

WISE¹ Database V2.2d

Users' Guide

Tony Mount & Chris Humphrey

Environmental Research Institute of Supervising Scientist
GPO Box 461 (1-3 Briggs St)
Darwin, NT 0801, Australia

A version of this database (written in PARADOX) was originally developed by *eriss* to analyse river health assessment QA/QC audit information. This current version of the database has been developed by *eriss* for the Commonwealth Government's National River Health Program (<http://www.ea.gov.au/water/programs>), which is funded through the Natural Heritage Trust (<http://www.nht.gov.au>). The database is designed to be used in conjunction with macroinvertebrate-based, river health assessment under the Australian River Assessment Scheme (<http://ausrivas.canberra.edu.au>).

This software and related documentation do not necessarily represent the views of the Department of the Environment and Heritage or the Commonwealth Government. The Department of the Environment and Heritage, its employees and agents, do not accept any liability for any damage caused by, or any economic loss arising from, reliance on this information. It is the responsibility of the user to make their own decisions about the accuracy, currency, reliability, and correctness of the information generated by the database.

What WISE does

The Whole of Individual Sample Estimate (WISE) database is a Microsoft Access database that compares the live-sorted component of a sample (LS) with an equivalent-sized component representative of the whole sample (prior to sorting) (WSE, or 'Whole Sample Estimate'). The WISE database automatically calculates the Live-Sort/Whole of Sample Estimate (LS/WSE) taxa number ratio and various Bray-Curtis dissimilarity indices – as described below.

The WISE database has been developed to: (i) enable users to assess their performance in recovering small/cryptic taxa; and (ii) provide a standardised assessment tool of operator performance, thereby enabling performance in live-sort taxa recovery over the entire NRHP program to be validly assessed.

¹ Whole of Individual Sample Estimate

The database may also be used to assess the performance of laboratory subsampling and sorting procedures. For this, the ‘live-sort’ component referred to in the descriptions below, is replaced with a laboratory subsampled and sorted component. Further details relevant to this particular application are provided in Humphrey and Thurtell (1997).

Changes in V2.2d

A series of proportional scaling-down procedures are used in the databases – as explained below (‘Calculations performed by WISE’). In the original PARADOX database, random selection of taxa with scaled counts between 0 and 1 was undertaken without accounting for the original abundances of the corresponding taxa. As shown in Humphrey and Thurtell (1997), this could lead to a slight overestimation of the number of taxa selected in the Whole Sample Estimate. In the new, current database, this problem is overcome (see Appendix 2, attached), with random selection of taxa with scaled counts between 0 and 1 now taking account of the abundance of the taxa. This is done by creating a table with these taxa repeated according to their abundance, making a random selection from this table and counting the results for each taxon.

Installation

Unzip (or copy) the files to the directory required. The database is written in Microsoft Access 97 and a copy of MS Access is required to use it. (It appears to run satisfactorily under Access 2000.)

Structure

The database contains the following tables and queries:

Table	Contents
Samples	Sample ID details (Site, Date, Habitat) and the fraction of residue subsampled.
Counts	The live sort and residue counts for each taxon for each Site, Date and Habitat.
Calc1	Contains the data and results of row calculations for the selected Site, Date, Habitat. The contents of this table are displayed in the report.
Results	Results for the calculation of Bray-Curtis dissimilarity stats for each iteration.
BrayCurtis	The Bray-Curtis results for each iteration.
ReptID	Details to be transferred to the report.
Habitats	A list of valid habitat names.
RandomisationList	List of taxa with scaled counts < 1 and with rows repeated according to pre-scaled frequency.
Errors	Table to hold validation errors.
TestOutput	Used sometimes for debugging.
twisetemp	Created and destroyed several times during the calculation. It would not normally appear in the Database Window.

Query	Result
NumerPrAbs, etc	Used in calculation of summary stats.
CountRandoms	SELECT twisetemp.Taxa AS Expr1, Count(twisetemp.Seq) AS CountOfSeq FROM twisetemp GROUP BY twisetemp.Taxa;
SortedRandList	SELECT RandomisationList.* FROM RandomisationList ORDER BY RandomisationList.Freq DESC;
qwisetemp	Created and destroyed several times during the calculation. It would not normally appear in the Database Window.

Operation

The startup screen allows you to select from existing data, to perform calculations and view the results, and to bring up a data entry screen.

Entering new data:

To enter new data, or modify existing data, click “Add/Edit Data”. This brings up a data entry screen at the current record. Records are navigated using the standard MS Access buttons. (A new record is added using the button labelled “>*” at the foot of the window.)

Data can also be obtained from linked spreadsheets – see below.

Alternatively, you can overwrite the existing “counts” and “samples” tables with imported data providing the format of the imported data is the same as the existing Access tables. (Use the “Get External Data” option in the File menu.)

Data Validation

Some automatic data validation is provided when data are entered via the data entry screen. Data which are obtained from linked spreadsheets or pasted from other sources must be validated separately using the “Validate Data” button. Since this is a little more comprehensive than the automatic validation, it is a good idea to run this on all data.

The following validation rules apply:

Site: Cannot be left blank because ‘site’ is a key field.

Date: Another key field. The format is determined by the Regional Settings in Windows.

Habitat: Must be contained in the Habitat lookup table.

Taxa: Cannot be blank.

LS, or *%Res*: Positive integer greater than 0, or blank.

Fraction: >0 and ≤100

Performing calculations on a single site:

The number of iterations is entered into the text box (located in the upper right-hand side of the screen). The default is 10.

Clicking the button “Summary Stats, Current site” causes the calculations to be performed, as detailed below, and creates the report. Progress of the calculation is indicated in the status bar at the foot of the screen.

The results, and the buttons to view or print the report do not appear until a calculation is performed. (Note that the computer must have a printer installed in order to view the report.)

The number of iterations can be set to 0 in which case only one iteration is performed, the randomisation is cancelled and the same results are obtained on each run. This repeatability aids comparison with manually calculated results.

Performing Calculations on all sites:

Clicking the button “Summary Stats, All Sites” will result in a calculation being performed for all site/date/habitat combinations. The calculations are performed in exactly the same way as for a single site for the specified number of iterations.

The report is not available using this option. The summary statistics for each site are located in the table called REPTID. This can be exported to Excel by opening the database window, clicking File | SaveAs/Export, selecting the appropriate xls file type and required directory.

Linking an Excel Spreadsheet to WISE.

It is possible to link to the data in an Excel spreadsheet without actually having to import the spreadsheet as tables. The advantage is that if the data in the spreadsheet is changed it does not have to be re-imported. The disadvantage is that more data entry is required and the program runs much more slowly.

Spreadsheet layout:

The spreadsheet must have at least the following columns:

Site, Date, Habitat, Fraction, Taxa, LS, %Res

The order of columns does not matter. Additional columns are irrelevant.

The first row must contain the column names, spelt exactly as above. Case (upper, lower etc) does not matter.

The remaining rows must contain data. There cannot be any blank rows, page breaks, column sums etc.

Every row must contain an entry for Site, Date, Habitat, Fraction and Taxa.

Step 1: Link to spreadsheet

In Access, click: File | Get External Data | Link Tables.

A Window labelled “Link” should open.

In the drop-down list labelled “Files of Type”, select xls.

Change to the directory where the spreadsheet is located and select the required spreadsheet. Click the button labelled “Link”.

Ensure “Show Worksheets” is selected and if there is more than one sheet, select the relevant sheet. (Only one sheet can be selected at a time.) Click “Next”.

Click “First Row contains column headings”. Click “Next”.

The imported table can have any name eg “Sheet1”. Click “Finish”.

Step 2: Close the 'Main' form otherwise you will not be able to rename the tables. Rename any existing tables called 'Samples' or 'Counts' to something else (eg exSamples and exCounts).

Step 3: Set up two queries on the linked spreadsheet.

Select the "Queries" tag from the database Window. (If the database window is not displayed click the appropriate icon in the toolbar).

Select "New". Select "Design View" and click "OK".

Select "Sheet1" (or whatever you called the linked spreadsheet), click "Add" and "Close".

Using the left mouse button drag and drop the field labels Site, Date, Habitat and Fraction into consecutive columns on the row labelled "Field".

Right-click anywhere in the top section of the query windows, select "Properties" and change "Unique Values" from "No" to "Yes".

Close the query Window and save it with the name "Samples".

Repeat the procedure (ie select Queries, New, select linked table, Add, Close) this time dragging and dropping the fields Site, Date, Habitat, Taxa, LS, %Res into consecutive columns in the row labelled "Field". Save this query with the name "Counts".

(The Counts table included with WISE has a column labelled "Seq" which is used to establish unique keys. It is not required when linking to a spreadsheet.)

You now have two queries, one called Samples and one called Counts which obtain their data directly from the spreadsheet. When you run the program it will do the calculations using these queries.

Open Forms | Main and perform required calculations.

Support

For problems with the existing program contact: tonym@eriss.erin.gov.au

For upgrades contact: chrish@eriss.erin.gov.au

Calculations performed by WISE

An example of the WISE Results report, is shown in Appendix 1. The broad procedures involved in the calculations, according to table columns in the WISE report, are:

1. A taxa list with relative abundances of macroinvertebrates for 2 fractions of the sample, live-sort (LS) component (col **A**) and a subsample of the preserved residue (col **B**) is compiled, the latter with total no. of macroinvertebrates \geq the LS component, but always with minimum sample size of 100 organisms.
2. A 'subsample' of organisms from LS component (cols **C & D**), equivalent in proportion to the subsampled residue, is added to the residue fraction.
3. The new combined LS and residue component (**E**) is scaled down and rounded to the sample size of the LS component for LS sample size > 100 (**F & G**), or to 100 for $LS < 100$ (**H & I**), this fraction (G or I) then being termed 'whole sample estimate' (WSE).

Proportional scaling is used between columns A & D, E & G and E & I. The rounding required within these three steps follows particular rules: (i) scaled taxa abundances ≥ 1 are rounded to the nearest whole number; and (ii) taxa fractions (< 1) are summed then, on the basis of the cumulative sum, randomly assigned a value '1' amongst the string of individuals from corresponding taxa, until N individual allocations have been made, where N is the nearest whole number of the summed fractions. A worked example of these calculations is provided in Attachment 2.

Random allocations of summed fractional occurrences to the 'rare' taxa referred to above have flow-on effects through the calculations of WSE taxa lists such that different results may occur depending upon the taxa allocation. Hence, WISE performs (by default) 10 iterations towards WSE calculation and subsequent calculation of the summary statistics; the summary statistics are averaged over the 10 iterations.

WISE Outputs

Summary statistics from WISE (Appendix 1) are provided for LS and WSE comparisons for: (a) a sample size equivalent to the LS sample size, listed under the '**Actual**' outputs (most common situation), and (b) a sample size normalised to 100 animals for situations where low live-sort sample size is encountered ($LS < 100$ animals), listed under the '**Normalised**' outputs. Under 'Actual' and 'Normalised' outputs are the following summary statistics:

- (i) Live-Sort/Whole of Sample Estimate ('**LS/WSE**') taxa number ratio;
- (ii) Bray-Curtis dissimilarity index for presence/absence ('**Pres/Abs**') or relative abundance ('**Rel. Abund.**') data. Within each of these data categories, the dissimilarity index is provided for two situations: firstly, an unadjusted value ('**Unadjusted Bray-Curtis**'), and secondly, a value where LS taxa that are unique to LS (i.e. not found in the residue) are eliminated from the ensuing calculations (i.e. prior to WSE calculation). (This latter value, '**Bray-Curtis less taxa unique to LS**', is provided because often the live-sorting has removed all large rare taxa so that it would not be expected that they would occur in the residue and WSE.)

As AUSRIVAS models are only based upon presence/absence data, relevant assessment criteria for agencies will be LS/WSE (analogous to the RIVPACS 'observed/expected' ratio) and B-C dissimilarity for presence/absence data.

The Bray-Curtis dissimilarity measure is calculated in WISE, for two samples, j and k , based on taxa 1 to N (indexed by i), as:

$$\left(\sum_i |x_{ij} - x_{ik}| \right) / \left[\sum_i (x_{ij} + x_{ik}) \right]$$

where x_{ij} is the abundance for taxon i in sample j .

References

Humphrey C & Thurtell L. 1997. External QA/QC of MRHI agency subsampling and sorting procedures. *In* Development and implementation of QA/QC protocols for sample processing components of the MRHI agency bioassessment program. *By* CL Humphrey, AW Storey & L Thurtell. Final Report to Land and Water Resources Research and Development Corporation (Reference No. ARR2), December 1997.

Microsoft Access 97. Microsoft Corporation.

WHOLE INDIVIDUAL SAMPLE ESTIMATE - RESULTS

Site: 1110017
Date: 13-Nov-95

Habitat RIFFLE

Fraction 5

Summary Statistics		
	Actual	Normalised
LS/WSE	1.09	1.24
Unadjusted Bray-Curtis:		
Rel. Abund.	0.44	0.59
Pres/Abs	0.24	0.26
Bray-Curtis less taxa unique to LS:		
Rel. Abund.	0.43	0.58
Pres/Abs	0.15	0.16

Calculations for last 10 iterations

Family	A	B	C	D	E	F	G	H	I
APTEZZZL	1		0.05	0	0.00	0.00		0.00	
BAETZZZN	1	12	0.05	0	12.00	4.86	5	2.53	3
CAENZZZN	1	6	0.05	0	6.00	2.43	2	1.27	1
CALAZZZL	2	2	0.10	0	2.00	0.81	2	0.42	1
CHIRZZZL	1	113	0.05	0	113.00	45.77	46	23.84	24
CONOZZZL		1	0.00	0	1.00	0.41	1	0.21	1
ELMIZZZA	3	10	0.15	0	10.00	4.05	4	2.11	2
ELMIZZZL	2	41	0.10	0	41.00	16.61	17	8.65	9
EMPIZZZL		2	0.00	0	2.00	0.81	1	0.42	
GRIPZZZN		2	0.00	0	2.00	0.81	1	0.42	
HBIIZZZX	6	7	0.30	1	8.00	3.24	3	1.69	2
HPHIZZZL	3	16	0.15	0	16.00	6.48	6	3.38	3
HPSYZZZL	41	93	2.05	2	95.00	38.48	38	20.04	20
HPTIZZZL	2		0.10	0	0.00	0.00		0.00	
LCERZZZL	1	2	0.05	0	2.00	0.81	1	0.42	1
LIBEZZZL	8	1	0.40	1	2.00	0.81		0.42	
LPHLZZZN	38	22	1.90	2	24.00	9.72	10	5.06	5
PALAZZZX	1		0.05	0	0.00	0.00		0.00	
PPOTZZZL	66	81	3.30	3	84.00	34.03	34	17.72	18
PSEFPZZL	3	4	0.15	0	4.00	1.62	2	0.84	1
PYRAZZZL	1	3	0.05	0	3.00	1.22	1	0.63	1
SCIRZZZL	4	7	0.20	0	7.00	2.84	3	1.48	1
SIMUZZZL	1	2	0.05	0	2.00	0.81		0.42	
STAPZZZA	1		0.05	0	0.00	0.00		0.00	
TABAZZZL		1	0.00	0	1.00	0.41		0.21	
TIPUIZZL	2		0.10	0	0.00	0.00		0.00	
UACAZZZX		24	0.00	0	24.00	9.72	10	5.06	5
UDIPZZZP		7	0.00	0	7.00	2.84	3	1.48	1
UOLIZZZX	3	5	0.15	0	5.00	2.03	2	1.05	1
UTRIZZZL		1	0.00	0	1.00	0.41		0.21	
Sum	192				474	192.00	193	100.00	100
Count	23						21		19

Legend

A: Live Sort (=LS) D: (C) rounded G: (F) rounded (=WSE)
B: Residual Fraction (=Res) E: (B) + (D) H: (E) x 100/Sum(E)
C: (A) x Fraction/100 F: (E) x Sum(A)/Sum(E) I: (H) rounded
(D), (G) and (I) are randomly selected where the corresponding unrounded values are < 1.0

Worked example of proportional scaling and rounding procedures

Table 1. Procedure used in proportional scaling and rounding of taxa abundances in WISE calculations: hypothetical example of calculations used to derive a 12% 'subsample' of agency live-sort component (columns A and C) with subsequent rounding (col D).

Family	A (original live-sort)	C (12% 'subsample' of live-sort)
A	1	0.12
B	7	0.84
C	8	0.96
D	9	1.08
E	4	0.48
F	5	0.6
G	4	0.48
H	1	0.12
I	4	0.48
J	4	0.48
K	3	0.36
L	3	0.36
M	10	1.2
N	2	0.24
O	2	0.24
P	1	0.12
Q	1	0.12
R	1	0.12
S	6	0.72
T	1	0.12
U	4	0.48
Cumulative sum for values in C < 1.0		7.44

Using the proportional scaling, rounding and randomisation rules:

- (i) Scaled taxa abundances ≥ 1 are rounded to the nearest whole number.

Thus families D and M round to 1 individual each.

- (ii) Taxa fractions (< 1) are summed then, on the basis of the cumulative sum, randomly assigned a value '1' amongst the string of individuals from corresponding taxa, until N individual allocations have been made, where N is the nearest whole number of the summed fractions.

The cumulative sum for values in column C < 1.0 is 7.44, which rounds to 7 individual allocations to be randomly assigned amongst the 62 individuals from the corresponding 19 families ($A_1, B_1 \dots B_7, C_1 \dots C_8, E_1 \dots E_4$ etc) – see following table. (Note of course that the 7 allocations will not necessarily correspond to selection of 7 different taxa. In the example below, 2 allocations are randomly assigned to Family C, so that of the original 7 allocations, only 6 families are now represented.)

**Possible outcome for column D after taxa re-allocations
(1 outcome from 10 iterations)**

Family	A (original live-sort)	C (12% 'subsample' of live-sort)	D (new rounded 12% 'subsample')
A	1	0.12	
B	7	0.84	1
C	8	0.96	2
D	9	1.08	1
E	4	0.48	
F	5	0.6	1
G	4	0.48	
H	1	0.12	
I	4	0.48	
J	4	0.48	1
K	3	0.36	
L	3	0.36	
M	10	1.2	1
N	2	0.24	
O	2	0.24	
P	1	0.12	
Q	1	0.12	
R	1	0.12	
S	6	0.72	1
T	1	0.12	
U	4	0.48	1
Total	81	9.72	9