3 SETTING UP TEST SITE DATA

3.1 What are test data?

A test data set is the physical and chemical information collected from any site that the user wishes to determine (or 'test') the physical or chemical character of. Thus, the definition of test data used here in the AUSRIVAS physical and chemical reporting software is the same as that used in the AUSRIVAS biological models.

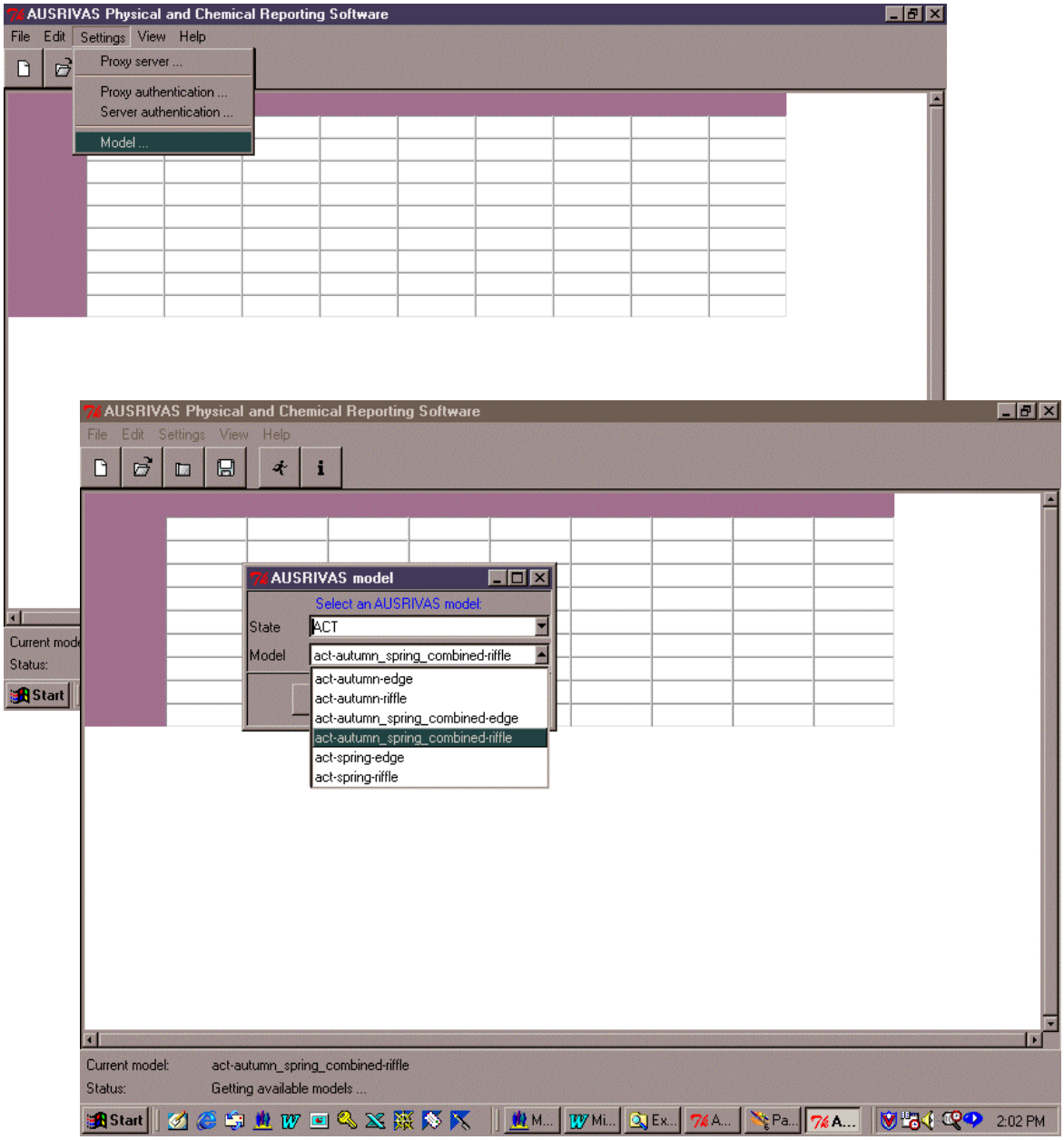
Reference data are the physical and chemical information collected from sites designated *a-priori* as reference sites, and which were used to construct the AUSRIVAS biological models. The reference site data is the information against which a test site is compared (see Figure 1.1). In the AUSRIVAS physical and chemical reporting software the reference site information is stored centrally and the variables included in the reference data are pre set. However, States and Territories may wish to modify or update the reference data sets and instructions for doing this are provided in Part 7.

3.2 How to set up a test data set

3.2.1 Step 1
Find out which models are available

The central AUSRIVAS server (see Figure 2.1) contains reference data for individual State and Territory models. Test data must be set up to match these models. That is, if using a spring model, test physical and chemical data must be collected in spring. Specific variables may also need to match the habitat type. In addition, the outputs of the physical and chemical reporting software are used to aid interpretation of the AUSRIVAS O/E scores (see Part 6) and it is important to choose the physical model that corresponds with the desired AUSRIVAS biological model.

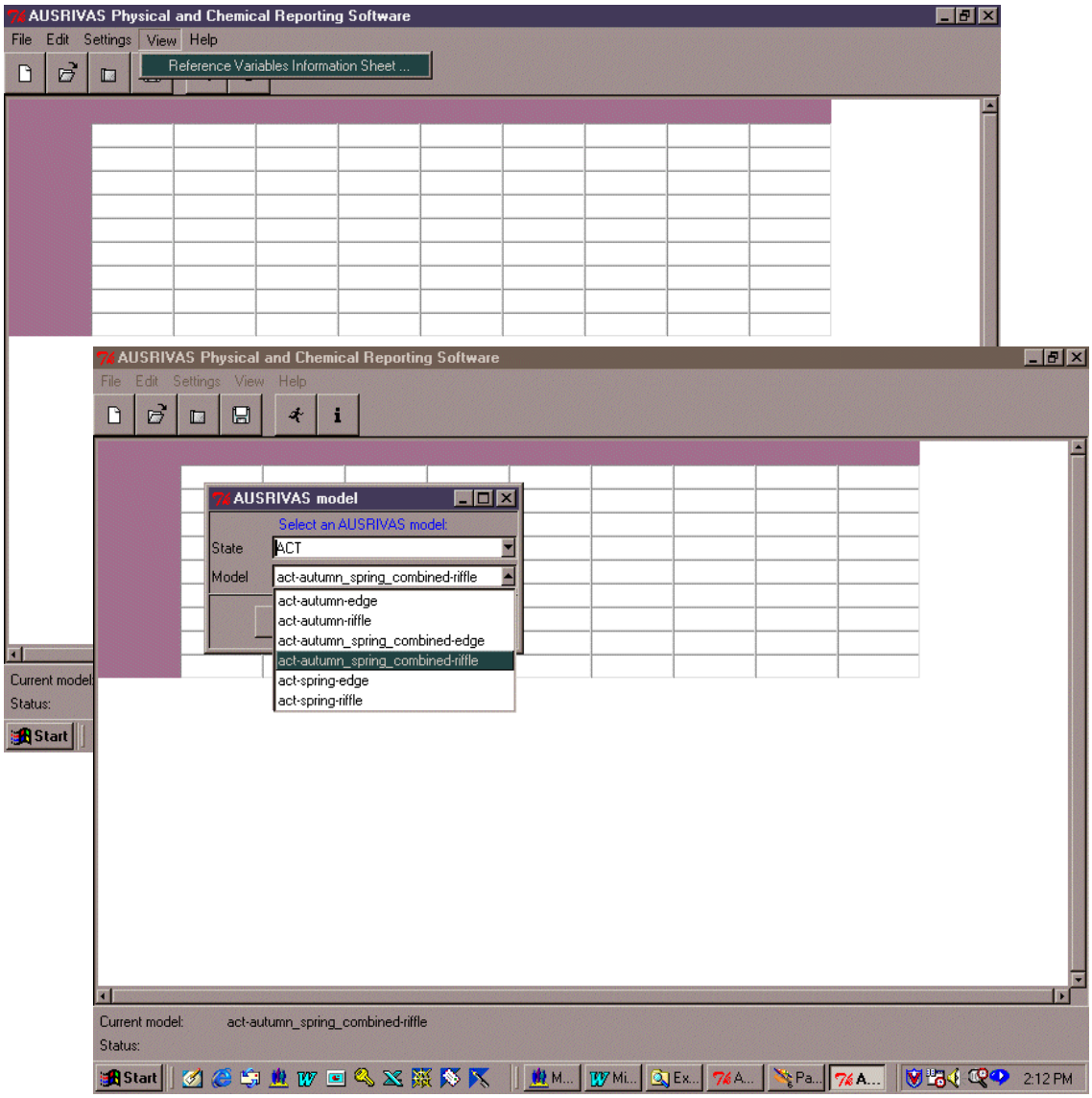
To find out which models are available in the AUSRIVAS physical and chemical reporting software, choose the "Settings" menu and then choose "Model". The model dialog box will appear. Choose the correct State or Territory and the correct model.



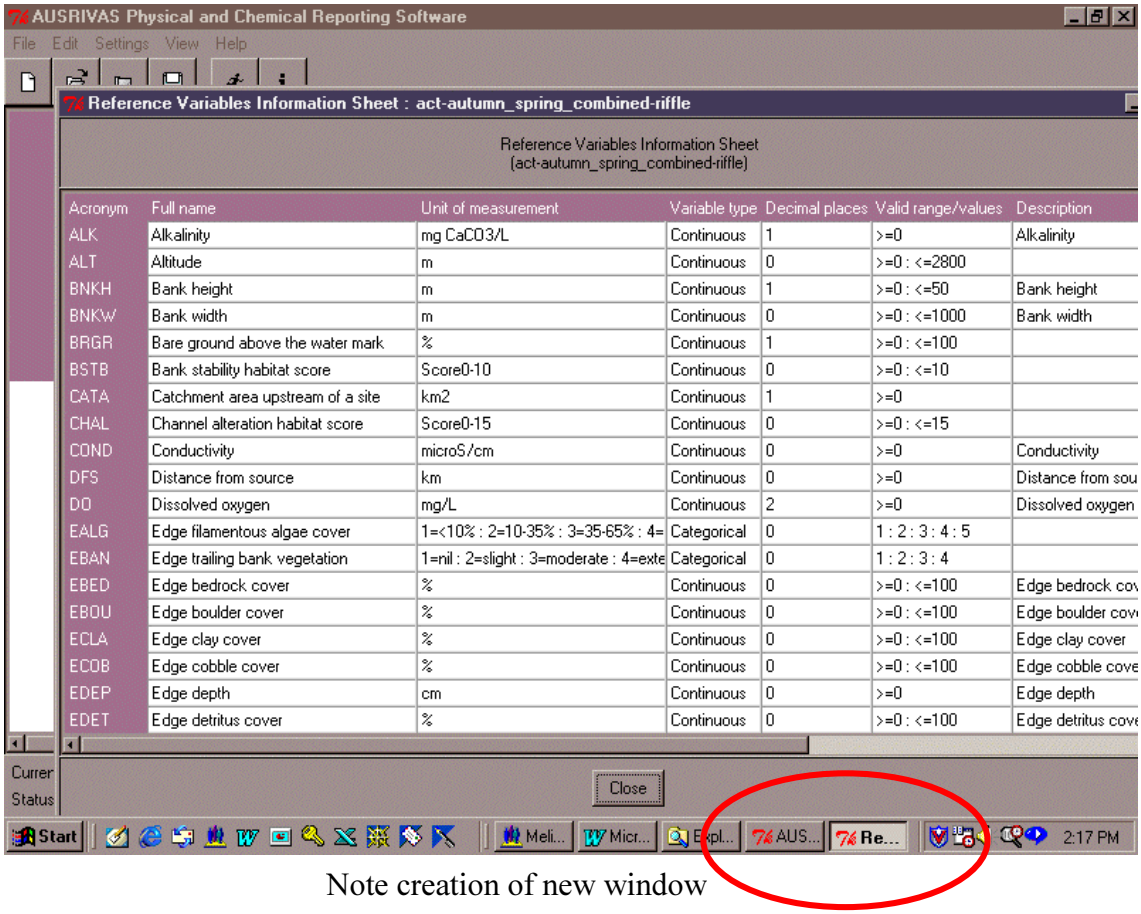
3.2.2 Step 2
Find out which variables are available in the corresponding reference data set

To set up a test data set, users will need to know which variables are available in the corresponding reference data set. The variables contained in each reference data set can be viewed via a sheet called the **REFERENCE VARIABLES INFORMATION SHEET.** This sheet contains information on the variables contained in a model, the acronym of each variable, the unit of measurement for each variable, the number of

decimal places used for each variable and the valid range of each variable (see Sections 3.2.3 and 7.3 for more information on the reference variables information sheet). To view the reference variables information sheet, choose the "View" menu and pick "Reference variables information sheet". The model dialog box will appear again. Select the correct model and hit OK.



The reference variables information sheet resembles the following. The variables included in the selected model (act-autumn_spring_combined-riffle) are listed in the left hand column. Note that the reference variables information sheet creates a new window.



Note creation of new window

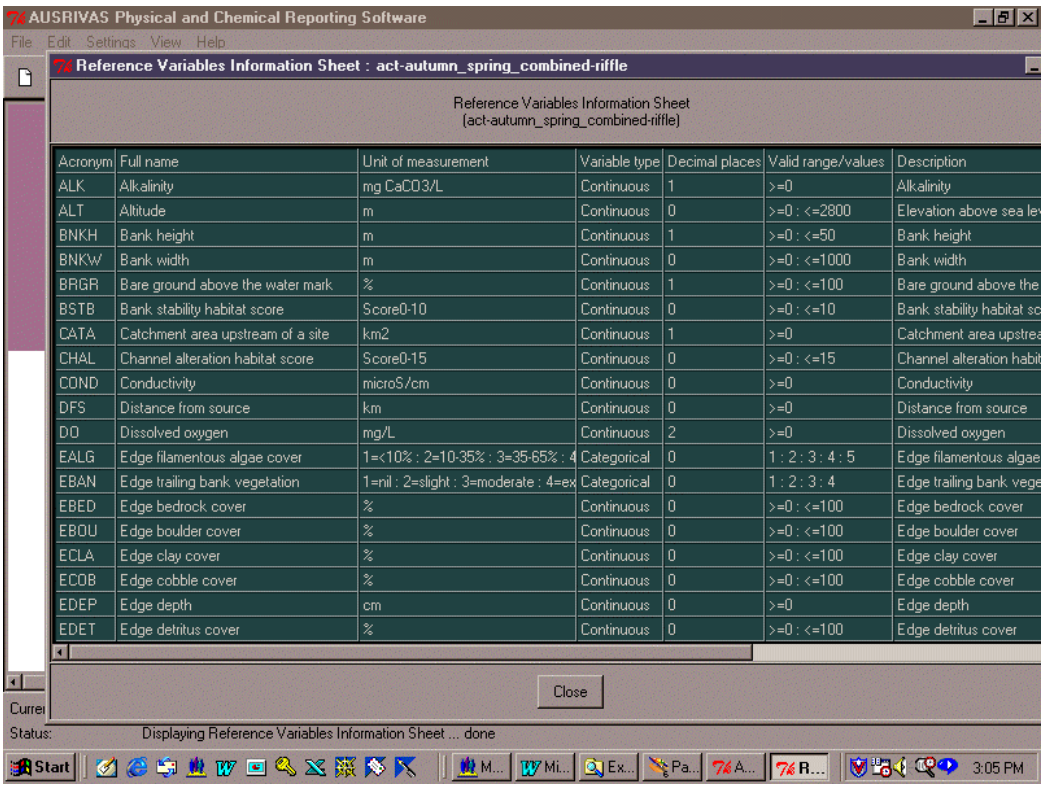
Special note on column widths

To change the width of any column for viewing, place the cursor on the edge of the column - the cursor will switch to a + symbol. Then, holding the RIGHT mouse button down, drag the column out to the desired width.

If required, steps for printing the reference variables information sheet are as follows:

Printing the reference variables information sheet - Step 1

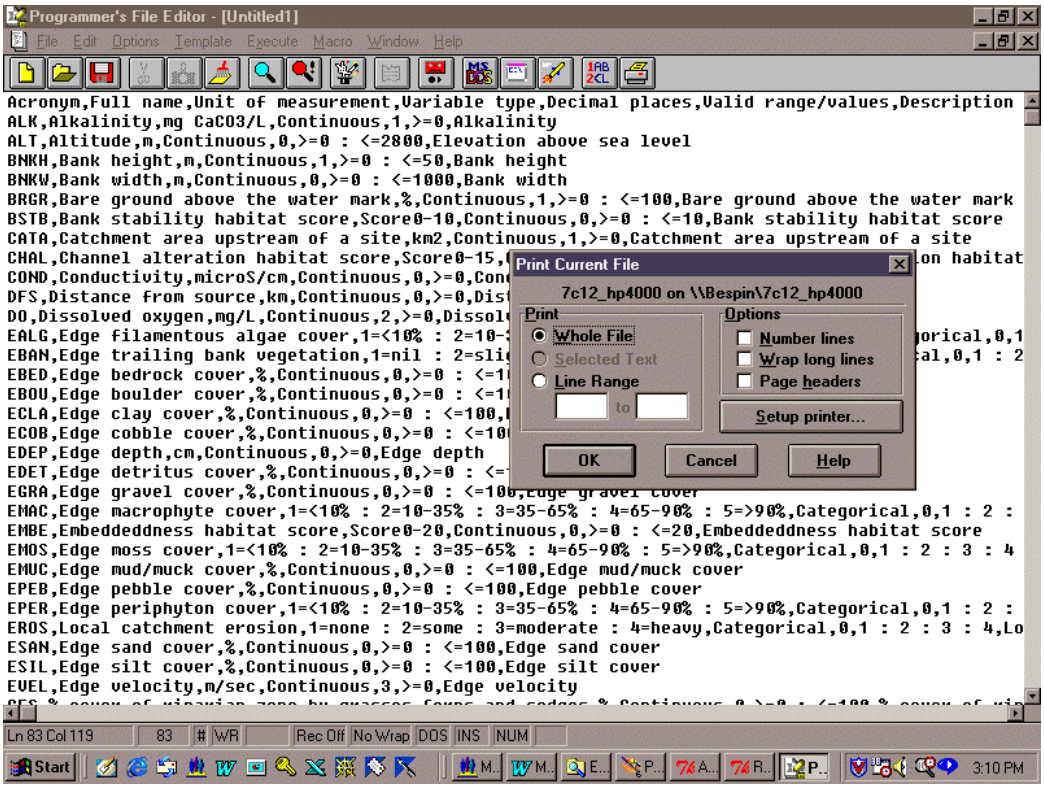
Select the sheet by clicking in the top left hand cell (acronym). The sheet will be highlighted as a green colour. Copy the selected cells using the <Control> and C keys.



Acronym	Full name	Unit of measurement	Variable type	Decimal places	Valid range/values	Description
ALK	Alkalinity	mg CaCO3/L	Continuous	1	>=0	Alkalinity
ALT	Altitude	m	Continuous	0	>=0 : <=2800	Elevation above sea le
BNKH	Bank height	m	Continuous	1	>=0 : <=50	Bank height
BNKW	Bank width	m	Continuous	0	>=0 : <=1000	Bank width
BRGR	Bare ground above the water mark	%	Continuous	1	>=0 : <=100	Bare ground above the
BSTB	Bank stability habitat score	Score0-10	Continuous	0	>=0 : <=10	Bank stability habitat sc
CATA	Catchment area upstream of a site	km2	Continuous	1	>=0	Catchment area upstrea
CHAL	Channel alteration habitat score	Score0-15	Continuous	0	>=0 : <=15	Channel alteration habit
COND	Conductivity	microS/cm	Continuous	0	>=0	Conductivity
DFS	Distance from source	km	Continuous	0	>=0	Distance from source
DO	Dissolved oxygen	mg/L	Continuous	2	>=0	Dissolved oxygen
EALG	Edge filamentous algae cover	1=<10% : 2=10-35% : 3=35-65% : 4	Categorical	0	1 : 2 : 3 : 4 : 5	Edge filamentous algae
EBAN	Edge trailing bank vegetation	1=nil : 2=slight : 3=moderate : 4=ex	Categorical	0	1 : 2 : 3 : 4	Edge trailing bank vege
EBED	Edge bedrock cover	%	Continuous	0	>=0 : <=100	Edge bedrock cover
EBOU	Edge boulder cover	%	Continuous	0	>=0 : <=100	Edge boulder cover
ECLA	Edge clay cover	%	Continuous	0	>=0 : <=100	Edge clay cover
ECOB	Edge cobble cover	%	Continuous	0	>=0 : <=100	Edge cobble cover
EDEP	Edge depth	cm	Continuous	0	>=0	Edge depth
EDET	Edge detritus cover	%	Continuous	0	>=0 : <=100	Edge detritus cover

Printing the reference variables information sheet - Step 2

Paste into a text editor program. The file will be in comma delimited (.csv) format. The file can be printed from the text editor or saved as a text (.txt) file and imported into Excel.



3.2.3 Step 3
Find out the attributes of each desired variable

The attributes of each variable listed on the reference variables information sheet are as follows:

Acronym	This is the exact acronym assigned to each variable
Full name	Gives the full name of the variable
Unit of measurement	Gives the unit of measurement for each variable. Different format for continuous or categorical variables.
Variable type	Identifies the variable as either continuous or categorical. Continuous variables are those for which data are arranged along a continuum. Categorical variables are those for which data are arranged in finite categories.
Decimal places	Gives the number of decimal places reported for each variable.
Valid range/values	Sets the valid range of values that can be expected for each variable.
Description	Provides additional information about the variable

Test data must be set up according to these attributes. Instructions for setting up a test data set are provided in the following section.

3.2.4 Step 4
Set up the test data set in an Excel spreadsheet

Once you know which models are available, which variables are available in these models, and the attributes of these variables it is easiest to set up the test data set in an Excel spreadsheet (or similar spreadsheet package). Any number of test sites can be included, but as discussed in Section 3.2.1 the test data must be collected in the same season and habitat type. Data have the following general properties:

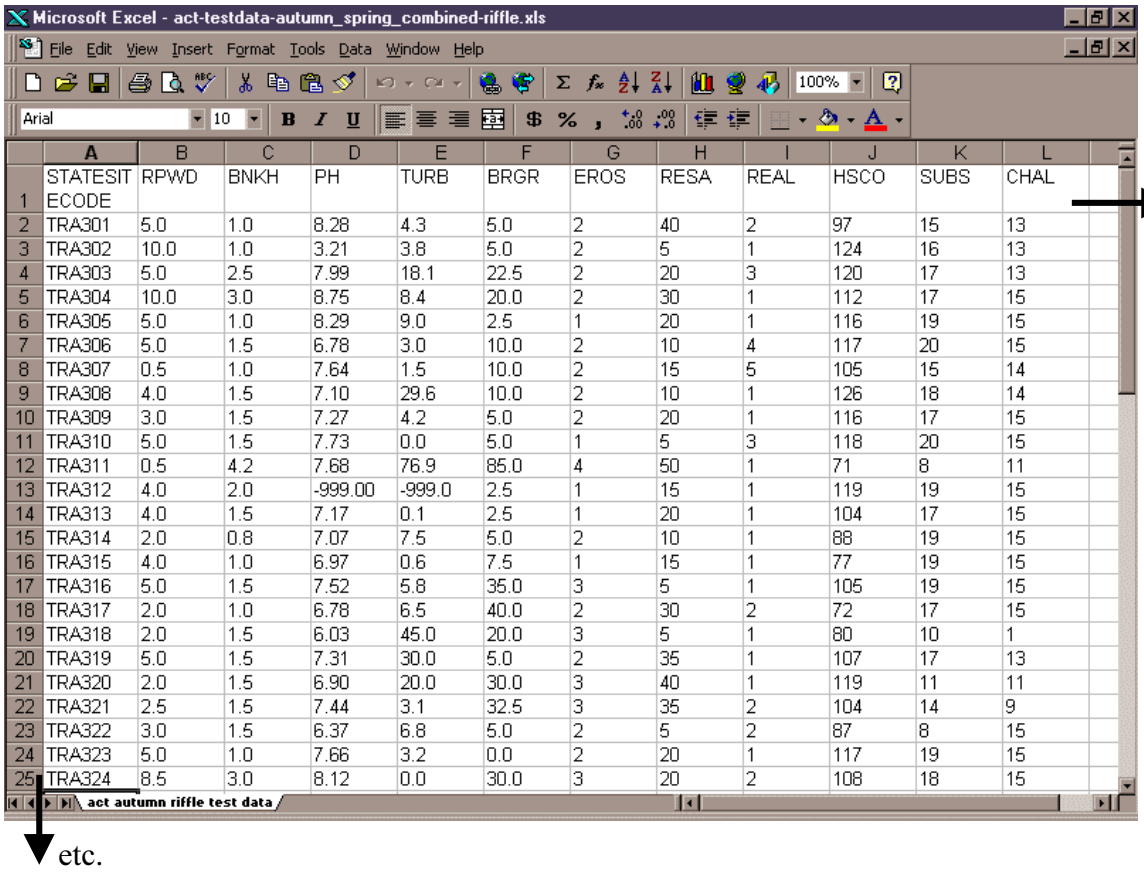
- Data are arranged with sites down the side (i.e. as rows) and variables across the top (i.e. as columns).
- The first column is the test site number, which MUST be labelled with the acronym STATESITECODE

  This acronym must be in capitals and contain no spaces.

- The spelling of acronyms must match EXACTLY with those listed in the reference variables information sheet. It is also important that the acronyms are in capitals with no spaces. Details on how to access the reference variables information sheet are provided in Section 3.2.2.
- Not all the variables available for any one model must be included in the test data set. The user can include, or 'test', as few or as many variables as desired.
- The variables that can be analysed in a test data set are limited to those that are available in the reference data set for a corresponding model. However, the test data set can contain extra variables that are not listed in the reference variables information sheet (this will aid in minimising the amount of manipulation required when setting up a test data set). Submission of these extra variables will return them as a list of "unused variables". This list of unused variables is one of the outputs, and is detailed in Section 5.1.
- It is critical that the units of measurement used for continuous variables in the test data set match those given on the reference variables information sheet, otherwise the comparison of test data against reference data will result in spurious outputs. Details on how to access the reference variables information sheet are provided in Section 3.2.2.
- It is also critical that the categories used for categorical variables in the test data set match those given on the reference variables information sheet, otherwise the comparison of test data against reference data will result in spurious outputs. Details on how to access the reference variables information sheet are provided in Section 3.2.2.
- The number of decimal places used for each variable in the test data set should be set according to the details provided on the reference variables information sheet. Details on how to access the reference variables information sheet are provided in Section 3.2.2.
- Cells with missing data are denoted with the value -999 Subsequently, this variable at this site can not be compared against reference site information.
- Cells can not contain formatting such as subscript, superscript, bold or italicised characters.
- Continuous variables can not contain characters such as < or >. Where chemical variables are listed as detection limits (e.g. <0.001) the < symbol must be removed (e.g. <0.001 becomes 0.001).

- Categorical variables can contain text (e.g. high, low, medium or bedrock, boulder, cobble, pebble, gravel, sand, silt, clay etc.).

The test data set will resemble the following spreadsheet:



	A	B	C	D	E	F	G	H	I	J	K	L
1	STATESITECODE	RPWD	BNKH	PH	TURB	BRGR	EROS	RESA	REAL	HSCO	SUBS	CHAL
2	TRA301	5.0	1.0	8.28	4.3	5.0	2	40	2	97	15	13
3	TRA302	10.0	1.0	3.21	3.8	5.0	2	5	1	124	16	13
4	TRA303	5.0	2.5	7.99	18.1	22.5	2	20	3	120	17	13
5	TRA304	10.0	3.0	8.75	8.4	20.0	2	30	1	112	17	15
6	TRA305	5.0	1.0	8.29	9.0	2.5	1	20	1	116	19	15
7	TRA306	5.0	1.5	6.78	3.0	10.0	2	10	4	117	20	15
8	TRA307	0.5	1.0	7.64	1.5	10.0	2	15	5	105	15	14
9	TRA308	4.0	1.5	7.10	29.6	10.0	2	10	1	126	18	14
10	TRA309	3.0	1.5	7.27	4.2	5.0	2	20	1	116	17	15
11	TRA310	5.0	1.5	7.73	0.0	5.0	1	5	3	118	20	15
12	TRA311	0.5	4.2	7.68	76.9	85.0	4	50	1	71	8	11
13	TRA312	4.0	2.0	-999.00	-999.0	2.5	1	15	1	119	19	15
14	TRA313	4.0	1.5	7.17	0.1	2.5	1	20	1	104	17	15
15	TRA314	2.0	0.8	7.07	7.5	5.0	2	10	1	88	19	15
16	TRA315	4.0	1.0	6.97	0.6	7.5	1	15	1	77	19	15
17	TRA316	5.0	1.5	7.52	5.8	35.0	3	5	1	105	19	15
18	TRA317	2.0	1.0	6.78	6.5	40.0	2	30	2	72	17	15
19	TRA318	2.0	1.5	6.03	45.0	20.0	3	5	1	80	10	1
20	TRA319	5.0	1.5	7.31	30.0	5.0	2	35	1	107	17	13
21	TRA320	2.0	1.5	6.90	20.0	30.0	3	40	1	119	11	11
22	TRA321	2.5	1.5	7.44	3.1	32.5	3	35	2	104	14	9
23	TRA322	3.0	1.5	6.37	6.8	5.0	2	5	2	87	8	15
24	TRA323	5.0	1.0	7.66	3.2	0.0	2	20	1	117	19	15
25	TRA324	8.5	3.0	8.12	0.0	30.0	3	20	2	108	18	15

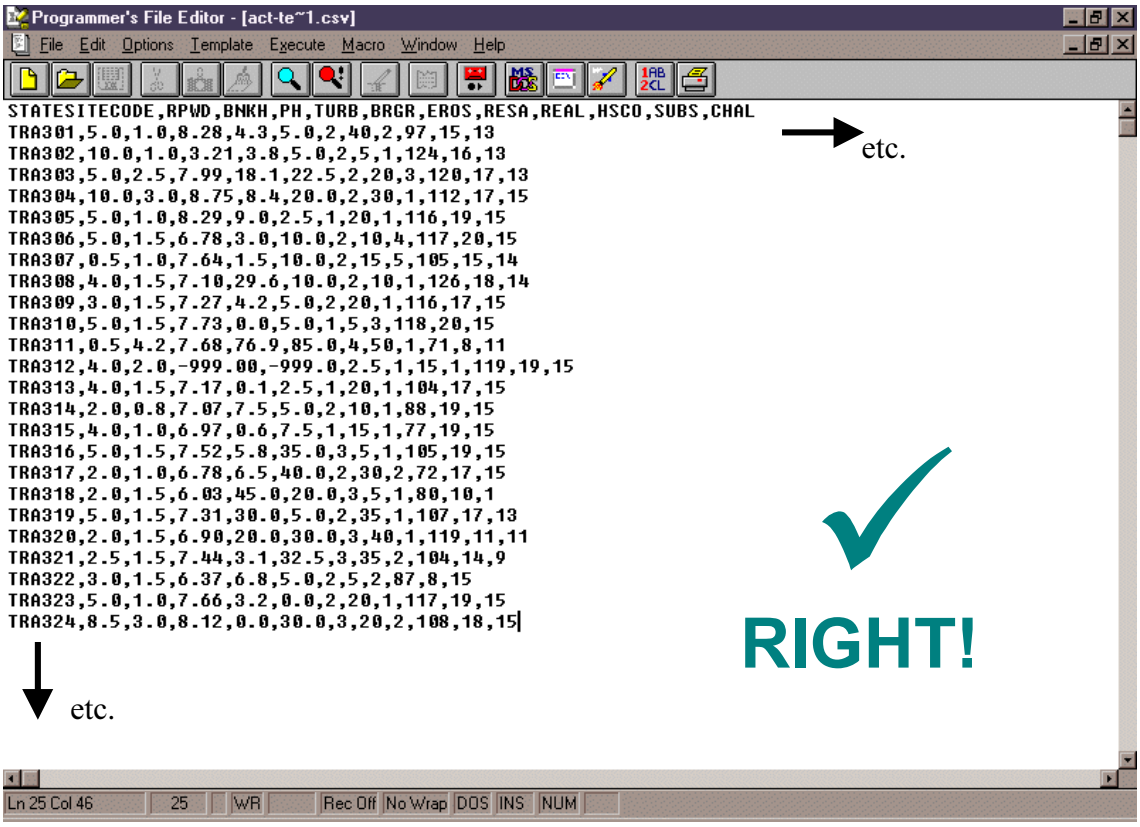
3.2.5 Step 5
Convert the spreadsheet to .csv file format

- For the program to be able to read the test data, the spreadsheet must be converted to a comma delimited .csv file. This is easily done using the "Save As" function in Excel – choose the file type called "CSV (comma delimited) (*.csv)".
- File names must follow the following format and be in lower case:

  state-testdata-model-habitat.csv

  Spaces within any portion must be underscored. For example, the ACT combined season riffle model test data file name is act-testdata-autumn_spring_combined-riffle.csv

When converted to .csv format the test site data will resemble the following:



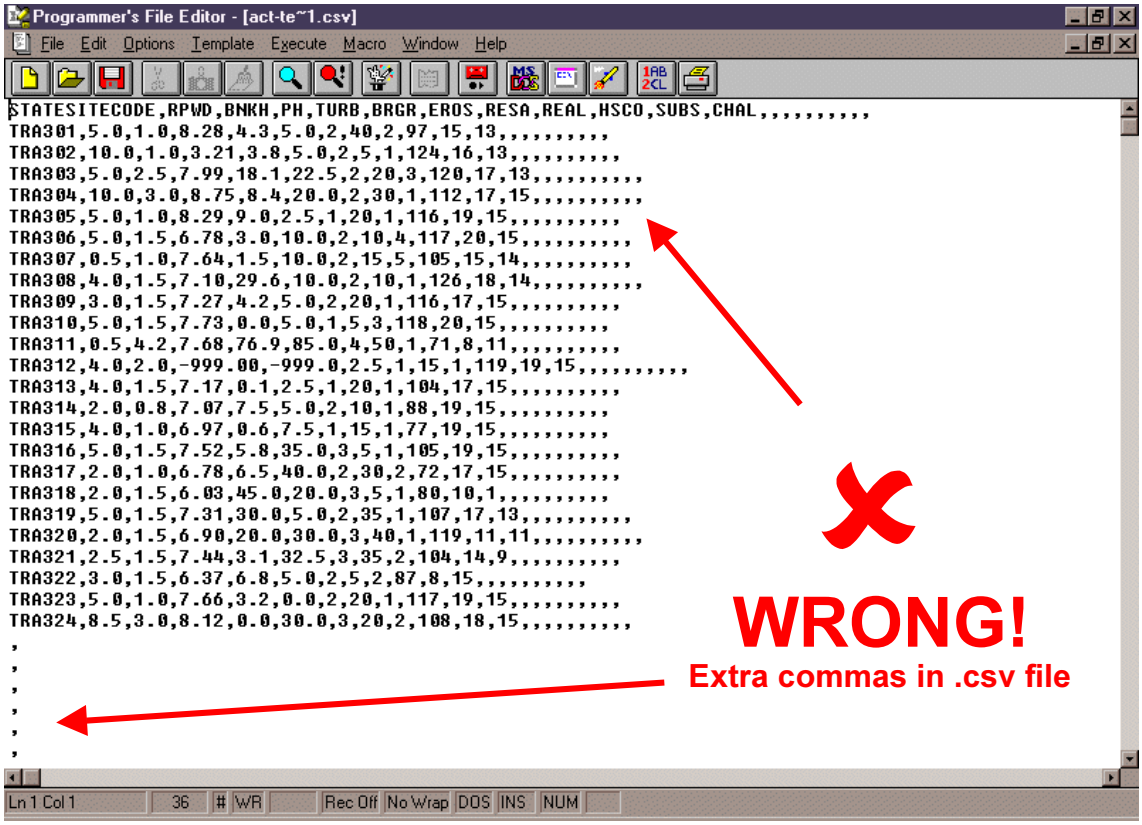
All .csv files **must be checked after conversion.** This can be done by viewing in a text editor such as Word Pad or similar.

3.2.5.1 Problems encountered when converting spreadsheets to .csv format

Some common problems encountered when converting spreadsheets to .csv files are outlined in the following sections.

Extra commas in .csv file

As far as we can tell, this problem occurs where certain formatting (such as bold) is switched on in empty rows and columns in the Excel spreadsheet. The resulting .csv file will look like this:

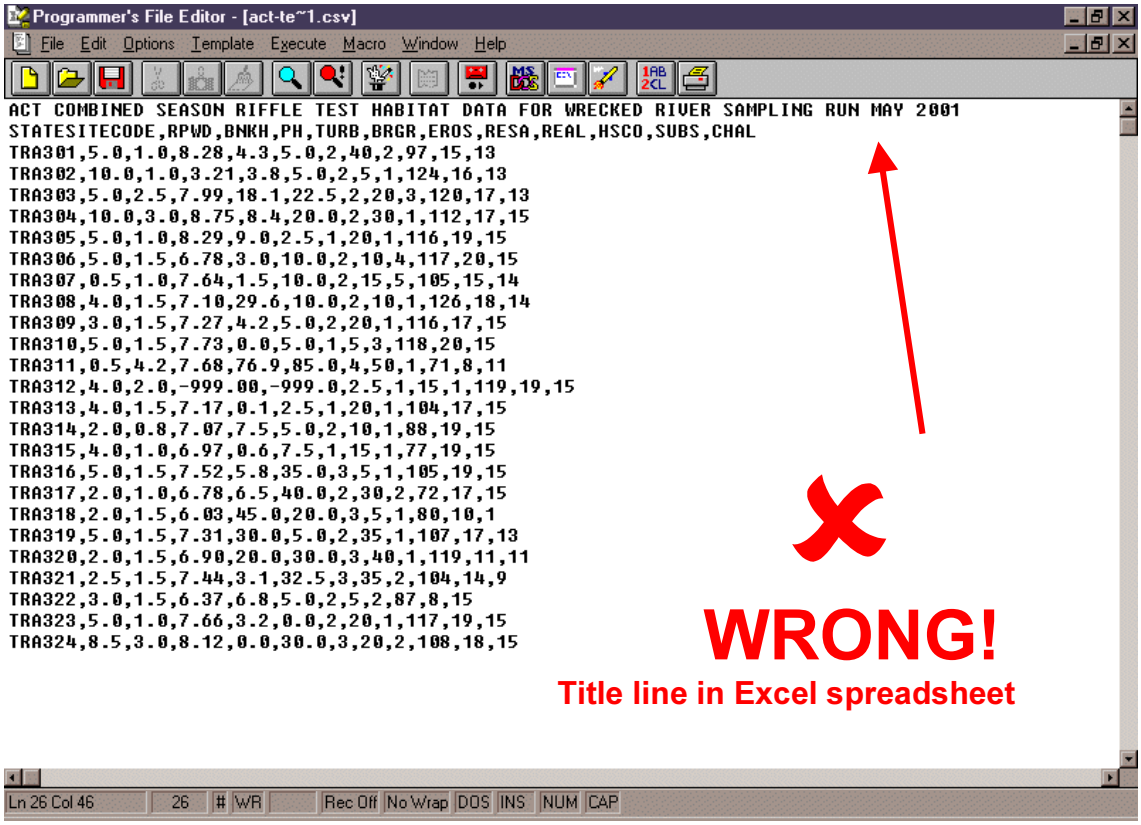


The **solution** is to either:

1. Remove any formatting from empty rows and columns in excel before saving as .csv format. Check the .csv file again after this is done though, as it is not 100% certain that this is the cause of the problem.
2. Remove the commas manually from the .csv file using a text editor such as Word Pad.

Title line in Excel spreadsheet

This problem occurs where the user has inserted a title line as Row 1 of the Excel spreadsheet. The resulting .csv file will resemble the following:



The **solution** is to remove the title line from the Excel spreadsheet before converting to .csv format. Row 1 should be the column title acronyms.

3.3 Checklist

At the end of setting up a test data set you should have:

1. a test data set for each desired model, saved in .csv format with NO ERRORS! Each variable in the test data set must conform to the attributes given in the reference variables information sheet.