NAT'L INST. OF STAND & TECH
A11106 138699
NIST
PUBLICATIONS

NBSIR 76-1147

A Combined Least Sums and Least Squares Approach to the Evaluation of Thermodynamic Data Networks

D. Garvin, V. B. Parker, D. D. Wagman and W. H. Evans

Physical Chemistry Division
Institute for Materials Research
National Bureau of Standards
Washsington, D. C. 20234

July 1976

Interim Report

Prepared for
Office of Standard Reference Data, NBS

QC
100
.U56
#76-1147
1976
c.2

NBSIR 76-1147

A COMBINED LEAST SUMS AND LEAST SQUARES APPROACH TO THE EVALUATION OF THERMODYNAMIC DATA NETWORKS

D. Garvin, V. B. Parker, D. D. Wagman and W. H. Evans

Physical Chemistry Division
Institute for Materials Research
National Bureau of Standards
Washington, D. C. 20234

July 1976

Interim Report

Prepared for
Office of Standard Reference Data, NBS

U.S. DEPARTMENT OF COMMERCE, Elliot L. Richardson, *Secretary*
Edward O. Vetter, *Under Secretary*
Dr. Betsy Ancker-Johnson, *Assistant Secretary for Science and Technology*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Acting Director*

Table of Contents

I.	Introduction	1
II.	The Type of Problem	5
Table I.	Types of Thermodynamic Measurements	8
III.	Data Evaluation	9
IV.	The Sequential Method	10
V.	Advantages of Using Computer Techniques	13
VI.	Computer Assisted Data Evaluation	15
VII.	Comparison of Methods: Goodness of Fit	18
Table 2.	Statistics on Comparisons of Sequential and Simultaneous Solutions. "Residuals" are Observed - Calculated Values on Measurements	21
VIII.	Discussion	22
References		24
Figures		26

A Combined Least Sums and Least Squares Approach to the Evaluation of Thermodynamic Data Networks*

David Garvin, Vivian B. Parker, Donald D. Wagman and William H. Evans

National Bureau of Standards

Washington, D. C. 20234

Abstract

A description is given of a system for computer-based evaluation of interrelated thermodynamic measurements of enthalpies of reaction, equilibria and entropies. This system is an extension of the CATCH program developed by J. B. Pedley, University of Sussex. In the new system linear least sums and least squares techniques are used to solve networks of thermodynamic equations to obtain the enthalpies and free energies of formation and the entropies of chemical substances. The least sums technique is shown to be useful in assessing the consistency of the data. A method combining least sums and least squares solutions, provides a weighted solution that reproduces closely the solutions that are obtained by a detailed analysis of the data using the customary sequential procedure. The results from tests on four large networks involving compounds of B, U, Rb and salts of Sn, Pb, Cd and Hg are discussed.

Keywords: Computerized Data Analysis, Data Evaluation, Least Squares (L2), Least Sums (L1), Thermochemistry, Thermodynamic Data Networks.

I. Introduction

Tables of thermodynamic properties of substances are used widely for the prediction of the energetics of chemical processes and of the yields at equilibrium. Typically these tables list enthalpy of

*Presented at the 5th International CODATA Conference, Boulder, Colorado June 28-July 1, 1976.

formation, ΔHf, Gibbs energy of formation, ΔGf, entropy, S, and heat capacity, Cp, at a reference temperature, usually 298.15 K, the enthalpy of formation at 0 K, $\Delta Hf°$, and the change in enthalpy between 0 and 298 K, $H°(298)-H°(0)$. They may also list properties at transition temperatures. Other tables give thermal functions, the thermodynamic properties H, G, S, and Cp as a function of temperature.

Construction of these reference data tables is an exercise in the art of data evaluation. Often the properties of a substance can be determined using several different sets of measurements. The experiments usually do not agree. Choices must be made in establishing the "best" values to be used by technologists.

Two methods are now being used to construct these tables. One is a sequential approach in which the data networks are solved piece by piece according to a preset strategy. This method has been used for decades. More recently simultaneous solutions of the data networks have been made using computer techniques.

Our concern in this paper is with the quality of the computer solutions of data networks. How well do the machine solutions compare with the hand-crafted selections? What human controls are required to assure high quality? What parts of the problem can be standardized? Are these techniques suitable for construction of tables now and in the near future? In particular, are they suitable for use in setting the CODATA Key Values for Thermodynamics?

We present here a novel method for a computer-based evaluation of thermodynamic data that combines the use of least sums and least squares techniques in the simultaneous solution of data networks. This is compared to the sequential method in four experiments. The new method appears to be suitable for general use.

Starting in the mid 1960's two groups in the USA and the USSR, have been preparing and publishing large up-dated tables. These tables are the National Bureau of Standards (US) "Selected Values of Chemical Thermodynamic Properties" [1] and the Institute for High Temperature (USSR) "Thermal Constants of Compounds" [2]. At the present time both of these programs are nearing completion.

These tables and all earlier ones have been constructed using the sequential method.

During the past ten years an alternative method for preparing tables of self consistent thermodynamic properties has been developed. This is the simultaneous solution of all of the measurement data using computer techniques. Guest, Pedley and Horn [3] used linear least squares analysis on the enthalpy measurements for boron compounds. They pointed out the advantages of the computer system: (a) a data base of evaluated measurements is accumulated for future use, (b) new measurements may be incorporated easily in a solution, and (c) the selected values may be revised quickly as new CODATA Key Values for Thermodynamics [4] are established.

Pedley expanded the scope of this effort with the establishment of the CATCH (Computer Analysis of Thermochemical Data) series of tables [5], of which five have been issued covering enthalpy data on compounds of nitrogen, phosphorus, the halogens, silicon, chromium, molybdenum and tungsten. Two features of the CATCH system are important: (a) cooperative data evaluation, the data for each table being assessed by thermochemists familiar with the particular subject, and (b) the possibility of issuing new, updated tables at frequent intervals.

Slightly earlier Syverud and Klein had developed a similar simultaneous solution system [6] at the Dow Chemical Company on contract from the National Bureau of Standards. It had the capability for handling mixed sets of enthalpy, free energy and entropy data and provided solutions based on either linear least squares or least sums techniques. Their system has been used in evaluation of data for the JANAF Thermochemical Tables [7] and for the alkali metal ions in the CODATA Key Values for Thermodynamics [4].

Since 1972 there has been active cooperation between the CATCH program at the University of Sussex, England, and the Chemical Thermodynamics Data Group at NBS. Jointly a single system is being developed for the manipulation of thermochemical data bases and the machine-based preparation of selected values.

II. The Type of Problem

The mathematical problem to which a network of thermochemical measurements reduces is a set of linear algebraic equations in many variables. The variables, or some of them, are overdetermined.

$$\Sigma_i a_{ij} x_i = y_j \pm u_j$$

$$i = 1, \ldots n, \; j = 1, \ldots m; \; m > n$$

where the a's are known coefficients, the x's are the variables to be solved for, y is a measured value and u is its uncertainty.

The thermochemical experiments from which this set of equations is derived all measure the change in a property of a system. These changes are, by convention, expressed in terms of properties of the individual substances involved in the process.

For example, consider the enthalpy of solution of one mole of rubidium oxide in water

$$Rb_2O(\text{crystal}) + H_2O(\text{liquid}) \rightarrow 2RbOH(\text{aqueous})$$

$$\Delta H = -338 \pm 3 \text{ kJ}$$

which is expressed as

$$\Delta H = 2 \cdot \Delta Hf(RbOH) - \Delta Hf(Rb_2O) - \Delta Hf(H_2O)$$

where each ΔHf represents the molar enthalpy of formation of the substance from the elements in its formula. In the mathematical problem this becomes

$$y_1 \pm u_1 = -338 \pm 3 = 2 x_1 - x_2 - x_3$$

Similar equations relating the measured changes in Gibbs (free) energies of formation, ΔGf, or the entropies, S, may also be written.

Very often in the solution of a thermochemical network the properties of compounds of one particular element are considered variables and those of other elements are held fixed. In the example above, when solving for properties of rubidium compounds, $\Delta Hf(RbOH)$ and $\Delta Hf(Rb_2O)$ are variables but $\Delta Hf(H_2O)$ would be held fixed at a preassigned value and lumped with the measurement.

When networks involving compounds of only one element as variables are examined several characteristics stand out. First, there are many different types of measurements to be combined. They use different techniques and measure different classes of properties. The principal types of measurements are listed in Table 1. It is very difficult to compare the reliability of these widely differing measurements; one is comparing apples and oranges. Second, when the measurements are reduced to mathematical form almost all have the following simple forms:

$$y = ax$$

$$y = a_1x_1 - a_2x_2$$

$$y = a_1x_1 - a_2x_2 - a_3x_3$$

with the third case being relatively uncommon. Third, the networks, although containing 10 to 100 variables, are linked loosely. Each variable appears in only a small percentage of the equations.

Fourth, thermodynamic laws place constraints upon the solution in the form of interrelations between variables. The most important of these is the relation

$$\Delta G = \Delta H - T\Delta S$$

This is the only linkage between the three main classes of measurements shown in Table 1.

The mathematical approach is simple: solve an overdetermined set of linear algebraic equations, using an accurate algorithm. Preparation of the input data is more difficult. Computational programs require substantial sections to handle chemical bookkeeping and quality control. And, most important, very careful examination of the experiments is required. This often involves extensive preliminary computations.

TABLE I. TYPES OF THERMODYNAMIC MEASUREMENTS

ENTHALPY (HEAT) CHANGES

- DIRECT FORMATION OF A COMPOUND FROM THE ELEMENTS
- COMBUSTION (OXIDATION) OF A COMPOUND
- REACTION BETWEEN TWO (OR MORE) COMPOUNDS
- DECOMPOSITION REACTIONS
- SOLUTION OF A COMPOUND
- DILUTION OF A SOLUTION
- PHASE CHANGES
- IONIZATION AND APPEARANCE POTENTIALS

GIBBS ENERGY (FREE ENERGY) CHANGES

- EQUILIBRIUM MEASUREMENTS (OFTEN AT HIGH TEMPERATURES)
  - SOLUBILITY
  - DILUTION
  - VAPORIZATION
  - REACTION
  - ELECTROMOTIVE FORCE (CELL DATA)

ENTROPY MEASUREMENTS

- ABSOLUTE ENTROPIES FROM HEAT CAPACITY DATA
- TEMPERATURE COEFFICIENTS OF EQUILIBRIA AND EMF DATA

III. Data Evaluation

There are two fundamental elements that are common to both of the methods discussed here for preparing tables of thermochemical data. In our opinion they should underlie any method.

The first is the careful examination of each measurement by the data evaluator. This is the heart of data evaluation and cannot be automated. It is necessary to determine, in the light of present knowledge, whether or not the interpretation of the chemistry was correct and whether the technique was suitable. Reinterpretation and correction of the data to standard conditions may be necessary. The overall reliability of each data item must be estimated if it is to be combined rationally with others. This estimate is difficult to make and often may be subjective.

The second is the application of a set of criteria for a reasonable solution for ΔHf, ΔGf, and S° for interrelated compounds. (1) The values selected must reproduce well those measurements considered to be reliable. (2) The values selected must be as consistent as possible with all other values in the tables and should be in reasonable accord with the properties of similar substances or with physico-chemical correlations. (3) A consistent set of auxiliary data should be used throughout the entire set of tables. These auxiliary data include both the values for physical constants and the properties of substances ubiquitous to thermodynamic measurements.

If these criteria are satisfied, one presumes that the individual ΔHf's etc. may be combined to predict the "best" value of the enthalpy, Gibbs energy or entropy change for any process, measured or not.

IV. The Sequential Method

How are the criteria for a reasonable solution applied in the traditional sequential or hand-crafted method? Values for certain compounds that are constantly and repeatedly needed, the "core" auxiliary data, such as those for H_2O, OH^- (aq. std. state), common acids and bases, CO_2, etc., are evaluated first. It is then customary to evaluate all of the data related to compounds of one element at one time.

To do this the data analyst assembles all of the data on compounds of an element and proceeds to evaluate and calculate the properties ΔHf, ΔGf, and S compound by compound, working through the network. The analysis starts with compounds whose properties can be determined independently, that is, they depend only on known auxiliary data but not on other compounds of the same element. Then the properties of other compounds dependent on these first selections are set. If several measurement paths lead to the same compound a confirmation of the choice may be obtained. On the other hand, some or all of the previous selections may have to be revised to obtain a reasonable overall fit. A sample network is shown in Figure 12. Compounds a, b, c and d are determined first. There are direct connections involving a and c and also c and d. Then compounds e to l are selected. Finally, m and n are set.

When discrepancies are noted, decisions have to be made as to how to resolve them. Should a path be ignored as being a poor measurement, or should a weighted average be taken? Is it possible that the problem is not that of erroneous measurements but of a poor value selected in an earlier step or in the auxiliary data? In any case, the evaluator must retrace his steps, find the suspect value and modify that selection, and all the subsequent values dependent upon it. Measurements that originally appeared to be reliable become suspect and are downgraded if they are highly inconsistent with values arrived at from other paths in the over-determined set. The skeleton of values for key compounds is built up carefully, taking care to reproduce well the measurements on which they are based. Because of these various factors, this manual sequential method is, in reality, iterative. More than one pass is involved in establishing the final values for the key network.

The major advantage to this way of evaluating data is that the evaluator works from positions of strength, that is, he emphasizes the definitive measurements and builds a framework with them, fitting and adjusting small networks until he arrives at a set of stable values.

This very advantage can be a disadvantage. If new and significant experimental data become available it becomes difficult

to incorporate or to modify selections without going through another complete pass for the network. This is time consuming. In addition, once the "selected" values are incorporated into other tables within the series, they cannot be changed without considering the effect upon all compounds evaluated after them. As a result it is a major problem to update. The system is static in between major revisions.

V. Advantages of Using Computer Techniques

There are advantages to machine manipulation of data that have led us to attempt to adapt them, with certain reservations, to thermochemical data evaluation. They include:

1. Machine techniques are ideal for manipulating large masses of data. Many calculations are routine and tedious. Let the computer do them!

2. Analyses can be made with the computer to determine the effect of particular items of data. This is particularly important for new measurements. For an extensive network these analyses would be too time consuming to do by hand.

3. A systematic diagnosis of the fit of the data can be obtained.

4. An analysis of the network can be made to be used as a guide to the strategy of data evaluation, and to indicate the importance of various substances.

5. Solutions can be made for the same data base with different sets of auxiliary data, such as those required for the CODATA Key Values, the NBS Technical Notes or the Institute for High Temperatures tables.

6. Most important, the evaluated measurements can be stored in easily reusable form for a future calculation.

But with these advantages come some potential disadvantages that may lead to loss of control by the data analyst. They are:

1. Large masses of data are treated at one time. They may overwhelm the data analyst and he may lose the "feel" for the measurements that is an important part of the sequential method. This could be particularly important when a stored data base is used some time in the future.

2. The evaluation of individual measurements separately followed by a solution of the entire network can conceal systematic errors in a series of related measurements that would appear early in the sequential solution.

3. It may be difficult to identify measurements that are suspect and require reexamination or reinterpretation.

4. The solutions may be mathematically acceptable but not consistent with physical chemical correlations and experience.

In the next section we present an approach that takes advantage of the machine capabilities and appears to minimize these disadvantages.

VI. Computer Assisted Data Evaluation

A multistage simultaneous solution technique is presented in this section. It has been designed to take into account the two features of the data emphasized by data analysts: reliability and consistency with other measurements. The procedure is outlined and discussed below. Then the results of some tests are given.

Step 1. The measurements are examined by a data analyst. An overall uncertainty is assigned to each one that is tentatively accepted. This step is common to all evaluation procedures.

Step 2. An equally-weighted simultaneous solution is made for the entire data set.

This is used to assess the overall consistency of the data and as a diagnostic tool to identify measurements that may require reexamination. The solution may or may not confirm the original judgement of reliability made by the data analyst.

Our preference for this first solution is the least sums technique in which the sum of the absolute values of the residuals is minimized. This technique selects a set of equations and solves them exactly. This set is analogous to the median of a group of numbers. This technique is much less sensitive to outlying values than is least squares.

Step 3. The residual on each measurement in the first solution, i.e. the difference between observed and calculated value, is combined with the preassigned uncertainty to give an overall

measure of goodness. We call this the "average fit". It is a simple average. It can be considered as an uncertainty modified to allow for consistency with other measurements.

Step 4. A second solution is made using a weighted linear least squares technique. The final answers are the output of this solution.

The weight used for each measurement is the reciprocal of the "average fit" developed in step 3. In addition maximum and minimum limits are imposed on the weights to keep them within a 200 to 1 range.

This weighting procedure departs from that commonly used in least squares analysis, i.e., assignment of a weight proportional to the square of the reciprocal of the uncertainty. This new procedure has been adopted to de-emphasize the wide range of uncertainties found when measurements made using different techniques are to be combined. Without this de-emphasis a solution tends to favor strongly a few selected experiments.

The multistage automated procedure described above is performed by a single computer program. But, in our experience, application of it alone to a set of measurements is not sufficient. Intervention by a data analyst is desirable. This consists of checking the input data (measurements) for errors of interpretation and transcription, reanalysis of individual items, particularly those

that fit poorly with others, examination of the core network of measurements emphasized in the solution, and testing the answers for thermodynamic and physical acceptability. These interventions lead to revisions in the input data and reapplication of steps 2 through 4. In order to avoid small scale adjustments , we try to restrict the interventions (other than error correction and reinterpretation) to removal of a measurement from the set, adding new data, or requiring that a measurement be fit exactly. These major but simple (and easily apparent) changes appear to be sufficient.

VII. Comparison of Methods: Goodness of Fit

Several tests of the sequential (manual) and computer solutions have been made for four test problems. These are crude statistical tests designed to determine how well each solution reproduces the measurements and the extent to which they agree.

The tests were:

(a) Calculation of the differences between the answers for each variable and comparing them to the uncertainties assigned by the data analyst. These are presented in Figures 1, 2, and 3 which show comparisons only for variables that are crucial to the solutions. The solid line (of unit slope) in these Figures is the locus of points for which the difference in values for a variable obtained by the two methods would be equal to the total uncertainty in the value as estimated by the data analyst. These tests are indicative of performance but are not as strong as those that follow.

(b) The fit of each solution to the measurements were examined. The patterns of residuals were developed using stem and leaf histograms [8]. The questions to be answered were "Is this solution a reasonable fit and are the outliers understood?". The patterns for the two solutions to each test problem are presented in figures 4, 6, 8, and 10.

(c) The correspondence of the pair of solutions to each problem was examined by studying the differences between corresponding residuals. The questions asked were "Are these solutions fitting the same data the sameway and what is the significance of mismatches?". Figures 5, 7, 9 and 11 display these differences.

The statistics developed in these tests are given in Table 2. In general, the sequential and simultaneous solutions are indistinguishable. They fit the measurements equally well, as shown by the pattern of residuals. Most of the data are fitted well, but some are not. This is to be expected --experiments disagree. In each case the solutions are strongly correlated, as shown on the difference-of-residuals histograms. But in each case there are data that were treated differently. Most of these outliers turn out to be unimportant cases. The answers agree to within the expected accuracy in the vast majority of cases. The four networks are discussed below.

Boron compounds (enthalpy data). This problem was used to develop the multiphase method. The network, Figure 12, is strongly cross-linked. The "sequential" solution, Figure 4a, was not prepared by hand but is a least squares solution with arbitrary weights, assigned by the data analyst, that reflect the emphasis given to each item in an earlier manual solution. The high correlation between solutions, Figure 5, was the basis for adopting the method.

Rubidium compounds (enthalpy, Gibbs energy and entropy data). The network has many replicate measurements but is almost a tree, i.e., has few cross-links. Even within the cross-linked region the feasible calculation paths are restricted. (In effect, this problem is based on a directed graph). The high degree of correlation, shown in Figure 7, between the solutions is not surprising.

Uranium compounds (enthalpy data). The network here is large with several regions within which there is strong cross linkage but between which the coupling is limited. The spread in the fits, Figure 8, is more apparent here than in other cases, but this is due to poorer data and larger problem size. The correlation between the solutions is strong.

Key compounds of Sn, Pb, Cd, and Hg (enthalpy, Gibbs energy and entropy data). This is an entirely different comparison than the other three. For them the simultaneous and sequential solutions were made concurrently. In the present case the sequential solution, Figure 10a, is composed of selected values from NBS Technical Note 270, as chosen in 1968, solving data for each element separately. The simultaneous solution (in 1976), Figure 10b, is based on the measurements used to set CODATA Key Values and the network includes data on compounds of all four elements. About one-fourth of the measurements are new. Residuals for the sequential solutions were calculated using this updated set of measurements.

Both solutions appear to fit the data reasonably well. The correlation between the solutions, Figure 11, is about the same as in the other tests. Three fourths of the answers agree to better than their uncertainties. The remainder show that new selections are indeed desirable.

TABLE 2
Statistics on Comparisons of Sequential and Simultaneous Solutions. "Residuals" are Observed - Calculated Values on Measurements

Problem	Boron	Rubidium	Uranium	Sn, Pb, Cd, Hg
Simultaneous Solution				
Equation	50	75	178	161
Variables	13	30	68	53
σ/kJ	2.69	3.04	4.72	2.90
Equations Compared with Sequential Solution	50	75	145	161
Residuals/kJ				
Simultaneous Solution				
1st Quartile	-0.540	-0.230	-1.464	-0.172
Median	0.0	0.0	-0.067	+0.004
3rd Quartile	+0.406	+0.188	+0.519	+0.180
Range	-2.9, +5.5	-16.3, +14.0	-30.8, +34.2	-6.0, +9.8
Sequential Solution				
1st Quartile	-0.845	-0.209	-0.920	-0.201
Median	+0.054	0.0	0.0	-0.008
3rd Quartile	+0.996	+0.209	+1.256	+0.188
Range	-3.6, +6.4	-13.4, +12.6	-30.5, +41.4	-5.9, +10.2
Difference Between Residuals/kJ				
1st Quartile	-0.159	-0.063	-0.255	-0.117
Median	-0.021	+0.017	+0.100	+0.126
3rd Quartile	+0.042	+0.096	+1.243	+0.134
Range	-1.0, +1.1	-5.2, +8.2	-11.9, +25.9	-5.5, +4.5

VIII. Discussion

We have shown that computer solutions can be obtained that agree with the solutions from the sequential method, at least to the level of testing used here. It is possible that more detailed testing would indicate a clear preference for one method or the other. But we do not believe that this will happen, because of the high degree of agreement in the treatment of individual measurements and in the answers. A definitive choice must be based on other than statistical considerations.

The advantages of the computer solution, stated in Section V, tip the balance in its favor for the long term. Already we have exploited them to study the same data base, using both CODATA and NBS Technical Note 270 auxiliary data. We have also used the computer technique to update selections in light of important new data. And, most important, the method has helped us to identify measurements for which reinterpretation was necessary. The application to compounds of Sn, Pb, Cd and Hg has shown that the restriction of analysis to compounds of one element at a time is not necessary or even desirable.

These computer solutions do not replace the data analyst. They are aids. At least three quarters of the work is in the careful analysis of the individual measurements. This will remain, and will become increasingly important because it will be the principal contact that the analyst has with the data. Additional

quality control measures will be needed to assure that errors in the data can be spotted at this stage. But, on balance, we recommend the use of the computer algorithm given here as an aid to improving the efficiency of thermochemical data evaluation.

REFERENCES

[1] Wagman, D. D., et al., "Selected Values of Chemical Thermodynamic Properties", Nat. Bur. Stand. (U.S.) Tech. Note 270-3 (1968), 270-4 (1969), 270-5 (1971), 270-6 (1971), 270-7 (1973), U. S. Government Printing Office, Washington, D. C. [Parts 1 and 2 are included in part 3].

[2] Glushko, V. P. (ed.) et al., "Termicheskie Konstanty Veshchestv", Vol. 1 (1965), Vol. 2 (1966), Vol. 3 (1968), Vol. 4, part 1 (1970), part 2 (1971), Vol. 5 (1971), Vol. 6, part 1 (1972), part 2 (1973), Vol. 7, part 1 (1974), part 2 (1974), VINITI, Moscow.

[3] Guest, M. F., Pedley, J. B., and Horn, M., J. Chem. Thermodynamics 1, 345 (1969).

[4] CODATA Task Group on Key Values for Thermodynamics, CODATA Bull. No. 17 (Jan. 1976), and CODATA Special Report No. 3 (Sept. 1975), CODATA, Paris. [These publications include all recommendations in earlier reports.] Parker, V. B., Wagman, D. D., and Garvin, D., "Selected Thermochemical Data Compatible with the CODATA Recommendations", Nat. Bur. Stand. (U.S.) NBSIR 75-968 (1976).

[5] Pedley, J. B., (ed.), "Computer Analysis of Thermochemical Data (CATCH Tables)", University of Sussex, Brighton. Data selectors and tables issued:
Cox, J. D., "Halogen Compounds" (1972)
Pilcher, G., "Nitrogen Compounds" (1972)
Pedley, J. B., and Iseard, B. S., "Silicon Compounds" (1972)
Head, A. J., "Phosphorous Compounds" (1972)
Barnes, D. S., "Cr, Mo and W Compounds (1974)

[6] Syverud, A. N. and Klein, M. [Periodic Reports on USDC-NBS Contract CST 1165], Dow Chemical Co., Midland, Mich., 1964-65.

Syverud, A. N., "Computer Analysis of Interrelated Data: Least Sums and Least Squares Applied to Thermodynamic Data", typescript, Dow Chemical Co., Midland, Mich., 1965.

Syverud, A. N., "Simultaneous Adjustment of Thermochemical Data. I. Enthalpies of Formation of Hydrofluoric Acid and Five Closely Related Fluorides", (Jan. 1969), "II. Enthalpies of formation of C_2F_4, CF_3 and six CF_3X molecules", (Aug. 1969), typescript, Dow Chemical Co., Midland, Mich.

Syverud, A. N., "Simultaneous Adjustment of Enthalpy of Formation, Gibbs, Energy of Formation and Entropy by Least Squares", typescript, Dow Chemical Co., Midland, Michigan

[7] Stull, D. R. and Prophet, H., (eds.), "JANAF Thermochemical Tables, 2nd edition, Nat. Stand. Ref. Data Series NSRDS-NBS 37, U. S. Government Printing Office, Washington, D. C., 1971.

[8] Statistical Engineering Laboratory, "Omnitab II News Letter No. 15", Nat. Bur. Standards (U.S.), unpublished (1975), citing Tukey, J. W., "Exploratory Data Analysis", Addison-Wesley (1970).

FIGURES

Figure 1. Absolute values of the differences between the enthalpies of formation of boron compounds in two solutions. The abscissa is the analyst's estimate of the reliability of the value in the sequential solution.

Figure 2. Absolute values of the differences between thermodynamic properties of rubidium and uranium compounds in two solutions. Abscissa as for Fig. 1.

Figure 3. Absolute values of the differences between thermodynamic properties for key compounds of Sn, Pb, Cd, and Hg. Abscissa based on CODATA Tentative Key Values, Part VI (1976).

Figure 4. Residuals on measurements, boron enthalpy data. (a) Sequential solution, (b) Simultaneous.

Figure 5. Differences between residuals in two solutions, boron enthalpy data.

Figure 6. Residuals on measurements, rubidium data. (a) Sequential solution, (b) Simultaneous solution

Figure 7. Differences between residuals, rubidium data.

Figure 8. Residuals on measurements, uranium enthalpy data. (a) Sequential solution, (b) Simultaneous

Figure 9. Difference between residuals in two solutions, uranium enthalpy data.

Figure 10. Residuals on measurements, Sn, Pb, Cd and Hg data.

(a) Sequential solution, NBS Tech. Note 270 (1968),

(b) Simultaneous solution (1976)

Figure 11. Difference between residuals in two solutions, Sn, Pb, Cd and Hg data.

Figure 12. Network for enthalpy data on key compounds of boron.

a = B(amorph), b = H_3BO_3(c), c = BN(c), d = BF_3(g),

e = B_2H_6(g), f = BCl_3(g), g = BCl_3(l) h = H_3BO_3(soln, aq),

i = B_2O_3(am), k = B_2O_3(c), l = HBF_4(aq), m = H_3BO_3(HCl soln),

n = $(CH_3)_3NBH_3$(c).

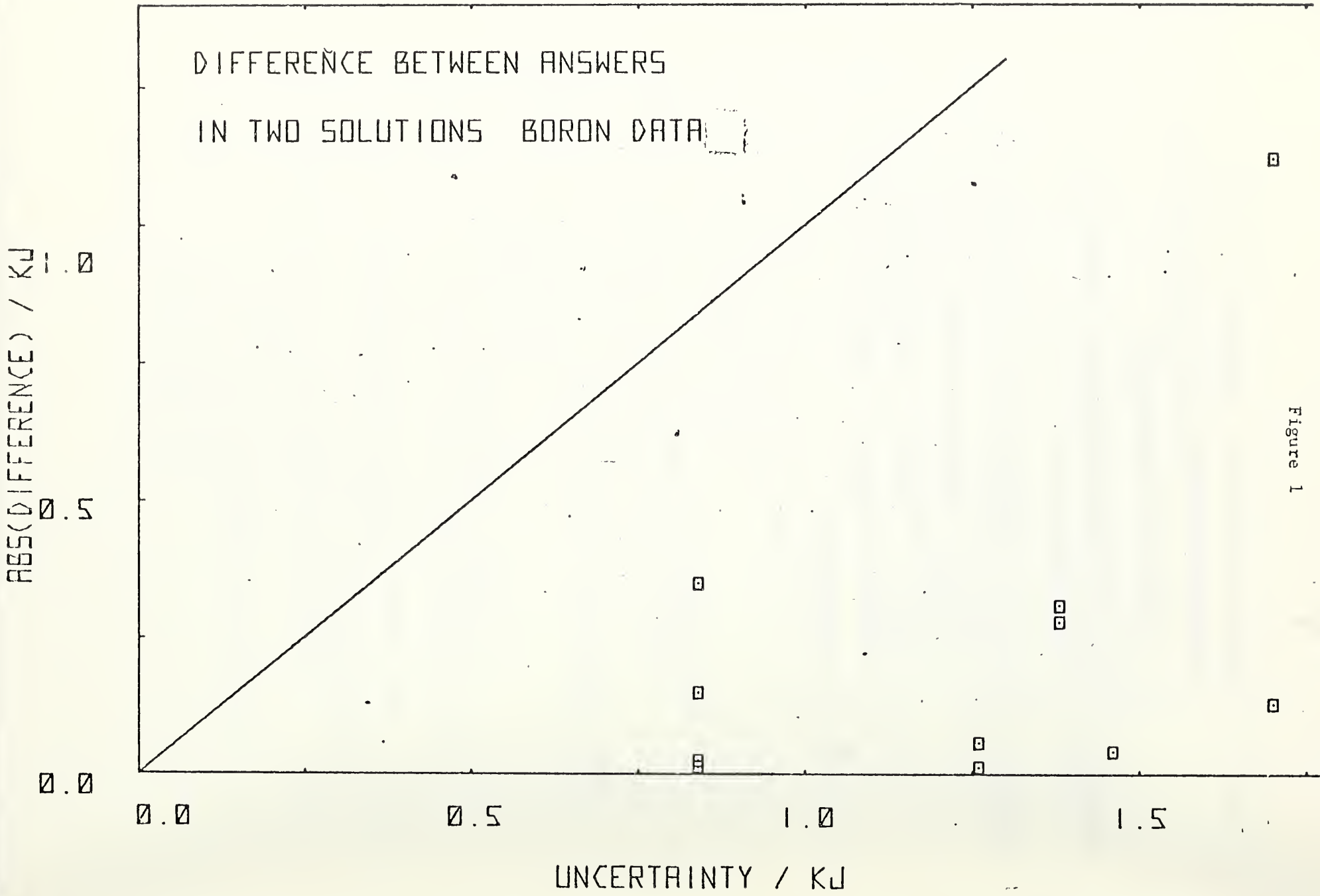

DIFFERENCE BETWEEN ANSWERS
IN TWO SOLUTIONS BORON DATA
ABS(DIFFERENCE) / KJ
1.0
0.5
0.0
0.0
0.5
1.0
1.5
UNCERTAINTY / KJ

Figure 1

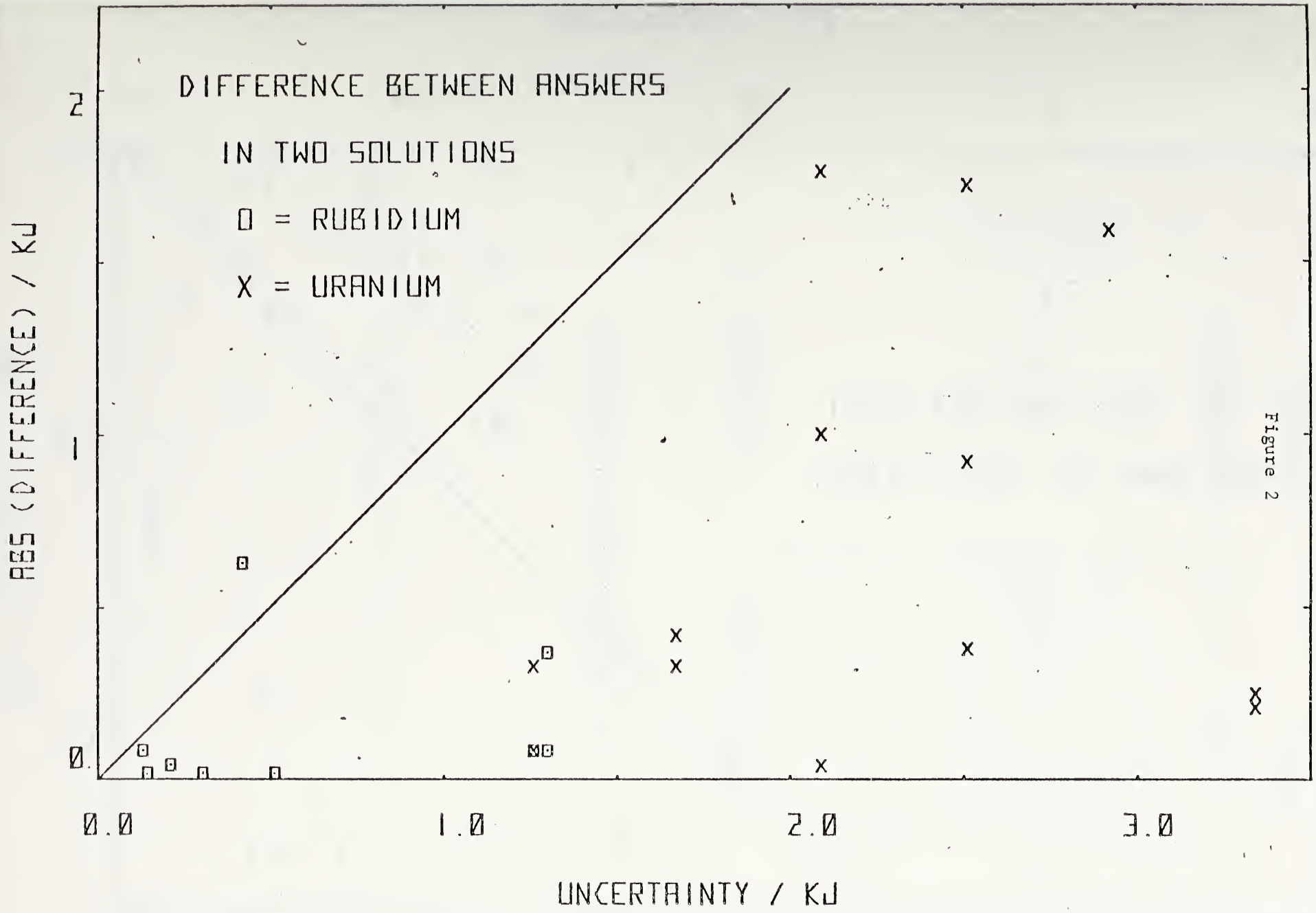

DIFFERENCE BETWEEN ANSWERS
IN TWO SOLUTIONS
0 = RUBIDIUM
X = URANIUM
ABS (DIFFERENCE) / KJ
2
1
0
0.0
1.0
2.0
3.0
UNCERTAINTY / KJ

Figure 2

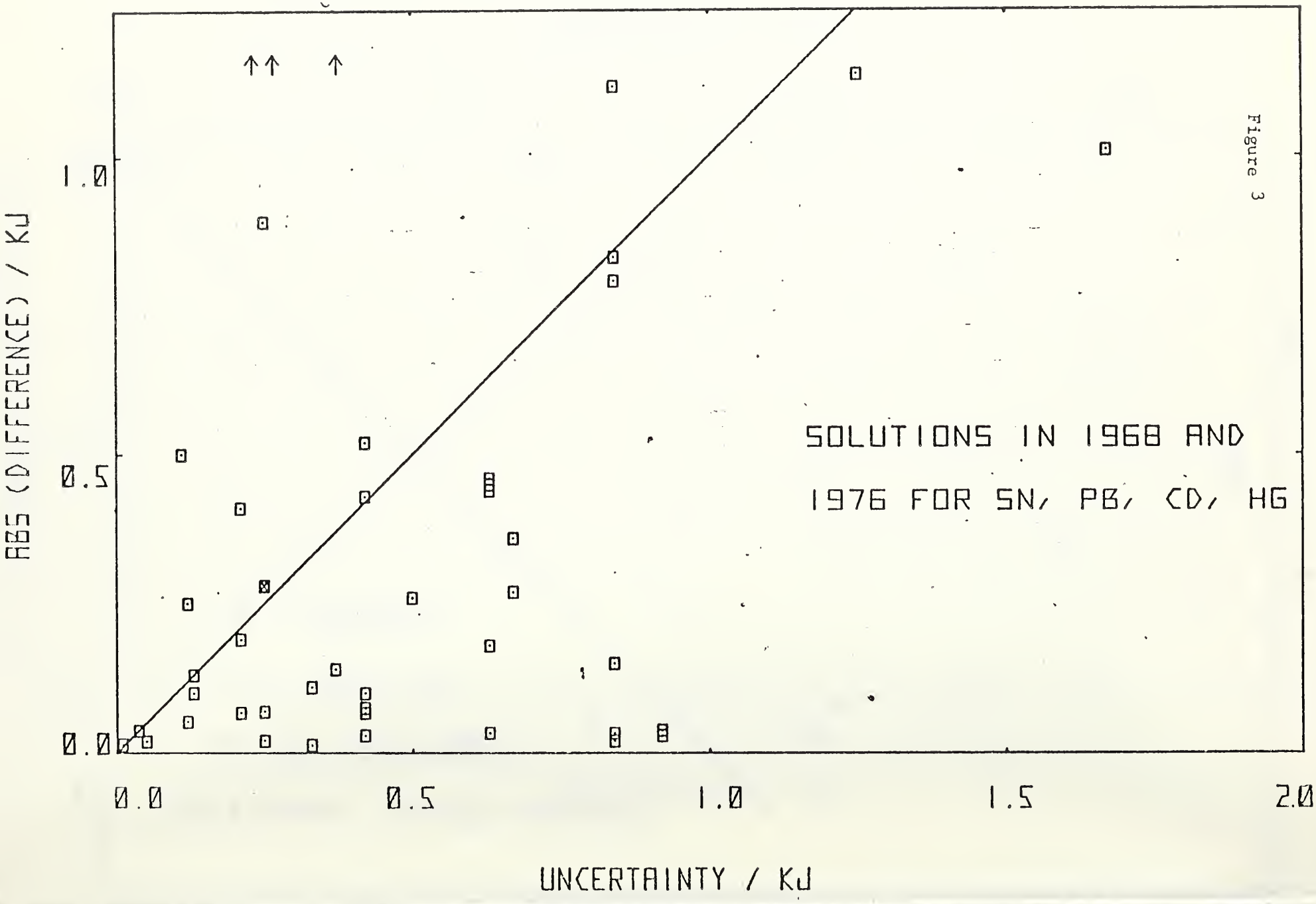
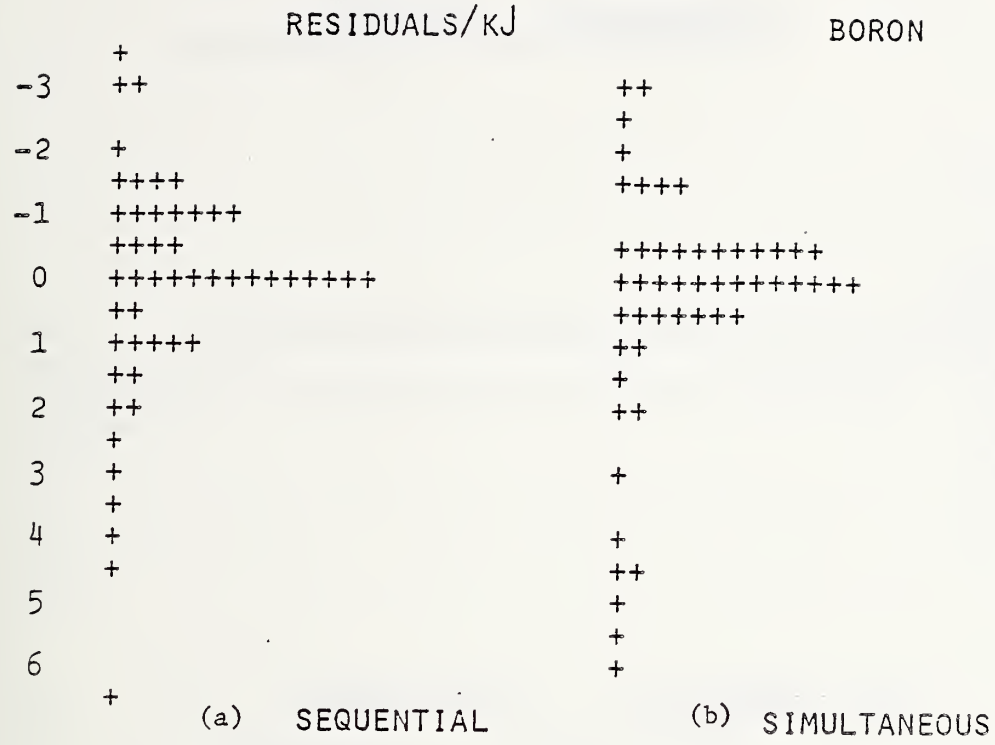

ABS (DIFFERENCE) / KJ
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
UNCERTAINTY / KJ
SOLUTIONS IN 1968 AND
1976 FOR SN, PB, CD, HG

Figure 3

Figure 4



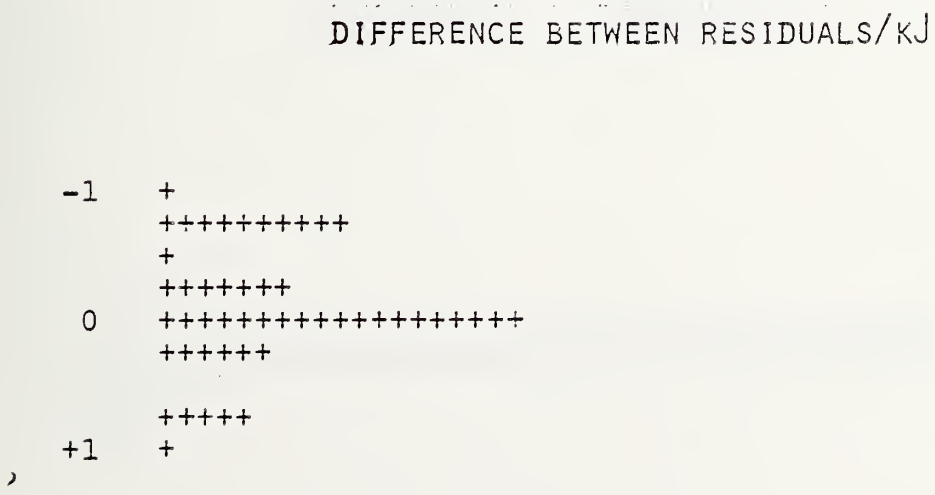
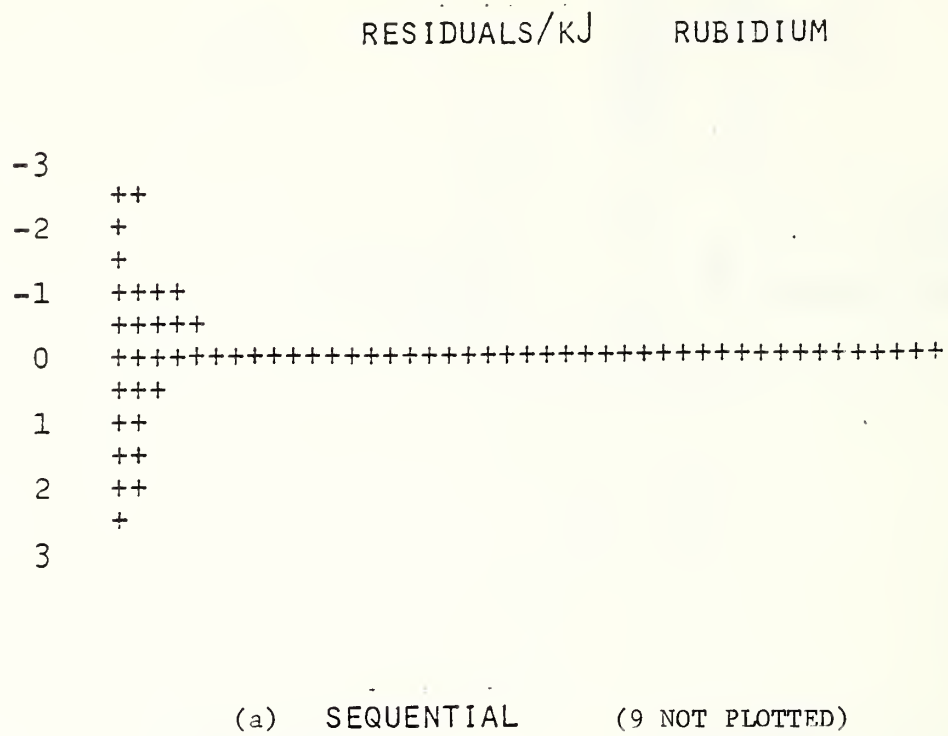
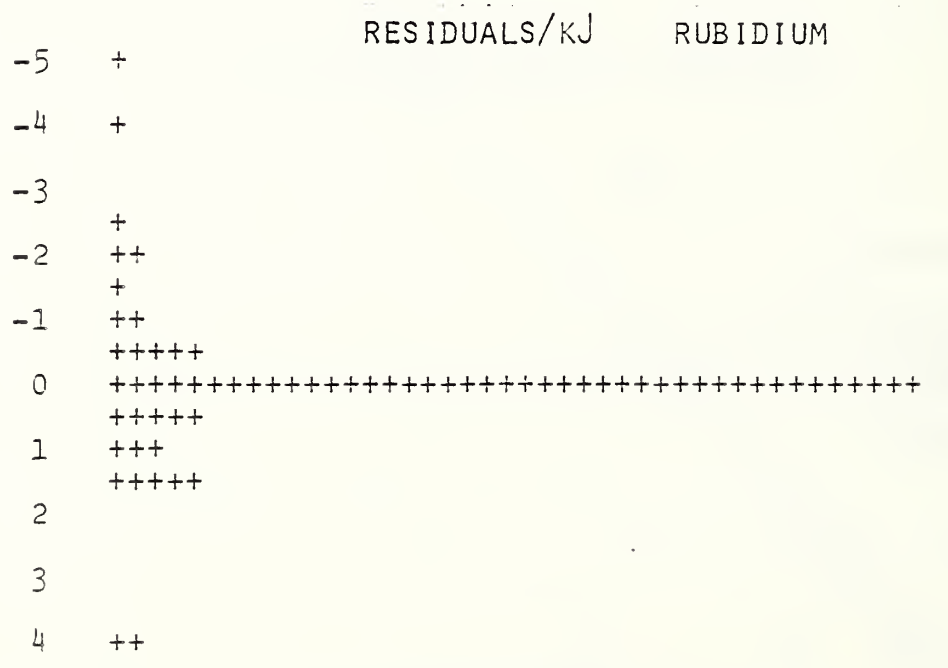


Figure 5

Figure 6 (a)



(a) SEQUENTIAL (9 NOT PLOTTED)



(b) SIMULTANEOUS (5 NOT PLOTTED)

Figure 7

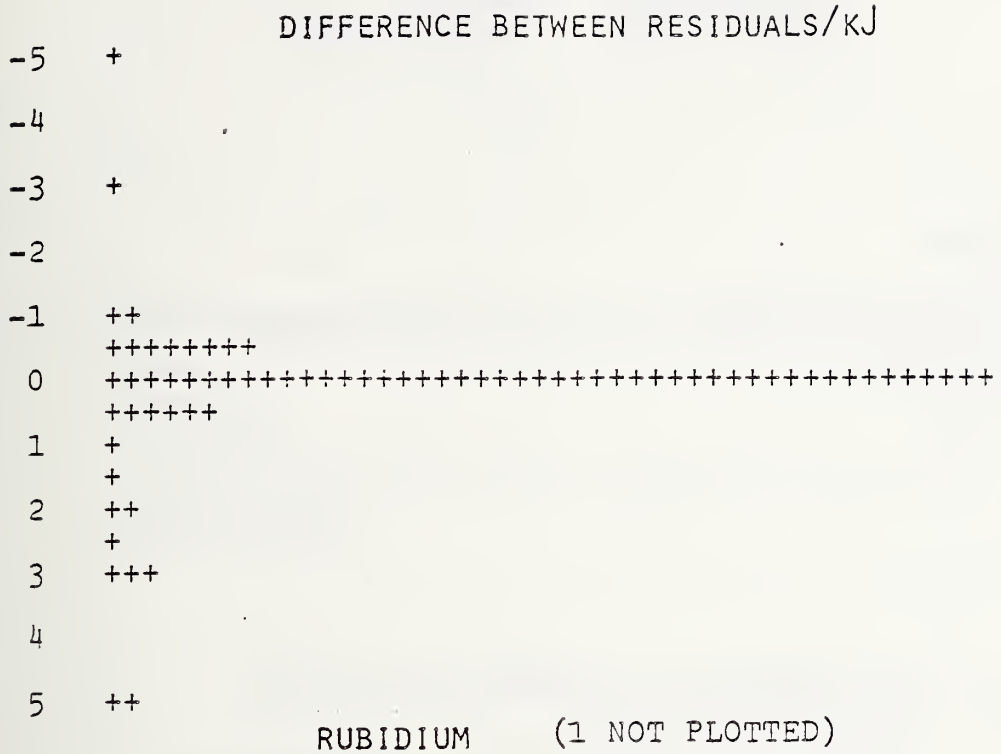
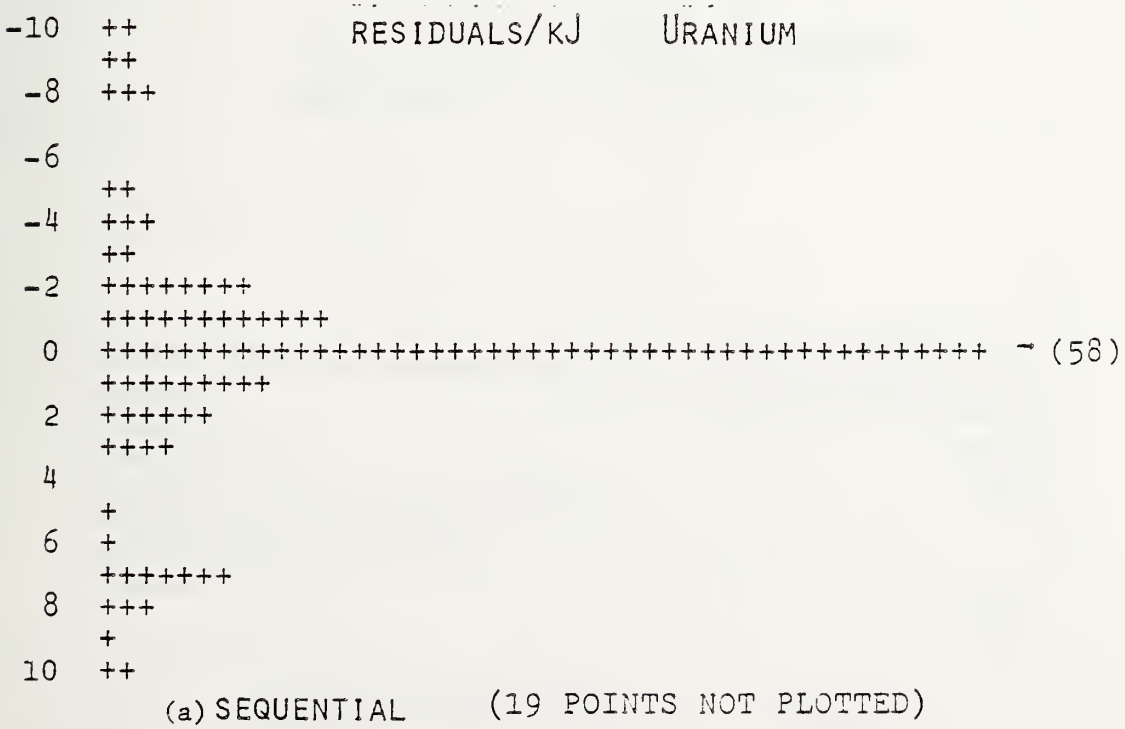
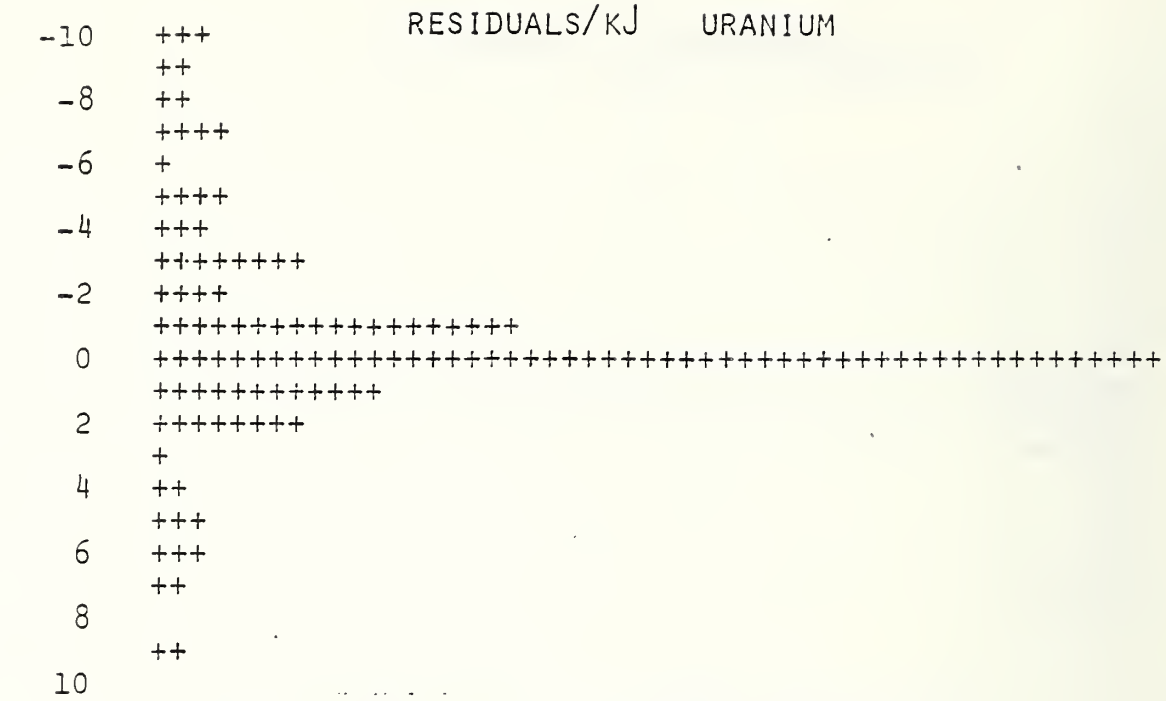
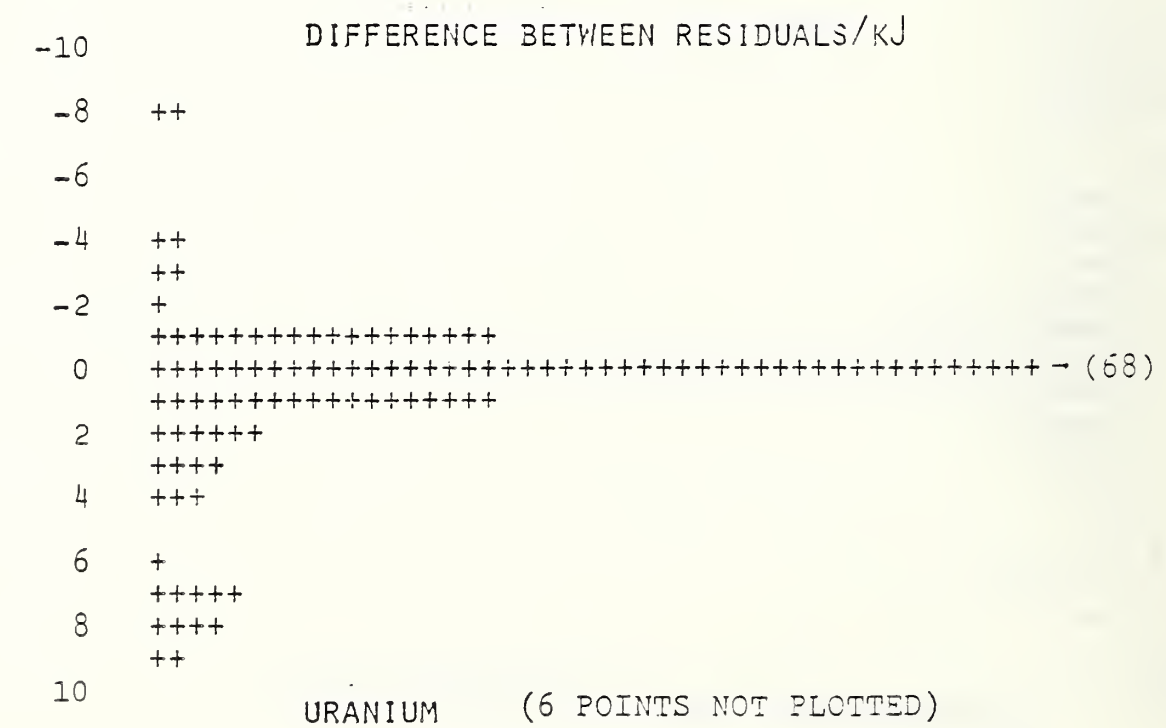


Figure 8 (a)





(b) SIMULTANEOUS (10 POINTS NOT PLOTTED)

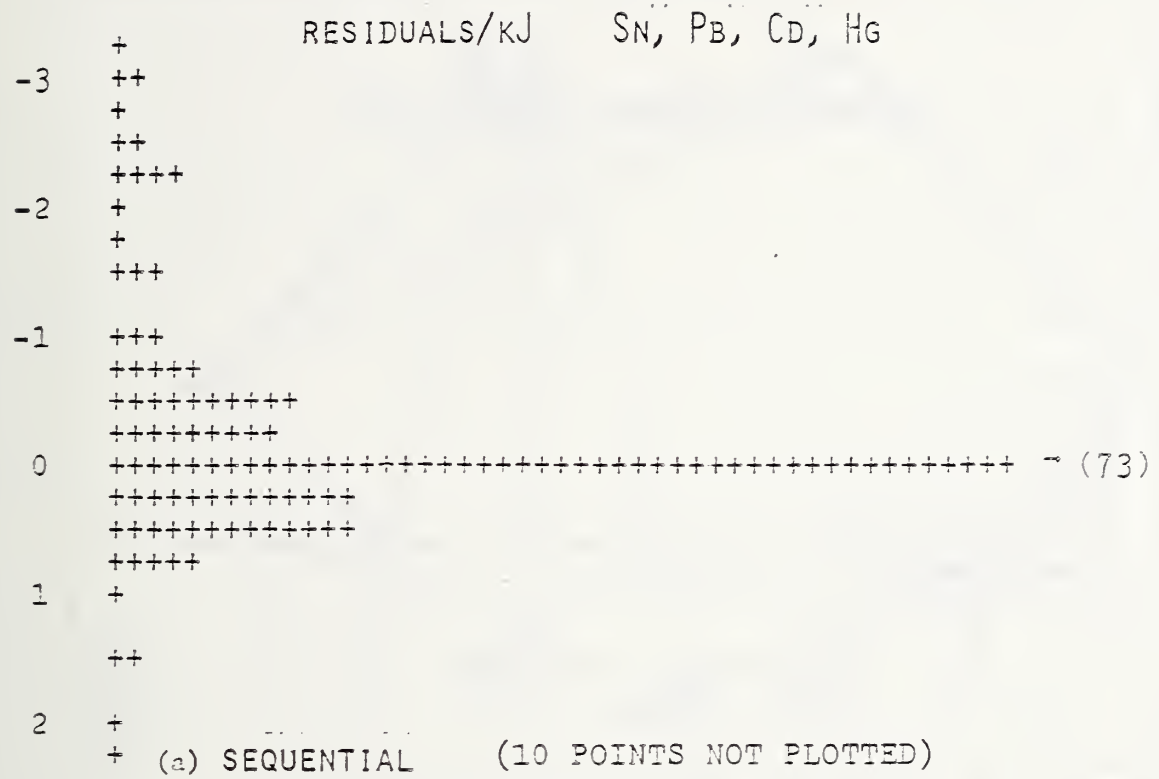


(6 POINTS NOT PLOTTED)

Figure 9

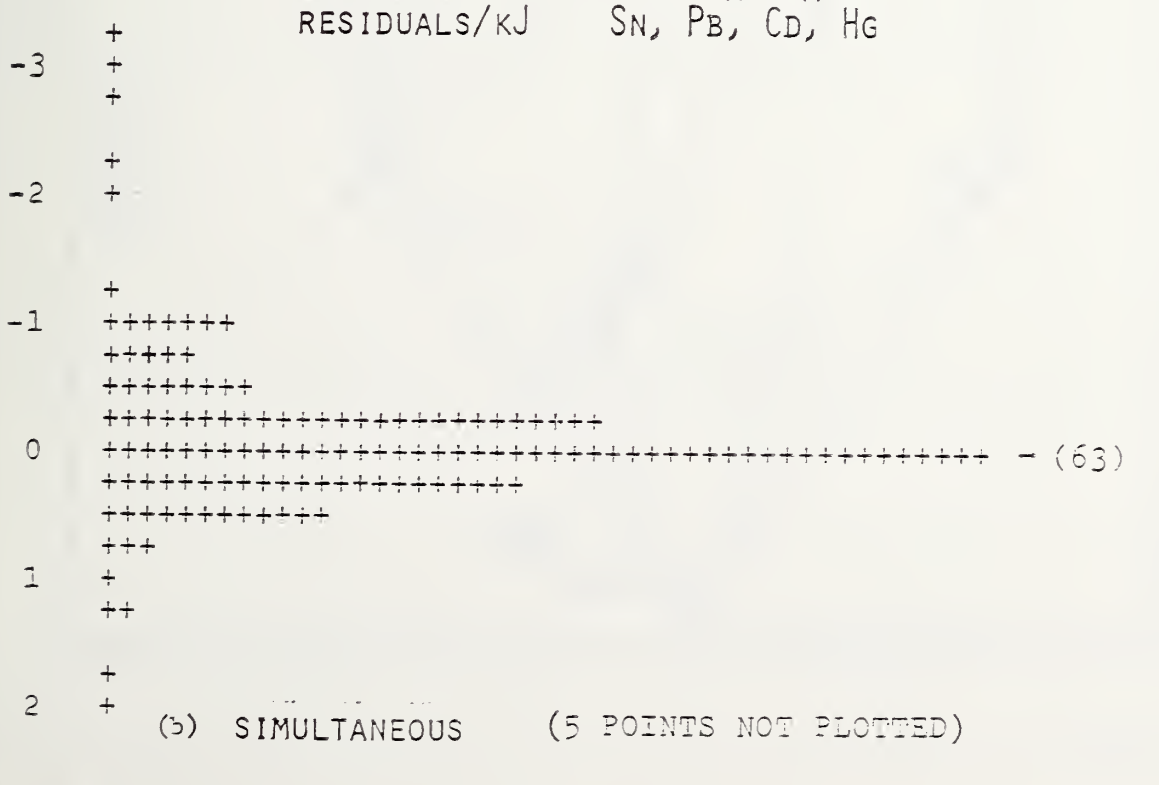
Figure 10

(a)



(a) SEQUENTIAL (10 POINTS NOT PLOTTED)

(b)



(b) SIMULTANEOUS (5 POINTS NOT PLOTTED)

Figure 11

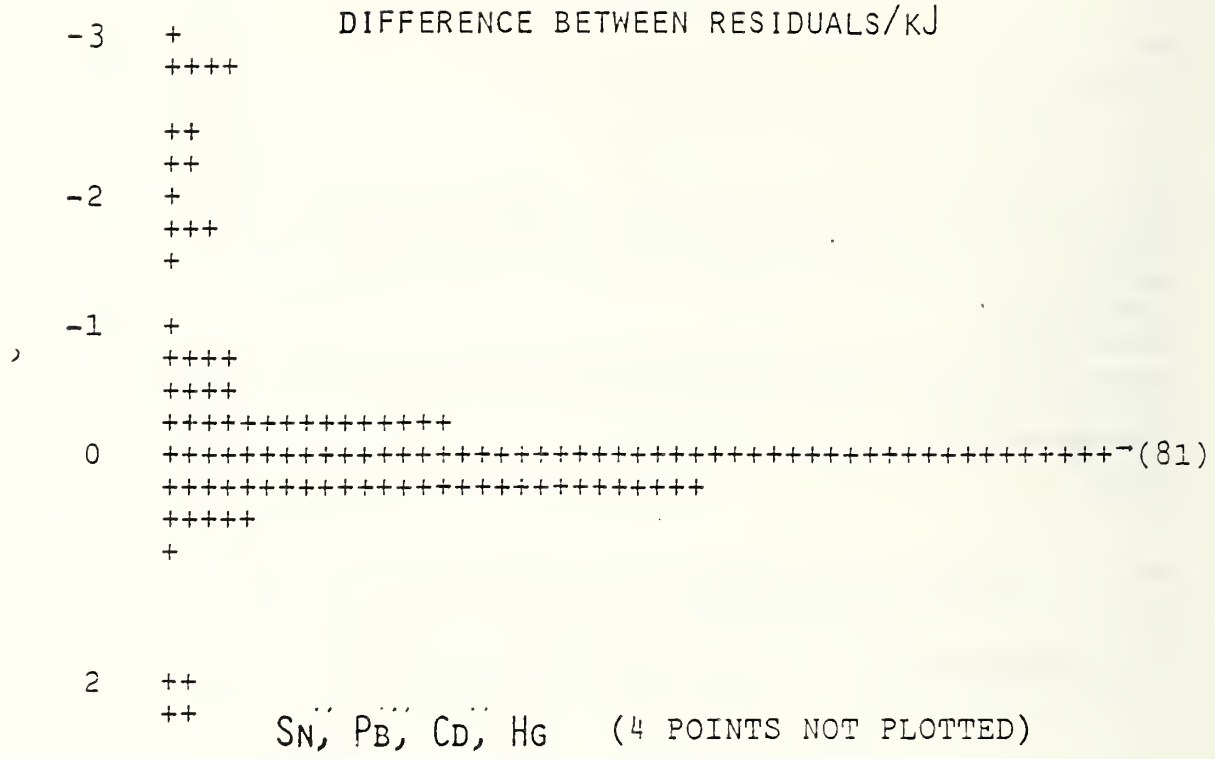
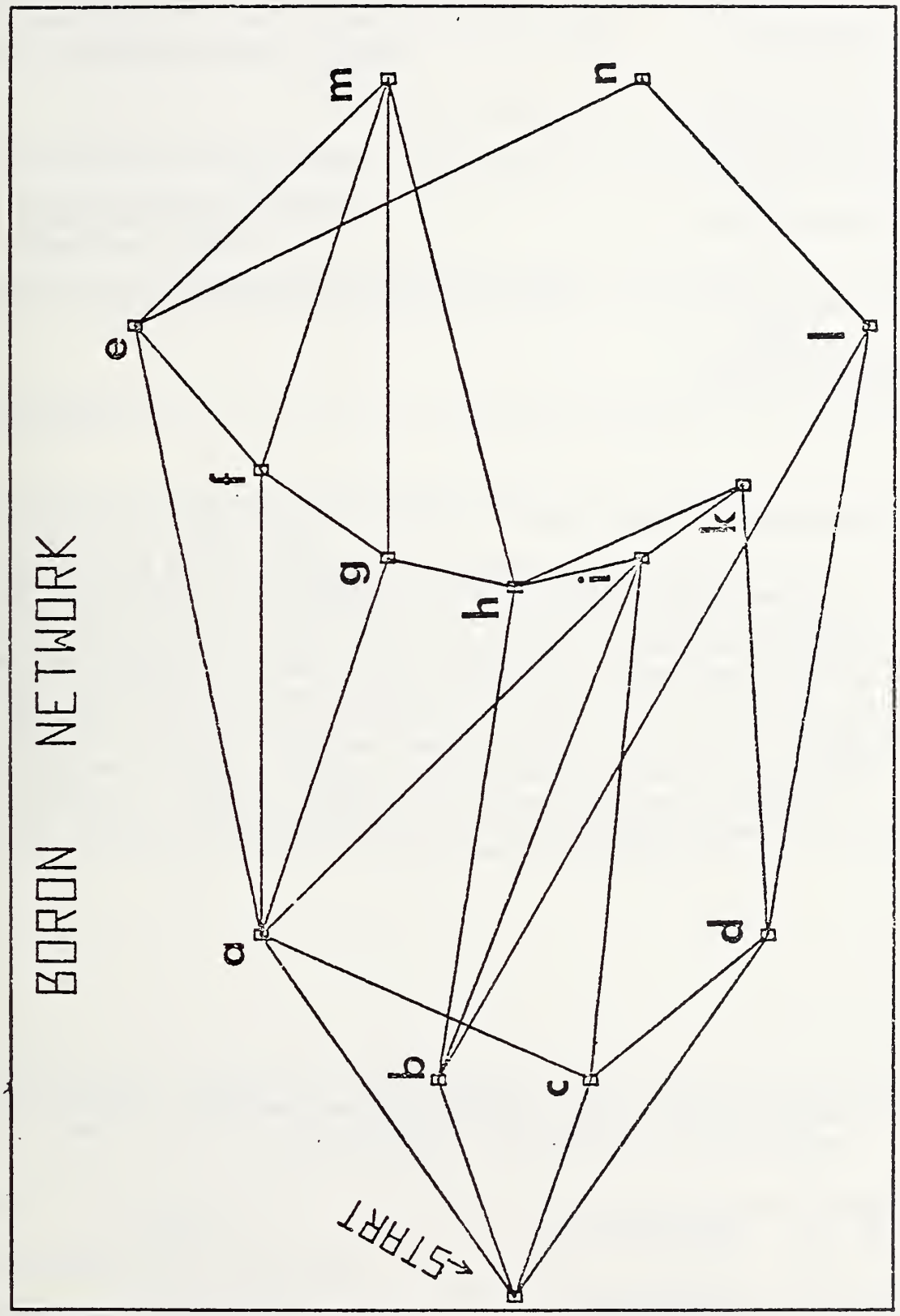

DIFFERENCE BETWEEN RESIDUALS/KJ
-3
-2
-1
0
2
(81)
SN, PB, CD, HG (4 POINTS NOT PLOTTED)

Figure 12



U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 76-1147	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

A COMBINED LEAST SUMS AND LEAST SQUARES APPROACH TO THE EVALUATION OF THERMODYNAMIC DATA NETWORKS

5. Publication Date: July 1976

6. Performing Organization Code

7. AUTHOR(S)

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)*

13. Type of Report & Period Covered: Interim

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

A description is given of a system for computer-based evaluation of interrelated thermodynamic measurements of enthalpies of reaction, equilibria and entropies. This system is an extension of the CATCH program developed by J. B. Pedley, University of Sussex. In the new system linear least sums and least squares techniques are used to solve networks of thermodynamic equations to obtain the enthalpies and free energies of formation and the entropies of chemical substances. The least sums technique is shown to be useful in assessing the consistency of the data. A method combining least sums and least squares solutions, provides a weighted solution that reproduces closely the solutions that are obtained by a detailed analysis of the data using the customary sequential procedure. The results from tests on four large networks involving compounds of B, U, Rb and salts of Sn, Pb, Cd and Hg are discussed.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)* Computerized data analysis; data evaluation; least squares (L2); Least sums (L1); thermochemistry; thermodynamic data networks

18. AVAILABILITY

[X] Unlimited

[ ] For Official Distribution. Do Not Release to NTIS

[ ] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13

[X] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151

19. SECURITY CLASS (THIS REPORT): UNCLASSIFIED

21. NO. OF PAGES: 37 41

20. SECURITY CLASS (THIS PAGE): UNCLASSIFIED

22. Price: $4.00