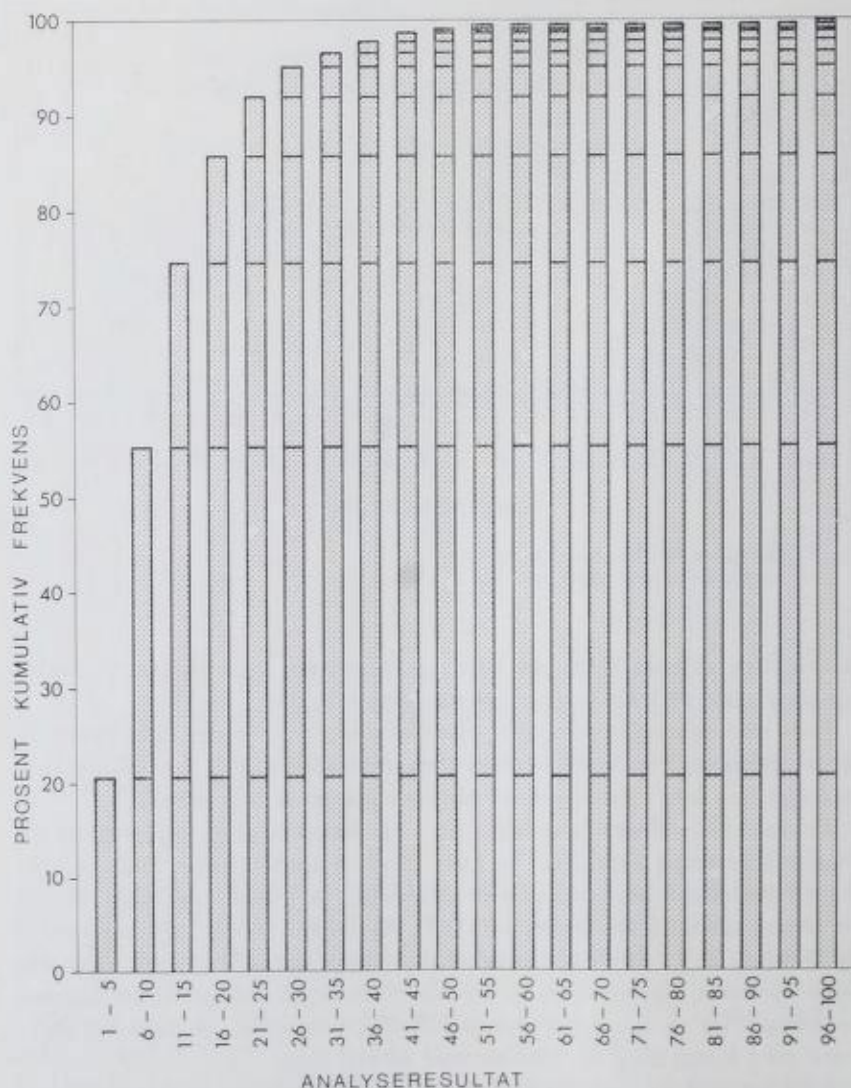


Fig. 3. Diagram for kumulativ frekvensfordeling av data tatt fra fig. 1.

Cumulative frequency distribution plot of data taken from Fig. 1.

metrisk middel (G) av den målte variable (som er det samme som antilogaritmen til det aritmetiske middel av logaritmene til den målte variable).

$$\text{Median (M)} = x_{50} \%$$

$$\text{Geometrisk middel (G)} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \text{antilog} (\Sigma \log x) / n$$

Standardavviket (S) for logaritmene til den målte variable (X) kan anslås grafisk ved f.eks. å subtrahere logaritmen til 50 prosentilen fra logaritmen til 84 prosentilen.

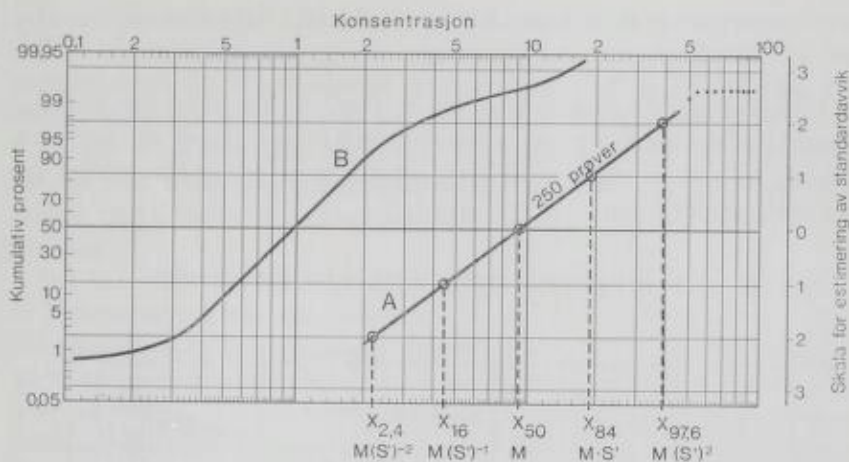


Fig. 4. Kumulative frekvensfordelinger.
 A. Lognormal fordeling.
 B. Tilnærmet lognormal fordeling.
 Cumulative frequency distributions.
 A. Lognormal distribution.
 B. Approximately lognormal distribution.

$$S = \log x_{84} - \log x_{50} = \log (x_{84}/x_{50})$$

En mer logisk beregning fåes hvis man også leser av 16 prosentilen og bruker både øvre og nedre del av kurven:

$$S = \frac{1}{2} [(\log x_{84} - \log x_{50}) + (\log x_{50} - \log x_{16})] = \frac{1}{2} \log (x_{84}/x_{16})$$

Ved lognormale fordelinger er det geometriske avvik (S') et hendig mål for spredningen. Det geometriske avvik kan defineres som antilogarithmen til S , og estimeres ved f.eks. å regne ut kvotienten mellom 84 og 50 prosentilen.

$$S' = x_{84}/x_{50}$$

Ønsker man å bruke både øvre og nedre del av fordelingskurven, blir den tilsvarende formel:

$$S' = \text{antilog } S = \sqrt{x_{84}/x_{16}}$$

I praksis kan følgende tilnærmede formel for S' være hensiktsmessig, idet medianverdien inngår, og 3 punkter på kurven blir benyttet:

$$S' = \frac{1}{2} (x_{84}/x_{50} + x_{50}/x_{16})$$

Den tilsvarende formel for S blir:

$$S = \log \frac{1}{2} (x_{84}/x_{50} + x_{50}/x_{16})$$

N.B. De angitte mål for sentralverdi og spredning ved lognormal fordeling må ikke forveksles med aritmetisk middel (α) og standardavvik (β) for ikke log-transformerte data. Størrelsen α og β kan beregnes etter følgende formel (Aitchison & Brown, 1966, s. 8),

$$\alpha = \exp. [2,3 S/n + \frac{1}{2} (2,3 S)^2]$$

$$\beta = \exp. [(2,3 S)^2 - 1] (\exp(\alpha) \text{ betyr } e^{\alpha})$$

Man skal spesielt merke seg at for et utvalg fra en populasjon med lognormal-fordelte verdier er det aritmetiske middel (α), regnet ut etter ovenstående formel, et bedre estimat av populasjonens aritmetiske middel enn det nærliggende estimat

$$\bar{x} = \sum_{j=1}^n x_j/n \text{ som vanligvis vil gi for høyt resultat.}$$

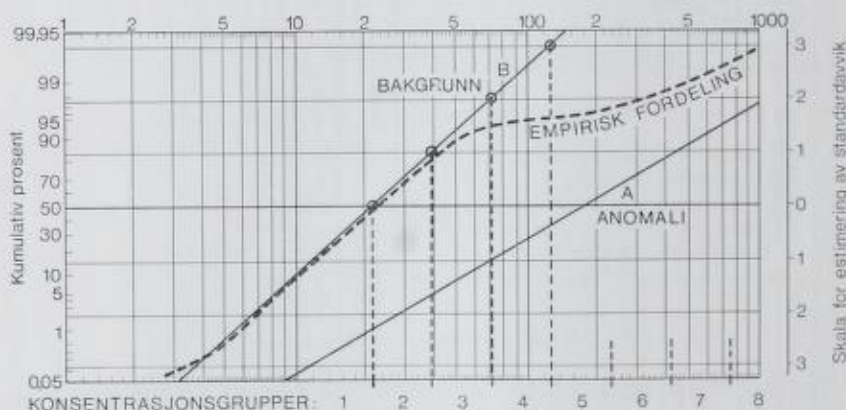


Fig. 5. Konsentrasjonsgruppering for karttegnig.
Grouping of concentrations for map presentation.

Ahrens (1954) mente at frekvensfordelingen for elementer i geologisk materiale vanligvis er lognormale. Flere forfattere har senere vist at regelen om lognormalitet ofte stemmer godt for sporelementer, og når geokjemiske sporelement-data fremstilles på logaritmisk sannsynlighetspapir, fåes ofte tilnærmet rette linjer (Kock & Link 1970 og 71).

Noen ganger er kurvene for geokjemiske data rettlinjet i midtpartiet på begge sider av ordinaten 50 %, men bøyer av i begge ender, ofte til en S-formet kurve. Denne avbøyning kan skyldes lavt prøveantall eller at analysemetodens relative nøyaktighet varierer med konsentrasjonen. Hvis prøveantallet er stort kan avbøyning, i fall den følger bestemte mønstre, indikere en kombinasjon av to eller flere lognormale fordelinger. Ved en kombinasjon av to fordelinger (bimodal fordeling), kan de to enkeltfordelinger la seg utskille ved grafisk løsning (Bolviken 1971). De to enkeltfordelinger kan da representere hver sin geokjemiske bakgrunn, eller den ene fordeling vil kunne betraktes som bakgrunn (fig. 5, kurve B), og den andre fordeling som den geokjemiske anomali (fig. 5, kurve A). Ofte kan man helt enkelt trekke en rett linje gjennom nedre og midtre del av den empiriske fordelingskurve og la denne linjen representere bakgrunnen.

Ved karttegnig av geokjemiske data er det vanligvis nødvendig å grup-

pere analyseresultatene. Man kan f.eks. gruppere slik at konsentrasjonsintervallene blir begrenset av medianen ganger hele potenser av bakgrunnens geometriske avvik (for logaritmene til konsentrasjonene tilsvarer dette aritmetisk middel pluss hele multiplum av standardavviket). I tillegg kan det forsvares å slå alle konsentrasjoner lavere enn medianen sammen i én gruppe hvis kartene skal brukes til malmløsing alene (fig. 5). Antall grupper vil forøvrig variere med hvordan bakgrunnsfordelingen ligger i forhold til den empiriske fordeling.

Det kan ellers ofte være hensiktsmessig å velge en universell gruppering av konsentrasjoner når man skal fremstille kart over geokjemiske data. Spesielt gjelder dette regionale undersøkelser som gjennomføres i flere etapper. Ved universell gruppering sikrer man da samme oppdeling av konsentrasjonene på et hvert stadium av undersøkelsene, slik at resultat-kartenes tegnforklaring blir ensartet over hele det undersøkte område. Siden fordelingene oftest er lognormale, bør en universell gruppering være slik at den gir like intervaller på en logaritmisk skala. Dette vil si at grensene for konsentrasjonsgruppene danner en geometrisk rekke. En hensiktsmessig fremgangsmåte vil det da være å dele hver tierpotens i et bestemt antall deler. Tabell 2 angir de konsentrasjonsgrenser og intervaller man får ved deling av logaritmene mellom hver tierpotens i henholdsvis 3, 4 og 5 like deler.

Uansett hvilken konsentrasjonsgruppering som velges, bør følgende iakttas:

Det bør alltid være et visst forhold mellom antall konsentrasjonsgrupper og det totale variasjonsområde for konsentrasjonene. Det er således rimelig at et element med store konsentrasjonssprang får flere grupper enn et med små sprang. I fall brukeren av det geokjemiske kart kan ventes å basere sin tolking på resultater i enkeltpunkter, bør konsentrasjonsintervallene ikke være for trange i forhold til prøvetakingsfeilen og analysemetodens presisjon. Analyselaboratoriet kan ha angitt analyseresultater under en egentlig rimelig følsomhetsgrense med endelige tall. Slike resultater bør ikke grupperes, men slås sammen med de resultater som er angitt som «0» eller «mindre enn» en bestemt grense. Dette innebærer at geokjemikeren bør ha klare data om prøvetakingsfeilen og om analysemetodens presisjon og følsomhet.

På geokjemiske kart i sort-hvitt blir vanligvis hver konsentrasjonsgruppe angitt ved skravur, isokonsentrasjonslinjer eller enkeltstående symboler. Frekvensfordelingen bør da alltid ledsage det geokjemiske kart. Spesielt gjelder dette hvis det er valgt en universell konsentrasjonsgruppering. Et mål for hvor høy en konsentrasjon er, fåes ved å undersøke hvor på frekvensfordelingen konsentrasjonen ligger. Man kan f.eks. angi antall geometriske avvik over medianen. Et mål for hvor «høyt» et utvalg av flere konsentrasjoner er fåes ved å sammenligne frekvensfordelingen for utvalget med frekvensfordelingen for den populasjon utvalget er tatt fra. Det er utarbeidet numeriske metoder for slik sammenligning (se f.eks. Kock & Link 1970, side 132—200). En oversiktlig visuell sammenligning kan man oppnå ved å fremstille både utvalget og hele populasjonen på sannsynlighetspapir.

I litteraturen støter man på begrepet terskel (engelsk threshold) som be-

tegnelse på grenseområdet mellom bakgrunn og anomali. Ofte defineres da terskel som den konsentrasjon som bare overskrides av 2,5 % av prøvene (ved uni-modale, normale og lognormale fordelinger er dette den konsentrasjon som svarer til 2 standardavvik over medianen). Det kan synes like rimelig å anvende begrepet terskel på den grense som overskrides av «høye prøver». Hvis begrepene bakgrunn og anomali oppfattes som statistiske fordelinger med enhver konsentrasjon mellom 0 og et felles maksimum som en mulig observasjon, blir nemlig den tradisjonelle definisjon av begrepet terskel meningsløs. Prinsipielt vil det da ikke eksistere andre minimumsverdier enn 0 og andre maksimumsverdier enn det teoretiske maksimum i noen av fordelingene. Det ligger i dette at begrepet geokjemisk anomali ikke knyttes ene og alene til forekomst av høye metall-konsentrasjoner, men at også særtrekk ved den geografiske fordeling av konsentrasjoner i mer sentrale deler av konsentrasjonsskalaen bør tillegges betydning ved tolkning av geokjemiske kart.

Tabell 1. Statistiske parametre beregnet ut fra kurvene i fig 4
Estimated statistical parameters based on Fig. 4

	Kurve A	Kurve B
Median, $M = x_{50}\%$	9.4	0.97
Geometrisk middel, $G = M$	9.4	~ 0.97
Aritmetisk middel av logaritmene, $\log G$	0.973	~ 0.013
Geometrisk avvik $S' = \frac{1}{2}(x_{94}/x_{20} + x_{20}/x_{16})$	2.07	~ 1.69
Standardavvik $S = \log \frac{1}{2}(x_{94}/x_{20} + x_{20}/x_{16})$	0.316	~ 0.228

Tabell 2. Konsentrasjonsgruppering for geokjemiske kart
Grouping of concentrations for geochemical maps

3-deling av hver tierpotens			4-deling av hver tierpotens			5-deling av hver tierpotens		
Grense	Avrundet	Intervall	Grense	Avrundet	Intervall	Grense	Avrundet	Intervall
1.00	1.00	1.01— 1.80	1.00	1.00	1.01— 1.80	1.00	1.00	1.01—1.60
2.15	2.20	2.21— 4.60	1.78	1.80	1.81— 3.20	1.58	1.60	1.61— 2.50
4.65	4.60	4.61— 10.00	3.17	3.20	3.21— 5.60	2.50	2.50	2.51— 3.90
10.00	10.00	10.1 — 22.0	5.63	5.60	5.61—10.0	3.89	3.90	3.91— 6.30
21.5	22.0	22.1 — 46.0	10.00	10.00	10.0 —18.0	6.30	6.30	6.31—10.00
46.5	46.0	46.1 —100.0	17.80	18.00	18.1 —32.0	10.00	10.00	10.01—16.00
100.0	100.0	osv.	31.70	32.00	32.1 —56.0	15.80	16.00	16.01—25.00
osv.	osv.		56.40	56.00	osv.	25.00	25.00	25.01—39.00
			osv.	osv.		38.90	39.00	osv.
						osv.	osv.	

LITTERATUR

- Ahrens, L. H. 1954: The lognormal distribution of the elements. *Geochim. Cosmochim. Acta* 5, 49—73.
- Aitchison, J. & Brown, J. A. C. 1966: *The Lognormal Distribution*. 176 p. Cambridge University Press.
- Boyle, R. W. & McGerrigle, J. I. (eds.) 1971: *Geochemical Exploration*. Proceedings, 3rd International geochemical exploration Symposium, Toronto, April 16—18, 1970. *Can. Inst. Min. Metall. Spec. Volume 11*, 594 p.
- Brooks, R. R. 1972: *Geobotany and Biogeochemistry in Mineral Exploration* 290 p. Harper & Row.
- Bolviken, B. 1971: A statistical approach to the problem of interpretation in geochemical prospecting. In: Boyle, R. W. & McGerrigle, J. I. (eds.): *Geochemical Exploration*. *Can. Inst. Min. Metall. Spec. Volume 11*, 564—67.
- Cameron, E. M. (ed.) 1967: Proceedings, Symposium on geochemical prospecting, Ottawa, April 1966, *Geol. Surv. Canada, Paper 66—54*.
- Canney, F. C. (ed.) 1969: International geochemical exploration Symposium. *Colorado Sch. Mines*. 64, No. 1.
- Duval, J. S., Schwarzer, T. & Adams, J. A. S. 1972: Lognormal distribution of trace elements in the environment. In: Hemphill, D. D. (ed.): *Trace Substances in Environmental Health — V. Proceedings of University of Missouri's 5th Annual Conference on Trace Substances in Environmental Health*, June 29 — July 1, 1971, pp. 120—31. University of Missouri.
- Hawkes, H. E. & Webb, J. S. 1962: *Geochemistry in Mineral Exploration*. 415 p. Harper & Row.
- Jones, M. J. (ed.) 1973: Geochemical prospecting. Proceedings of the 4th International geochemical exploration Symposium. London 1972. *Inst. Mining Metall.* (in press).
- Koch, G. S. & Link, R. F. 1970, 1971: *Statistical Analysis of Geological Data*. Vol. I 1970, 375 p. Vol. II 1971, 438 p. John Wiley & Sons.
- Lepeltier, C. 1969: A simplified statistical treatment of geochemical data by graphical representation. *Econ. Geol.* 64, 538—50.
- Tennant, C. B. & White, M. L. 1959: Study of the distribution of some geochemical data. *Econ. Geol.* 54, 1281—90.
- Williams, X. K. 1967: Statistics in the interpretation of geochemical data. *N.Z. J. Geol. Geoph.* 10, 771—97.