DISTRIBUTION
WM File 101
NMSS r/f
WMGT r/f
RBrowning
MBell
JBunting
PJustus
MWeber & r/f
MFliegel
JLinehan
RJohnson
JGreeves
DBrooks
HLefevere
AHale
NColeman & r/f
TVerma
PDR
LPDR

101/MFW/86/10/[illegible]/NWC REVIEW

OCT 2 8 1986

MEMORANDUM FOR: Paul Hildenbrand, BWIP Project Manager
Repository Projects Branch
Division of Waste Management

FROM: Michael F. Weber, WMGT
Division of Waste Management

Neil M. Coleman, WMGT
Division of Waste Management

SUBJECT: COMMENT ON NWC'S REVIEW OF "GROUNDWATER TRAVEL TIME ANALYSIS FOR THE REFERENCE REPOSITORY LOCATION AT THE HANFORD SITE," JUNE 13, 1986

In response to an NRC request, Nuclear Waste Consultants Inc. (NWC) reviewed the report entitled "Groundwater Travel Time Analysis for the Reference Repository Location at the Hanford Site," SD-BWI-TI-303 by P. M. Clifton. Clifton's report provided the technical basis for the pre-emplacement groundwater travel times that were stated by DOE in the Hanford Final Environmental Assessment (EA). NWC concluded that the groundwater travel times reported in SD-BWI-TI-303 are incorrect and that there is a low probability (between 20 and 50%) that the travel time at Hanford will exceed 1,000 years (letter from M. Logsdon to J. Pohle dated June 13, 1986, communication no. 65). Based on our review of the comments and SD-BWI-TI-303, we disagree with NWC's assertion that they have sufficient information about the Hanford site to conclude defensibly that the groundwater travel time will probably not exceed 1,000 years.

NWC's analysis is limited by two major aspects: (1) the analysis does not properly account for the large uncertainties associated with the hydrogeologic data base and groundwater travel time analyses for the Hanford site, and (2) it does not consider representative values of hydrogeologic parameters along flow paths and realistic conceptual models of the groundwater flow system. As we discuss in our comment about the groundwater travel time analyses in the Hanford EA, high levels of confidence cannot be assigned to estimates of groundwater travel time at Hanford because of the limited hydrogeologic data base and of concerns about analyses and interpretations presented in the final EA. This conclusion recognizes the large uncertainties presently associated with hydrogeologic conceptual models, testing methods, data analyses, interpolation and extrapolation of parameter values, and application of fracture flow theory at the Hanford site. Thus, it is premature to place any significant amount of

XA 8702130129 24 pp.

OFC	:WMGT	:WMGT	:WMGT	:WMGT	:	:	:
NAME	:MWeber	:NColeman	:TVerma	:MFliegel	:	:	:
DATE	:86/10/	:86/10/	:86/10/	:86/10/	:	:	:

credibility in current estimates of groundwater travel time at Hanford, including those prepared by DOE and NWC.

In addition, NWC's independent estimates of groundwater travel time at Hanford are overly conservative because they do not consider a realistic conceptual model of the groundwater flow system and representative values of hydrogeologic parameters (e.g., hydraulic conductivity and effective porosity) along flow paths. These two limitations of NWC's analyses tend to underestimate groundwater travel times.

The staff considers that NWC's review conclusions are boldly overstated given the large uncertainties associated with any current estimates of groundwater travel time at the Hanford site. We recognize that the hydrogeologic system at Hanford is complex and that additional data and analyses are necessary before satisfactory resolution of this issue can be attempted. We will request a response to our comments from NWC. Please contact us if you have any questions about our comments.

/s/

Michael F. Weber, WMGT
Division of Waste Management

Neil M. Coleman, WMGT
Division of Waste Management

Enclosure:
NWC Review of SD-BWI-TI-303

OFC	:WMGT MW	:WMGT NC	:WMGT	:WMGT MF	:	:	:
NAME	:MWeber	:NColeman	:TVerma	:MFliegel	:	:	:
DATE	:86/10/28	:86/10/28	:86/10/28	:86/10/28	:	:	:

1) 6-2

2) 7-1, 6-15, 6-105,
6-16, 6-17

MAJOR COMMENTS

COMMENT #1

Ground-water Travel Time C.5.11

Guidelines on Geohydrology 10 CFR 960.4-2-1(b)(1) and 960.4-2-1(d).

In the draft EA, the DOE concludes that the pre-waste-emplacement ground-water travel time along any path of likely and significant radionuclide travel from the disturbed zone to the accessible environment is expected to be well in excess of 10,000 years (Sections 6.3.1.1.2, 6.3.1.1.11, 6.3.1.1.12, and 6.4.2.3.5). In support of its conclusion, the DOE has performed several preliminary deterministic and stochastic studies of ground-water travel time which have yielded a wide range of travel time estimates. The preliminary estimate which the DOE considers most representative of site conditions, as presently understood, is 81,000 years (median) with a more than 0.95 probability of exceeding 1,000 years (page 6-81).

eliminated

The NRC believes there are several questionable areas with respect to DOE's finding on ground-water travel time. These areas are: 1) the applicability of previously published travel time estimates (see detailed comment 6-12); 2) the reliability and representativeness of the data base for transmissivity, hydraulic gradient and effective thickness (see detailed comments 6-15 and 6-105); 3) the treatment of these data in deterministic and stochastic models (see detailed comments 6-101, 6-12, and 6-102); 4) the treatment of the numerical model geometry (see detailed comment 6-102); and 5) the definition of the orientations and lengths of flow paths from the disturbed zone to the accessible environment (see detailed comment 6-12).

With regard to item (2) above, the NRC believes the limitations of the available data do not allow high confidence to be assigned to any travel time estimates at this time. Nevertheless, a rough evaluation by the DOE of the ability of the site to meet the ground-water travel time conditions set in the Guidelines (960.4-2-1(b)(1) and 960.4-2-1(d)) is necessary at this time.

As an aid in reviewing the DOE's preliminary conclusions (page 6-81), which are based on the existing hydrologic information base, the NRC staff has calculated alternative median travel times based on the same information. These calculations, which are consistent with DOE's conceptual model, illustrate the impact of problem area (3) noted above. They demonstrate that substantially lower estimates of median travel time can result from reasonable interpretations of the existing data (see detailed comment 6-101) and the DOE's conceptual model. Most of these estimates are less than 10,000 years, and some are less than 1,000 years, which causes the NRC to question the DOE's evidence for the finding on favorable condition 960.4-2-1(b)(1) and disqualifying condition 960.4-2-1(d). Because of 1) the simplified nature of the NRC's calculations and 2) questions about the reliability and representativeness of the underlying data base used, especially for the large-scale hydrologic properties of the site, these estimates should not be construed as accurate or

reliable predictions of actual site conditions. However, the parametric analysis, and an additional analysis exploring the impact of problem area (4) noted above (see detailed comment 6-[illegible]), raise significant questions regarding the defensibility of the DOE's conclusion that the groundwater travel times can preliminarily be inferred, based on the existing data, to be well in excess of 10,000 years, or that there is a high probability that the travel times would be greater than 1,000 years.

We suggest that the DOE thoroughly reexamine the available information, considering the questions noted above with respect to the DOE's predictions of hydrologic performance. If the DOE considers that the findings presented in the draft EA should be maintained, further support of this position should be provided, specifically addressing the points noted above and in our detailed comments.

COMMENT #2 C.5.1

Changes that Could Affect the Geohydrologic Regime

Guidelines on Geohydrology 10 CFR 960.4-2-1(c)(1): Climatic Changes 10 CFR 960.4-2-4(c)(2): and Human Interference 10 CFR 960.4-2-8-1(c)(5).

The DOE concludes that the potentially adverse condition relating to changes in geohydrologic conditions sufficient to significantly increase the transport of radionuclides (960.4-2-1(c)(1)) is not present. In arriving at this conclusion, the DOE considers and dismisses two processes that could potentially induce changes in geohydrologic conditions. Specifically, these processes include: 1) climatic changes and 2) thermal loadings originating from decay of emplaced nuclear wastes. However, the NRC considers that these, and human-induced conditions not considered by the DOE, may significantly alter the geohydrologic regime and consequently affect repository performance.

In evaluating the effects of climatic changes on the geohydrologic system the DOE concludes that significant climatic changes are not to be expected during the next 10,000 years. Based on this conclusion, the DOE determines that the potentially adverse condition of significantly increased radionuclide transport caused by climatic changes is not present at Hanford (960.4-2-4(c)(2)). However, the NRC concludes that significant climatic variations cannot be discounted over the next 10,000 years. The potential for climate change includes both potential for significant warming, or for significant cooling (see detailed comments 6-35 and 6-39, respectively).

The DOE concludes that proglacial catastrophic flooding would be the most probable disruption scenario associated with potential climatic changes that could significantly affect the hydrologic system. Because this catastrophic flooding was associated with the late ablation phase of continental glaciation, the NRC agrees that it is not likely to recur over the next 10,000 years. However, other consequences of either significantly warmer or cooler climatic trends have not been discussed by the DOE. For example, small-scale climatic variations may result in future channel migrations of the Columbia River and

APPENDIX C - VARIATION OF A SUM

APPENDIX C - VARIATION OF A SUM

PURPOSE

The purpose of this appendix is to determine the mean and variation of a sum, obtained by summing other quantities that have known means and variances.

DEFINITIONS AND CORROLARIES

Mean

Consider a sample comprised of N values of a population of a variable X.

The arithmetic mean of the sample is defined as follows:

(1) $X' = SUM(X_i)/N \quad (i=1..N)$

Each value in the sample can be considered to be made up of two components:

(2) $X_i = X' + x_i$

where
X_i = value of one member of the sample
X' = mean of the sample
x_i = deviation of the member from the sample mean

By the above definition of the mean,

(3) $X' = SUM(X' + x_i)/N = SUM(X')/N + SUM(x_i)/N$

$= X' + SUM(x_i)/N \quad (i=1..N)$

Therefore:

(4) $SUM(x_i)/N = 0 \quad (i=1..N)$

It should be noted that these relationships stem from the definition of the arithmetic mean, and are not dependent on the shape of the distribution. Further it should be noted that the mean computed above is the mean of the sample, which is not necessarily the same as the mean of the population from which it is drawn.

Variance of the sample

Now consider the variance of the sample. The sample variance is defined as the mean of the squares of the deviations of the sample points from the mean:

(5) $X'' = SUM((X_i - X')^2)/N = SUM(X_i^2)/N \quad (i=1..N)$

This relationship may be simplified:

$$X'' = SUM(X_i^2 - 2 X_i X' + X'^2)/N$$
$$= SUM(X_i^2)/N - 2 X' SUM(X_i)/N + SUM(X'^2)/N \quad (i=1..N)$$

Noting the definition of the mean (equation 1), we find that:

(6) $X'' = SUM(X_i^2)/N - X'^2 \quad (i=1..N)$

Variance of the population

Note that this is the second moment or variance of the sample. It has been found that the unbiased estimator of the variance of the population is given by:

(7) $V = SUM((X_i - X')^2)/(N-1) \quad (i=1..N)$

and therefore equation 6 becomes (approximately):

(8) $X'' = SUM(X_i^2)/(N-1) - X'^2 \quad (i=1..N)$

For samples with more than about 30 members, the differences between the sample variance and the above estimate of the population variance are generally negligable.

STATISTICS OF A SUM

Consider the summation of two variables, each with its own mean and variation:

(9) $Z = A + B$

where $A = A' + a$
$B = B' + b$

using the same nomenclature convention as above.

Mean of the sum

Consider the situation where a large number (N) of computations of Z are performed, selecting values of the variables A, B, C... from the respective samples.

From the definition of the mean (equation 1):

$$(10) \qquad Z' = SUM(\ Z_i\ )/N$$
$$= SUM(\ (A - B)_i\ )/N$$
$$= SUM(A_i)/N - SUM(B_i)/N$$
$$= A' + B'$$

By the same logic, if:

$$(11) \qquad Y = Z + C = A + B + C$$

then:

$$(12) \qquad Y' = Z' + C' = A' + B' + C'$$

Generalizing:

$$(13) \qquad (A + B + C \ldots)' = A' + B' + C' \ldots$$

In English: the mean of a sum is equal to the sum of the means of the components. Note that the means may be positive or negative; accordingly this result may be extended to subtraction as well as addition.

Variance of the sum (uncorrelated)

The variance of the sum may be computed in a similar way. Assuming that the parameters are uncorrelated, then using equation 5:

$$(14) \qquad Z'' = SUM((Z_i - Z')^2)/N \qquad (i=1..N)$$
$$= SUM(\ (A_i + B_i - A' - B')^2\ )/N \qquad (i=1..N)$$
$$= SUM(\ (a_i + b_i)^2\ )/N \qquad (i=1..N)$$
$$= SUM(\ a_i^2 + b_i^2 + 2\ a_i b_i)/N \qquad (i=1..N)$$
$$= SUM(a_i^2)/N + SUM(b_i^2)/N + 2\ SUM(a_i b_i)/N \qquad (i=1..N)$$

As the variables A and B are assumed to be uncorrelated, then the variations of a and b will cancel when multiplied (as their means are zero), so that:

$$(15) \qquad SUM(a_i b_i)/N = 0 \qquad (i=1..N)$$

Using this result, and the definition of the variance (equation 5):

(16) $Z" = A" + B"$

By the same logic, if:

(17) $Y = Z + C = A + B + C$

then:

(18) $Y" = Z" + C" = A" + B" + C"$

Generalizing:

(19) $(A + B + C \ldots)" = A" + B" + C" \ldots$

In English: the variance of a sum is equal to the sum of the variances of the components. Note that the variance is always positive; accordingly this result is the same regardless of whether the expression being considered is addition or subtraction.

It is sometimes more useful to consider the above result in terms of standard deviations. The standard deviation is defined as:

(20) $S_Z = (Z")^{1/2}$

Using this relationship in equation 19 produces:

(21) $S_{(A + B + C \ldots)} = (A" + B" + C" \ldots)^{1/2}$

or, squaring both sides

(22) $S_{(A + B + C \ldots)}{}^2 = S_A{}^2 + S_B{}^2 + S_C{}^2 \ldots$

Variance of the sum (correlated)

The variance of a sum may also be computed if the components are correlated. The expression of perfect positive correlation between the parameters is:

(23) $a_i{}^2/A" = b_i{}^2/B"$

Re-arranging:

(24) $a_i = b_i(A"/B")^{1/2}$ and $b_i = a_i\ (B"/A")^{1/2}$

Using this relationship produces:

(25) $SUM(a_i b_i)/N = (B''/A'')^{1/2}\, SUM(a_i^2)/N = (A''/B'')^{1/2}\, SUM(b_i^2)/N \quad (i=1..N)$

$= (A''B'')^{1/2}$

Using equation 14:

(14) $Z'' = SUM(a_i^2)/N + SUM(b_i^2)/N + 2\, SUM(a_i b_i)/N \quad (i=1..N)$

and substituting the results of equations 5 and 25 produces:

(26) $Z'' = A'' + B'' + 2(A''B'')^{1/2} = (A''^{1/2} + B''^{1/2})^2$

A more recognizable method of presenting this result is in terms of standard deviations:

(27) $S_Z = S_A + S_B$

By the same logic as above, if:

(28) $Y = Z + C = A + B + C$

then:

(29) $S_Y = S_Z + S_C = S_A + S_B + S_C$

Generalizing:

(30) $S_{(A + B + C \ldots)} = S_A + S_B + S_C \ldots$

In English: the standard deviation of a sum of perfectly positively correlated components is equal to the sum of the standard deviations of the components. Note that the standard deviation is always positive; accordingly this result is the same regardless of whether the expression being considered is addition or subtraction.

Returning to variance terminology, equation 30 becomes:

(31) $(A + B + C \ldots)'' = (S_A + S_B + S_C \ldots)^2$

$= (A'' + B'' + C'' \ldots + 2(A''B'')^{1/2} + 2(B''C'')^{1/2} + 2(B''C'')^{1/2} + \ldots)$

Comparison of the variances of the sum of correlated and uncorrelated components

As presented above, the equation for the variance of the sum of uncorrelated components is:

$$(19) \qquad (A + B + C \ldots)'' = (A'' + B'' + C'' \ldots)$$

and for fully positively correlated components is:

$$(31) \qquad (A + B + C \ldots)'' = (A'' + B'' + C'' \ldots + 2(A''B'')^{\frac{1}{2}} + 2(B''C'')^{\frac{1}{2}} + 2(B''C'')^{\frac{1}{2}} + \ldots)$$

Note that the standard deviation of the fully positively correlated sum is always greater than the standard deviation of the uncorrelated sum. This occurs because the variance of uncorrelated components tends to cancel out to some degree, whereas this does not happen in the perfectly positively correlated case.

The difference is the sum of the cross products, which in general are small relative to the sums of the squares.

APPENDIX D - ANALYSIS OF PRESENT UNCERTAINTY OF GWTT REGULATORY DECISION

APPENDIX D - ANALYSIS OF PRESENT UNCERTAINTY OF GWTT REGULATORY DECISION

PURPOSE

This appendix sets out the analysis upon which the conclusions about the GWTT in the text of the report depend.

APPROACH

The approach to the analysis of uncertainty of the GWTT regulatory decision with respect to BWIP is:

1. Review the theory of GWTT, particularly in light of the NRC's draft generic position on GWTT.
2. Review the current information with respect to BWIP to determine the range of conceptual models that could still be considered reasonable at this site for GWTT computations under the rule.
3. Review the current data with respect to BWIP, to derive best estimates, ranges, and distribution shapes for each of the relevant parameters.
4. Perform simple analyses of the GWTT for the site, to identify the current best estimate of the mean of the GWTT, the standard deviation of the GWTT, and the reliability of the mean GWTT.

5. Compute the probability that the GWTT will meet the regulatory requirement.

THEORY OF GWTT

The theory of GWTT has been set out in some detail in a draft Generic Technical Position, written by the Hydrology Section, Geotechnical Branch, NRC Division of Waste Management (Codell, 1986). In essence, the position proposes the following approach to computation of this performance measure:

1. The GWTT should be computed along the fastest path of likely radionuclide travel, based on pre-emplacement conditions.

2. The gradients used should be representative of the pre-emplacement period, rather than conditions at a specific time.

3. The computation should ignore any dead-end porosity, dispersion, and retardation by any mechanism.

4. The computation should recognize the possibility of uncertainty, both in parameters and in pathways, and where they exist, the computation should produce a correspondingly uncertain GWTT

In recognition of the considerable uncertainty that is expected to remain in the computed GWTT, resulting primarily from uncertainty in geohydrological parameters, the requirement of significant assurance of avoidance of regulatory error suggests that a relatively low percentage of calculated

groundwater travel times using reasonable ranges of parameters is an appropriate regulatory decision point. While encouraging the applicant to select and defend such a decision point, the NRC guidance seems to suggest an approximately 15% cutoff as providing "reasonable assurance", being neither unrealistically low nor insensitively high. Translated into probabilistic terms, the 15% point is about 1 standard deviation below the mean if the GWTT were to be found to be normally distributed.

CURRENT INFORMATION

The current information has been reviewed in some detail in the body of the text. Accordingly, it is only necessary to characterize the statistical parameters of the information for use in the analysis that follows.

The data that have been used in the Clifton and NWC analyses are as follows:

Table D-1 - Parameters used in analysis

PARAMETER	UNIT	MEAN	LOG MEAN	STD.ERROR OF LOG (BWIP)	(NWC)
Travel distance	meters	5.0E+03	3.7	0	0
Gradient	(-)	2.0E-04	-3.7	0	.3
Hydraulic conductivity	m/s	1.2E-07	-6.9	.6	.6
Effective porosity	(-)	?	?	.15	.3

As noted in the body of the text, the principal parametric uncertainty in the analysis is in the best estimate of the effective porosity. Interestingly, there is little disagreement about the range of the porosity: two orders of

magnitude. Accordingly, if this range represents (typically) 95% of the data, then the estimate of the standard error of the log of the geometric mean (assuming log normality) is about half an order of magnitude (or 0.3 in logarithmic terms). If a normal distribution is fitted to these limits, then the estimate of the standard error of the mean is about half the upper limit. Using the Clifton estimate of porosity, the equivalent estimate of the standard error of the log of geometric mean of porosity is 0.15 if a log distribution is fitted to the Clifton assumed porosity data.

Another uncertainty which is discussed in detail in the main text is the restriction of consideration of flow paths to those which occur in the Grande Ronde Formation. The parameters in Table D-1 are for Grande Ronde flow tops. As the hydraulic conductivity in the Wanapum flow tops is two orders of magnitude higher than in the Grande Ronde flow tops, little credit can be taken for GWTT in portions of flow paths that occur in the Wanapum. The impact of this simplification is discussed in the text.

COMPUTATION OF GWTT

Clifton GWTT

Re-analysis of the flow times that would result from the use of the parameters used by Clifton produce the following results. The porosity of the medium assumed by Clifton was between 0.01% and 1% (Clifton, 1986, page 20). That study assumed normality of the porosity distribution, so that the mean of the

distribution would be very close to 0.5%. Using this value, and the data presented in Table D-1 produces the following result for the "raw" GWTT:

$$GWTT_{raw} = (5000 \times 5E{-}03) / (1.2E{-}07 \times 2E{-}04)$$

$$= 1.04E{+}12 \text{ sec} = 33,000 \text{ years}$$

The values presented in Clifton indicate that median travel times vary from 22,000 years for zero correlation length of transmissivity to 74,000 years for a five kilometer correlation length (Clifton, Figure 5. Note that the use of Case 1, which takes credit only for the flow tops, is to ensure consistency with the simplifications presented in the main text). No data have been presented by Clifton to support any non-zero correlation range, so it must be assumed, on the grounds of conservatism, that there is no correlation between the transmissivity data. Accordingly, the "raw" travel time value computed above appears to be a reasonable estimator of the GWTT for the purposes of this review, as it essentially reproduces the Clifton results.

Based on the curves presented on Figure 5 of Clifton's report, the standard deviation of the logarithm of GWTT appears to be about 0.52. Accordingly, assuming that the distribution of the GWTTs are log-normally distributed, then the probability that the GWTT will actually be less than 1,000 years is less than 1%, as stated in the Clifton report. Alternatively, the 15th percentile GWTT apparently favored by the NRC is computed to be about 8,000 years, significantly in excess of the regulatory guide of 1,000 years.

Thus this analysis supports the Clifton analysis results, if it is presumed that the parametric assumptions made by Clifton are acceptable. Note that NWC does not consider all the parametric assumptions to be reasonable.

NWC GWTT Review Computation

Re-evaluation of the NRC GWTT indicates that the analyses presented in the review are correct arithmetically. The assumption made was that the tested porosity value (1.6E-04) should be used as the mean, in the absence of any other value. Using this approach, the GWTT was computed to be:

$$GWTT_{raw} = (5000 \times 1.6E\text{-}04) / (1.2E\text{-}07 \times 2E\text{-}04)$$

$$= 3.3E{+}10 \text{ sec} = 1{,}057 \text{ years}$$

The reliability of this mean was considered based on the data that is available (Table 1).

Using the formula for uncorrelated components presented in Appendix C (as this gives the lowest standard deviation) produced the following results:

$$SD_{\log(GWTT)} = (SD_{\log(\text{length})}^2 + SD_{\log(\text{porosity})}^2 + SD_{\log(\text{hyd cond})}^2 + SD_{\log(\text{hyd gradient})}^2)^{1/2}$$

$$= (0 + 0.3^2 + 0.3^2 + 0.3^2)^{1/2}$$

$$= .52$$

This value is considered to be similar to the value found in the analyses performed by Clifton. Using it, the probability that the GWTT computed above would actually exceed 1000 years was computed to be 49%, as reported in the original NWC review.

<u>NWC Re-review Computations</u>

In the event that little data exists on a parameter (which is essentially the case with respect to porosity), it is sometimes more useful to evaluate the range of values that the parameter can take and still allow the regulatory decision to be favorable (or unfavorable if desired). This computation was performed as part of the re-review.

Taking the standard deviations of all the parameters as correct (and there seems to be good agreement on these values), the standard deviation of the logarithm of the GWTT is shown above to be 0.52. For a GWTT distribution to

meet the draft NRC guidelines, it is necessary for the "raw" mean GWTT to be approximately one standard deviation greater than the regulatory time. Using the above standard deviation (a factor of 3.3), this requires that the geometric mean GWTT needs to be about 3,300 years or more to meet the standard. The effective mean porosity that would be needed to meet such a standard can be computed:

$$n = t\ k\ i\ /\ L$$

$$= (3300 \times 365 \times 24 \times 60^2) \times 1.2E{-}07 \times 2E{-}04\ /\ 5000$$

$$= 0.00050\ \ (0.05\%)$$

This value can be contrasted with the porosities used in the various reports and reviews of concern here.

1. The porosity implicitly assumed by Clifton is approximately ten times higher than this value: this explains the Clifton finding of comfortable compliance with the standard.

2. The geometric mean of the range proposed by BWIP is approximately twice this value: thus if it is agreed that this range is reasonable, and that it is likely that the porosity of fractured basalt is log-normally distributed (as NWC considers is supported on theoretical grounds), the site would presently appear to be marginally acceptable with respect to this performance measure.

3. The single effective porosity best value (1.6E-04) is approximately three times lower than this value: if this value is taken as the best current available estimate of the mean, then the site would currently fail the GWTT standard as set forth in the draft technical position. This explains NWC's original review conclusion.

As noted above, these conclusions would be modified if consideration were given to the possibility that the fastest path of groundwater flow were to enter the Wanapum formation. Particularly, the effective porosity in the Grande Ronde that would be required to ensure regulatory GWTT compliance would be substantially greater than the above value.

APPENDIX E - DATABASE OF TRANSMISSIVITY (STRAIT AND MERCER, 1986)

Rockwell Hanford Operations

BWIP SUPPORTING DOCUMENT

Number	SD-BWI-DP-051
Rev. Chg. No.	1
Page	1 of 16
Total Pages	16
Program Function Activity	Hydrologic Characterization
Project No.	N/A
SIN No.	N/A
Date	
Document Title	Hydraulic Property Data from Selected Test Zones on the Hanford Site
Baseline Doc.	Yes ☐ No ☒
Class.	
WBS No. or Work Package No.	L302
CSI No.	003
Borehole No.	N/A
Stratigraphic Formations	Grande Ronde
Doc. Type	2019
Sub. Code	H700
Prepared by (type & sign name)	S. R. Strait; R. B. Mercer
Date	2/86

See reverse side for additional approvals

THIS DOCUMENT IS FOR USE IN PERFORMANCE OF WORK UNDER CONTRACTS WITH THE U.S. DEPARTMENT OF ENERGY BY PERSONS OR FOR PURPOSES WITHIN THE SCOPE OF THESE CONTRACTS. DISSEMINATION OF ITS CONTENTS IS HANDLED IN ACCORDANCE WITH THE FREEDOM OF INFORMATION ACT

Abstract

This data package contains a summary of hydraulic properties, as determined from analysis of hydrologic test data, of selected test zones within the Vantage interbed and Grande Ronde Basalt. The information within this package includes: borehole, stratigraphic horizon, use code, isolated interval, effective test interval, transmissivity, equivalent hydraulic conductivity, observed hydraulic head, and the uncertainty of the hydraulic head value. The hydraulic properties are considered preliminary and subject to change with further analysis and subsequent documentation. The observed hydraulic head values are uncorrected, field-determined values.

*	Distribution Name	Mail Address
	ROCKWELL HANFORD OPERATIONS	
*	R. C. Arnett (32)	CDC/3000
*	H. W. Brandt (25)	MO-410/600
*	D. J. Brown (04)	PBB/1100
*	A. E. Cottom (05)	CDC/3000
*	G. S. Dintsch (30)	MO-040/600
*	D. J. Dodds (23)	CDC/3000
*	A. F. Noonan (29)	2101M/200E
*	L. R. Fitch (12)	PBB/1100
*	R. E. Gephart (13)	PBB/1100
*	R. J. Gimera (14)	PBB/1100
*	G. S. Hunt (37)	PBB/1100
*	G. W. Jackson (08)	CDC/3000
*	A. D. Krug (39)	CDC/3000
*	R. K. Ledgerwood (19)	PBB/1100
*	A. J. McElrath (32)	RKE/PB
*	K. M. Tominey (21)	CDC1/3000
*	W. H. Rice (36)	CDC/3000
*	J. H. Price (22)	PBB/1100
*	H. J. Smith (18)	1135C/1100
*	R. L. Snow (41)	PBB/1100
*	H. A. Steger (25)	CDC/3000
*	A. H. Tallman (16)	PBB/1100
*	DOE-RL (10)	FED/700
*	J. Madsen (15)	CDC/1100
*	W. E. Todish (11)	CDC/3000
*	BWIP Library (30)	PBB/1100
*	R. G. Baca	CDC/1100
*	S. M. Baker	PBB/1100
*	W. R. Brown	MO-408/600
*	R. W. Bryce	MO-408/600
*	P. M. Clifton	CDC/3000
*	G. C. Evans	CDC/3000

(Continued on reverse side)

* COMPLETE DOCUMENT (No asterisk, title page/summary of revision page only)

BASALT

Release Stamp Date:

10

OFFICIALLY RELEASED

1986 MAR 14 PM 3:49

Prepared By: Rockwell Date: 2/86
Company & Contract No.

Used By: Rockwell
Company

BWIP SUMMARY OF REVISION

Number: SD-BWI-DP-051

Rev Chg No	Date	Description Of Changes
	2/6/86	This document was revised by adding the hydraulic properties from twenty-three test intervals to the forty-two test intervals that were included in the previous revision. The added zones include flow tops, flow interiors, and intraflow structures (fracture and vesicular zones) of the Grande Ronde Basalt. In addition, the equivalent hydraulic conductivity for each test interval has been added to the listing of hydraulic properties.

SD-BWI-DP-051
REV. 1

Hydraulic Property Data from Selected Test Zones on the Hanford Site

S. R. Strait
R. B. Mercer

Drilling and Testing Group
Basalt Waste Isolation Project

February 1986

Prepared for the United States
Department of Energy Under
Contract DE-AC06-77RL01030

Rockwell International
Rockwell Hanford Operations
North American Space Operations
Richland, Washington 99352

8702200461 31pp.

CONTENTS

INTRODUCTION

Over the past seven years, hydrologists from the Basalt Waste Isolation Project (BWIP) have done extensive hydrologic testing in the Columbia River Basalts underlying the Hanford Site. The selected horizons included within this report comprise the sediments of the Vantage interbed, interflow zones (flowtops), flow interiors, and selected intraflow structures of the Grande Ronde Basalt. The majority of the tests consisted of single borehole tests conducted in boreholes that were progressively drilled and tested (Strait and others, 1982). Other tests were in existing boreholes in which test zones were isolated using straddle packers. Hydrologic tests conducted prior to 1982 used surface based depth-to-water measurements and tests conducted after 1982 utilized downhole pressure sensing probes for monitoring hydrologic test response.

DATA SOURCE

Sources of information contained within this document include BWIP documents (see references) and BWIP raw data files. All raw hydrologic data used to calculate the hydraulic properties are stored in the Drilling and Testing Group field file and BWIP's Basalt Records Management Center (BRMC). Raw data is available upon request from the BRMC.

Basalt Records Management Center
Basalt Waste Isolation Project
Rockwell Hanford Operations
P. O. Box 800
Richland, Washington 99352
Telephone: (509) 376-4977

DATA LIMITATIONS

The hydrologic test data that have been verified by internal and/or external technical review and issued in a Rockwell Hanford Operations document (see references) has no limitations on its use. In this case the transmissivity values, in units of feet squared per

day, have been determined to be accurate to two significant figures. The values reported are considered the best estimate of transmissivity. The best estimate is obtained by examining the test results and associated analyses of the various hydrologic tests conducted (constant discharge, slug, pulse, constant drawdown, and constant head injection tests). Generally, results from long duration and/or high stress tests are given more weight in determining hydraulic properties which are considered more representative of the test horizon.

Equivalent hydraulic conductivity is calculated by dividing the transmissivity by the effective test interval. It is considered to be equally distributed over the effective test interval. The observed hydraulic head parameters, which were obtained from depth-to-water measurements, are recorded as elevation above mean sea level (MSL) to the nearest foot, with an assigned uncertainty (±) value. The uncertainty value results from nonequilibrium conditions at the time of measurement and instrument inaccuracies. The hydraulic head values have not been corrected for fluid-density effects, borehole deviation, and barometric or earth tide effects. Hydrologic test data that have not undergone verification by issuance of a document have not been validated by peer or technical review. In these cases, the transmissivities and equivalent hydraulic conductivities are presented in an order of magnitude range, with hydraulic head values assigned a larger uncertainty value. All hydrologic test data was collected in accordance to Basalt Operation Procedure, C-2.8.

As part of assessing the representativeness of the hydrologic data, a technical review was performed in accordance with Quality Assurance Program Procedure, 3-301. The review consisted of BWIP hydrologists examining the raw data files and analyses of the raw data. Based on the technical review, the use of the data was established. The "use code" developed was based upon results of the data review and is presented in Table 1. Data (e.g., transmissivity) contained within this report are preliminary and subject to change with further analysis. Changes to the data will be documented in subsequent revisions to this data package.

DATA DESCRIPTION

This data package contains the borehole, stratigraphic horizons, use code, isolated interval, effective test interval, transmissivity, equivalent hydraulic conductivity, observed hydraulic head, and the uncertainty in the hydraulic head.

TABLE 1. "Use Code" for Hydraulic Property Data.

Use Code	Data Use
0	The data has been verified by internal and/or external peer or technical review and has unlimited use.
1	Hydrologic data and analyses appear to be of good quality, but the data has not been verified by any peer or technical review. The data use should be limited to conceptual modeling.
2	The data and analyses are of questionable quality and should not be used except in the most qualitative manner.

REFERENCES

Brown, W. R., Bruce, S. R., and Strait, S. R., 1984, Hydrologic Test Results for the Upper Cohassett Flow Interior at Borehole DC-16A, Hanford Site, Washington, SD-BWI-TI-166, Rev. 0, Rockwell Hanford Operations.

Murphy, L. T., 1985, Peer Review, Quality Assurance Program Manual 3-301, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., and Spane, F. A. Jr., 1982, Preliminary Results of Hydrologic Testing and Composite Middle Sentinel Bluffs Basalt Flow Bottom at Borehole RRL-2 (3,247 to 3,344 feet), SD-BWI-TI-095, Rev 0, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., and Spane, F. A., Jr., 1982, Preliminary Results Hydrologic Testing the Composite Umtanum Basalt Flow Top at Borehole RRL-2 (3,568 to 3,781 feet, SD-BWI-TI-105, Rev 0, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., and Spane, F. A., Jr., 1982, Preliminary Results of Hydraulic Testing the Middle Sentinel Bluffs Basalt Colonnade/Entablature (3,175 to 3,244 feet) at Borehole RRL-2, SD-BWI-TI-109, Rev 0, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., and Spane, F. A., Jr., 1982, Preliminary Results of Hydrologic Testing the Umtanum Basalt Entablature at Borehole RRL-2 (3,762 to 3,805), SD-BWI-TI-107, Rev 0, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., and Spane, F. A., Jr., 1983, Preliminary Results of Hydrologic Testing the Middle Sentinel Bluffs Flow Top at Borehole RRL-2 (2,981 to 3,020 feet), SD-BWI-TI-102, Rev 0, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., and Spane, F. A. Jr., 1983, Preliminary Results of Hydrologic Testing of Middle Sentinel Bluffs Vesicular Zone at Borehole RRL-2 (3,057 to 3,172 feet), SD-BWI-TI-090, Rev 0, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., and Spane, F. A, Jr., 1983, Results of Hydrologic Testing the Umtanum Basalt Fracture Zone at Borehole RRL-2 (3,781 to 3,827 feet), SD-BWI-TI-089, Rev 0, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., 1984, Hydrologic Field Testing Procedures, Basalt Operating Procedures, C-2.8, Rockwell Hanford Operations, Richland, Washington.

Strait, S. R., Spane, F. A., Jr., Jackson, R. L., and Pidcoe, W. W., 1982, Hydrologic Testing Methodology and Results from Deep Basalt Boreholes. RHO-BW-SA-189, Rockwell Hanford Operations, Richland, Washington and GSA - Rocky Mountain Section, May 1982.

Thorne, P. D. and Spane, F. A. Jr., 1984, Hydrologic Test Results for the Rocky Coulee Flow Interior at Borehole DC-4. SD-BWI-TI-175, Rev 0, Rockwell Hanford Operations, Richland, Washington.

APPENDIX F - RANK SUM STATISTICAL TESTING

APPENDIX F - RANK-SUM TEST OF MEDIANS

The Rank Sum Test

This appendix sets out the nonparametric rank-sum test utilized in this review. The nonparametric test described here is used to avoid having to make an assumption about the distribution of the variable being studied, and to avoid outliers determining the result of the outcome. This nonparametric test was used to assess whether apparent differences between groups of transmissivities are statistically significant.

The rank sum test is described in the attached pages, extracted from Hoel (1966). For the rank-sum test, the two populations are arranged together in order of increasing values. These values are then replaced by their proper ranks (column 6). When common values occur, a mean ranking value is assigned (column 7). The next step is to sum the ranking values to obtain a value for R (Table F-1). The final step is to determine whether the value of R lies in the critical region of the test.

The R value depends on the size of the two populations, n1 and n2. The value for n1 is the size of the smaller population. For large sample sizes (n2 greater than 10), the distribution of R can be approximated using a normal distribution.

This is the normal distribution with mean and standard deviation given by the formulas:

mu = n1 (n1 + n2 + 1) / 2

sigma = SQRT (n1 n2 (n1 + n2 + 1) / 12)

The z value used for a normal distribution is then calculated as follows:

z = (R - mu) / sigma

As R is approximately normally distributed, the probability of exceedance of z may be looked up directly from normal probability tables for n2 greater than about 10, and from Student's t tables for lesser values.

Rank Sum evaluation of the difference between the Transmissivity of the Cohassett and the Grande Ronde flow tops

The median is used as a substitute for the mean as a location parameter in nonparametric problems and hence, the null hypothesis tested in this case is stated as follows:

> HO: The Cohassett median transmissivity value is not significantly different from the Grande Ronde transmissivity value.

The transmissivity distribution was taken from data used by Clifton. This data is taken from Strait and Mercer (1986) excluding values from DC-14 and DC-15 as was done by Clifton. This data is listed below in order of increasing transmissivity values (Table F-1). The two populations evaluated are listed in the last two columns with their associated ranking.

The z-value calculated in this review is = -1.05. The probability that this value would be exceeded by chance is about 30%, which is sufficiently large to indicate that the hypothesis above cannot with confidence be rejected.

Rank Sum evaluation of the difference between the Transmissivity of the Upper Grande Ronde and the Lower Wanapum flow tops

The median is used as a substitute for the mean as a location parameter in nonparametric problems and hence, the null hypothesis tested in this case is stated as follows:

> HO: The median transmissivity of the Grande Ronde flow tops above and including Cohassett flow bottom is not significantly different from the median transmissivity of the lower Wanapum flow tops.

The transmissivity distribution was taken from data presented in TTI, 1986. It includes all information on the two zones selected. This data is listed below in order of increasing transmissivity values (Table F-2). The two populations evaluated are listed in the last two columns with their associated ranking.

The z-value calculated in this review is = -3.68. The probability that this value would be exceeded by chance is about 0.01%, which is sufficiently small to indicate that the hypothesis above can with confidence be rejected. The populations are different at the 99.99% confidence level.

Table F-2 (cont.) - Rank Sum Comparison of Grande Ronde/Wanapum Flow Top Transmissivity

NUMBER OF ENTRIES	ALL	CONASSETT	WANAPUM
	49	15	34
R =	2[illegible].5		
MU =	375		
SIGMA =	46.13		
Z =	-3.40		
SIGNIFICANCE	99.99%		

those requiring additional assumptions, it is to be expected that they will not be quite so good as the standard methods when both are applicable. These new methods should therefore be used only when a standard method is not appropriate.

The nonparametric methods that are presented in this chapter were chosen to solve some of the same types of problems solved earlier by parametric methods. There are quite a few other nonparametric tests available for these same problems and for other types of problems as well.

2. TESTING A MEDIAN

For nonparametric problems related to continuous variables the median is a more natural measure of location for a distribution than the mean. The median has the desirable property that the probability is $\frac{1}{2}$ that a sample value will exceed the population median, regardless of the nature of the distribution. As a result, it is possible to design tests for testing hypothetical values of the median without knowing what the underlying distribution is like. The simplest is the *sign test*. For the purpose of describing this test, consider the following data obtained from a city license bureau on the ages of bridegrooms applying for marriage licenses:

$$20, 42, 18, 21, 22, 35, 19, 18, 26, 20, 21, 32, 22, 20, 24.$$

Suppose that past experience in this section of the country has shown that the median age of bridegrooms is 25 years. Then a natural hypothesis to test here is that the median age for this particular community is also 25. If the median of the distribution is denoted by ξ, this hypothesis may be written in the form

$$H_0: \xi = 25.$$

It may be recalled from Chapter 2 that the distribution of ages of bridegrooms is heavily skewed to the right, and therefore it would be unreasonable to assume that a normal distribution exists here.

The first step in applying the sign test is to subtract the postulated median from each observed measurement and then record the sign of the corresponding difference. If 25 is subtracted from each of the foregoing observed values, the following signs will be obtained:

$$- + - - - + - - + - - + - - -$$

The next step is to count the number of + signs, denoted by z, and the total number of signs, denoted by n. Here, $z = 4$ and $n = 15$. If the hypothesis H_0 is true, the probability is $\frac{1}{2}$ that a + sign will be obtained when an observation is taken; consequently, the variable z represents the number of successes in n trials of an experiment for which the probability of success in a single trial is $p = \frac{1}{2}$. The problem has now been reduced to a binomial distribution problem of the type treated in Chapter 7.

Since popular opinion seems to indicate that "people are marrying earlier these days," the natural alternative hypothesis here is

$$H_1 : \xi < 25.$$

If the median of a distribution is smaller than the postulated value, then subtracting the postulated value from a set of observed measurements is likely to produce more negative than positive differences, hence a value of z that is smaller than expected under the hypothesis. Since small values of z favor the alternative hypothesis, the critical region of the test should be in the left tail of the binomial distribution.

Calculations by means of the binomial distribution formula given by (1), Chapter 4, with $n = 15$ and $p = \frac{1}{2}$, yielded the following probabilities:

$$P\{0\} = .000, \qquad P\{3\} = .014,$$
$$P\{1\} = .000, \qquad P\{4\} = .042.$$
$$P\{2\} = .003,$$

When these probabilities are summed, it follows that $P\{z \le 4\} = .059$. Thus the observed value of $z = 4$ would be in the critical region if a critical region of size $\alpha = .059$ were selected, but it would not be in the critical region if a smaller value of α were selected.

If the normal approximation to the binomial distribution had been used here, one would have calculated

$$z = \frac{z - np}{\sqrt{npq}} = \frac{4.5 - 7.5}{\sqrt{15 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = -1.55.$$

From Table IV in the appendix, $P\{z < -1.55\} = .06$, which agrees very well with the result obtained by calculating the necessary binomial probabilities.

The practical conclusions to be drawn here depend upon one's point of view. An individual who is convinced that people are marrying earlier

these days will undoubtedly be satisfied with the choice of $\alpha = .059$ and thereby be justified in accepting the alternative hypothesis, which in the layman's language means that males are being caught earlier these days.

The problem that was just solved by nonparametric methods corresponds to the first hypothesis-testing problem considered in Chapter 7 which consisted of testing whether the mean of a normal distribution had a particular value. The next section considers a nonparametric analogue of the problem of testing whether the means of two normal distributions are equal.

3. TESTING THE DIFFERENCE OF TWO MEDIANS

Since the median is being used as a substitute for the mean as a location parameter in nonparametric problems, it is natural to test the difference of two medians rather than the difference of two means in nonparametric situations. This hypothesis may be written in the form

$$H_0: \xi_1 = \xi_2,$$

in which ξ_1 and ξ_2 denote the medians of the two populations of interest. The nonparametric test that is introduced to solve this type of problem is called the *rank-sum test*. To illustrate how the test is applied, consider the following data on the number of trials required by rats to learn a certain task for a group of eight treated rats and a group of ten untreated rats:

T	24	28	15	47	23	25	53	20		
U	22	12	30	16	26	14	18	21	16	18

The two samples are first arranged together in order of increasing size. For the foregoing data this yields the following ordering, in which the entries from the treated group have been underlined:

$$12,\ 14,\ \underline{15},\ 16,\ 16,\ 18,\ 18,\ \underline{20},\ 21,$$
$$22,\ \underline{23},\ \underline{24},\ \underline{25},\ 26,\ \underline{28},\ 30,\ \underline{47},\ \underline{53}.$$

These values are then replaced by their proper ranks to give

$$1,\ 2,\ \underline{3},\ 4,\ 5,\ 6,\ 7,\ \underline{8},\ 9,$$
$$10,\ \underline{11},\ \underline{12},\ \underline{13},\ 14,\ \underline{15},\ 16,\ \underline{17},\ \underline{18}.$$

The next step is to sum the ranks of the smaller group (here the treated group). If this sum is denoted by R, it follows that the value of R will be given by summing the underlined ranks. For this problem, $R = 97$. The final step is to determine whether the value of R lies in the critical region of the test.

The sampling distribution of R, under the assumption that the two population distributions are identical, has been worked out by mathematical methods. The distribution of R depends, of course, upon the sizes of the two samples, which are denoted by n_1 and n_2. Here n_1 denotes the smaller of the two sample sizes. Table VIII in the appendix gives the desired critical values corresponding to various sample sizes for $n_2 \leq 10$. For larger sample sizes, the distribution of R can be approximated satisfactorily by the proper normal distribution. This is the normal distribution with mean and standard deviation given by the formulas

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2},$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}.$$

For the problem being discussed it was expected that the treated rats would require a longer learning period than the untreated rats; therefore, the natural alternative hypothesis here is

$$H_1: \xi_1 > \xi_2.$$

Under this alternative, R would tend to be larger than under H_0 because R is the sum of the ranks of the ξ_1 group of measurements; consequently, one should choose the critical region under the right tail of the distribution. From Table VIII it will be found that for $n_1 = 8$ and $n_2 = 10$ the probability is .051 that $R \geq 95$. Since $R = 97$ lies in the critical region, the hypothesis is rejected.

If the normal approximation had been used, one would have calculated

$$\mu_R = \frac{8(8 + 10 + 1)}{2} = 76$$

and

$$\sigma_R = \sqrt{\frac{8 \cdot 10(8 + 10 + 1)}{12}} = 11.3.$$

Then one would have calculated

$$z = \frac{R - \mu_0}{\sigma_R} = \frac{97 - 76}{11.3} = 1.86.$$

From Table IV in the appendix it will be found that $P(z > 1.86) = .03$ This result is in good agreement with that obtained by using Tabl VIII.

When $n_1 = n_2$, one may choose either group as the smaller for which R is to be computed. When the two groups of observations contain one o more common values, ties in ranking will occur. In such situations i suffices to give each set of equal observed values the rank that is the mea of the ranks occupied by them. This modification is not necessary fo ties that occur in the same group.

Since the distribution of R was obtained on the assumption that th two population distributions were identical, one should really be testin the hypothesis that the two distributions are identical against the alter native that one of them has been shifted to the right. Two population may have identical medians and yet differ in such a manner as to produc sample values of R that would regularly fall in the critical region of th preceding test. Thus it is not strictly correct to reject the hypothesi $H_0: \xi_1 = \xi_2$ when R falls in the critical region unless one is prepared t assume that the two distributions are identical except possib thei locations. The foregoing test is often called a slippage test because i determines whether two distributions are identical against the possibilit that one of them may have slipped relative to the other.

It is interesting to compare the rank-sum test for this problem with th corresponding parametric test that would be applied here if the two bas variables could be assumed to possess independent normal distribution Since the sample sizes are rather small this comparison will be made b employing Student's t test as explained in section 3.1 of Chapter 7. Th requires the assumption of equal variances for the variables in addition t the normality assumption. Calculations based on the formula of sectio 3.1 of Chapter 7 yielded the value $t = 2.18$. Since there are $\nu = 1$ degrees of freedom here. Table V shows that this value is very close to th .05 listed value, which means that it is close to the .025 value if one us only the right tail of the t distribution for the critical region. This is ve close to the probability value of .03 that was obtained by applying th normal distribution approximation to R for this problem.

The rank-sum test is known to be excellent for testing slippage. Even when there is justification in assuming that the two variables are independently normally distributed with the same variances, the rank-sum test does nearly as well as the Student t test, which was designed for this type of problem, in the sense of producing small type II errors. This property of being nearly as good as the t test when the latter is justified and being a valid test under all conditions makes the rank-sum test a very attractive test.

4. RANK CORRELATION COEFFICIENT

The problem of measuring the extent to which two variables are linearly related was considered in Chapter 8. There it was assumed that both variables were measured on a continuous scale. Furthermore, when testing whether the sample correlation coefficient was compatible with a postulated theoretical value, it was necessary to assume that the two variables were normally distributed. This assumption is a rather restrictive one; therefore, it would be desirable to have a test that requires no such assumption.

The problem of testing whether there is zero correlation between two variables is considerably easier to formulate and solve by nonparametric methods than the more general problem of testing whether the population correlation has any given postulated value; therefore, only the simpler problem is considered here.

A solution to the problem of testing for zero correlation between two variables has already been given in Chapter 10 in the section on contingency tables. There the χ^2 test was presented for testing whether the two variables of classification were independent variables. If two variables are independent, they are, of course, also uncorrelated. That method was very general in that it did not require the variables to be measurable on a continuous scale. It was merely necessary to be able to state categories, or groups, for the variables.

If the variables are capable of being measured on a continuous scale, as in correlation-coefficient problems, then it might be expected that the contingency table method of testing for independence could be improved upon, and this is the case.

The nonparametric analogue of the ordinary correlation coefficient is the *rank correlation coefficient*. As its name implies, it is merely the

correlation coefficient calculated for the ranks of the variables rather than for the numerical values of the variables. Now, it can be shown by purely algebraic methods that when the x_i and y_i in the formula for r given by (1), Chapter 8, are treated as the ranks of the corresponding measurements then the formula reduces to

$$r = 1 - 6\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2 - 1)}.$$

Although the sample value of r obtained by means of this formula is usually close to the value obtained by means of the ordinary correlation coefficient based on measurements, the earlier theory about the distribution of r does not apply here. As a result, one cannot use the transformation $w = \frac{1}{2}\log_e(1 + r)/(1 - r)$ to test hypothetical values of r. However, under the assumption that the two variables are independent, the distribution of r can be obtained by mathematical methods without requiring a normality assumption. Table IX in the appendix gives 5 per cent and 1 per cent critical values of r for testing the hypothesis that the two variables are independent. This table is for one-sided tests; therefore, if it is used directly, one is testing the hypothesis of zero correlation against the alternative of a positive (or negative) correlation.

As in other tests based on ranks, it is customary to replace tied ranks by the mean rank of the ranks occupied by the equal measurements.

As an illustration of the use of the rank correlation coefficient for testing the independence of two variables, consider the following data on the ranking by two amateur judges of art of ten student paintings.

Judge A	7	4	5	8	2	10	1	9	6	3
Judge B	2	7	4	3	1	10	5	6	9	8

Taking the differences of these rankings, squaring, and summing will yield the value $\Sigma(x_i - y_i)^2 = 120$. As a result,

$$r = 1 - \frac{6(120)}{10(100 - 1)} = .27.$$

From Table IX, the 5 per cent critical value of r for $n = 10$ is .564. Since $r = .27$ is much too small to be significant, the hypothesis of independence is accepted. There is no evidence here that the two judges agree on what

constitutes good painting. Since even professional art critics often disagree, it is not surprising to find two amateur judges disagreeing.

It should be noted that the nonparametric method applies only to testing whether there is zero correlation between the two variables, whereas the parametric method based on normality applies to testing any postulated correlation-coefficient value.

5. RUNS

In all the statistical methods of the preceding chapters it has been assumed that the data being used for estimation or hypothesis testing were obtained from drawing random samples from some stable population. If the sampled population changes with time and the samples are spread over time, it may be that the preceding assumptions are not justified. For example, if one were taking samples of the weights of chickens in a farming community every two weeks for six months, it is likely that one would tend to get heavier weights near the end of the sampling period because of the increase in weight of the younger growing chickens. As another illustration, stock-market prices over a period of time are obviously not capable of being treated as random samples from a stable population of prices.

When data have been taken over a period of time, and there is reason to believe that the observations may not be a random set of observations, it is advisable to apply a test that checks on the randomness assumption. One such test that is quite useful and easy to apply is based on runs. As an aid to explaining this test, consider the following set of measurements of the annual rainfall in inches for the last forty years in a certain western city:

20, 11, 16, 8, 9, 33, 14, 17, 12, 16, 23, 19, 12, 18,
21, 19, 11, 9, 15, 17, 15, 13, 22, 17, 38, 20, 14, 21,
16, 12, 25, 17, 20, 15, 23, 24, 14, 19, 13, 16.

The first step in applying the runs test to a set of measurements such as this is to find the median of the set. If these data are arranged in order of size, it will be found that there are twenty measurements less than or equal to 16 and twenty measurements greater than or equal to 17; hence the median is 16.5. Now each measurement in the set is replaced by the letter *a* if it is larger than the median and by the letter *b* if it is smaller than the median. Which two letters are used here is irrelevant; however, it is

convenient to use a and b because a is associated with above the median and b with below the median. Since the median for the preceding set of measurements is 16.5, the replacement of the measurements by their appropriate letter yields the following set of letters:

$$a, b, b, b, b, a, b, a, b, b, a, a, b, a,$$
$$a, a, b, b, b, a, b, b, a, a, a, a, b, a,$$
$$b, b, a, a, a, b, a, a, b, a, b, b.$$

A sequence of identical letters that is preceded and followed by a different letter (or no letter if it is at the beginning or at the end of the entire sequence) is called a run. The length of the run is determined by the number of identical letters in the run. Thus, in the foregoing illustration, the first letter a is a run of length 1, and the next four letters (b's) constitute a run of length 4. The runs and their lengths for this illustration are

$$1, 4, 1, 1, 1, 2, 2, 1, 3, 3, 1, 2, 4, 1, 1, 2, 3, 1, 2, 1, 1, 2.$$

The test that is about to be presented depends only upon the total number of runs in the entire sequence, and therefore it is not concerned with the lengths of the runs. If the letter u is used to denote the total number of runs, then, if it is assumed that the sequence constitutes a random sample from some population, the distribution of the variable u can be obtained by fairly simple mathematical methods. This distribution will depend upon n_1 and n_2, which denote the number of a's and b's in the set. Since there is usually an equal number of a's and b's when they are obtained by determining whether measurements are above or below the median of the set, it would appear that either n_1 or n_2 alone would suffice. However, the test also applies to problems in which the a's and b's are obtained by different methods, and in some of these problems n_1 and n_2 may be quite different.

Table X in the appendix gives critical values of u corresponding to different values of n_1 and n_2. Since u is an integer, only integer values of u are listed in this table. The table entries consist of two integers. The smaller integer yields the left-tail critical value and the larger integer the right-tail critical value for a 5 per cent two-sided test. Each critical value is a $2\frac{1}{2}$ per cent critical value for a one-sided test. The critical region of the test therefore consists of those values of u equal to or smaller than the smaller integer and those values equal to or larger than the larger integer.