

**Dealing with common RECCE and Permanent Plot data  
quality issues during data entry or checking  
Version 1**

J.M. Hurst, H. Broadbent and M. McKay

Landcare Research  
PO Box 69, Lincoln 8152  
New Zealand

Landcare Research Internal Report: LC0506/128

DATE: June 2006



ISO 14001

© **Landcare Research New Zealand Ltd 2006**

This information may be copied and distributed to others without limitation, provided Landcare Research New Zealand Limited is acknowledged as the source of the information. Under no circumstances may a charge be made for this information without the express permission of Landcare Research New Zealand Limited

---

## Contents

---

1.	Introduction .....	5
2.	General guidelines for data-entry .....	5
2.1	How should digital data files for NVS datasets be named? .....	5
2.2	What is the ‘text file’ associated with a NVS dataset, and what information should it contain? .....	5
2.3	How should error corrections to NVS data be documented? .....	6
2.4	What happens when a six-letter species code could denote more than one species? .	6
2.5	How can I check that the species codes I am entering are correct? .....	7
2.6	Why is my species not on the NVS list of species codes? .....	7
2.7	What should be done if the value for any data is outside the expected range? .....	7
2.8	Are there any data parameters that I should consider entering that are not used by the PC-RECCE, PC-DIAM, or PC-USTOREY packages? .....	7
2.9	What should be done if data are illegible? .....	7
2.10	Should corrections to data be made on the field sheets? .....	8
3.	RECCE data-entry guidelines .....	8
3.1	What should I do if the field team forgot to collect data for any parameter on the RECCE site description? .....	8
3.2	How should plot aspect be entered for flat plots? .....	8
3.3	What should be done if there are two entries for a species on a RECCE description? .....	8
4.	Stem diameter data-entry guidelines .....	9
4.1	What format should I use to enter tree tag numbers? .....	9
4.2	What should be done when trees on the same plot have been recorded against the same tag number? .....	9
4.3	What should be done if a tree has been retagged with a different tag number during a plot remeasurement? .....	10
4.4	What should be done about species identification inconsistencies for trees between surveys? .....	10
4.5	What should be done when two stems that were previously measured separately have fused together, and field staff remeasured them as one stem? .....	11
4.6	What should be done if a tree was missed at a previous plot measurement? .....	11
4.7	What should be done if field staff found two tags on a tree stem during a plot remeasurement? .....	12
4.8	What should be done if the tree has been tagged but the species name is not recorded, or the tree is tagged and the species recorded but the diameter is missing? .....	12

5.	Seedling and sapling data-entry guidelines .....	13
5.1	What should be done if there are two entries for a species in a seedling subplot? ...	13
5.2	What should be done if there is no data recorded against a species name recorded on a seedling subplot? .....	13
5.3	How should the presence of non-woody species in seedling tiers be entered?.....	13
5.4	What should be done if there are two entries for a species for a sapling subplot? ...	13
6.	References .....	14

---

## 1. Introduction

---

This document outlines a set of guidelines for minimising errors and inconsistencies in RECCE and permanent-plot (20 × 20 m) data archived in the National Vegetation Survey databank (NVS). These ‘rule of thumb’ guidelines summarise and build upon those formulated in earlier reports (Wiser et al. 1999; Newell & Baldwin 2000), but focus specifically on problems identified during data entry or checking.

It is important to follow consistent procedures for correcting errors or anomalies in data stored in NVS. These recommendations are therefore intended to be applicable regardless of the procedure or software used to enter data. However, because software or templates used for data entry differ, the detection of the kinds of errors outlined here may occur at either the data-entry stage or data-checking stage. We recommend that where data-entry facilities used do not allow point-of-entry identification of errors or anomalies (e.g. most MS-Excel templates), that an extensive checking exercise is conducted after data entry, by somebody familiar with the data and plot methods used.

---

## 2. General guidelines for data-entry

---

### 2.1 How should digital data files for NVS datasets be named?

Each computer file associated with a particular survey should be named using the survey name recorded on the RECCE field sheets. For example, ASCII format data files (see Hall 1992, 1994a, 1994b) containing data from the Anatoki permanent-plot survey conducted in 2002 would be named ‘ANATOK02.REC’ (RECCE data), ‘ANATOK02.QUA’ (diameter data), and ‘ANATOK02.UND’ (seedling and sapling data). Filenames within NVS are kept to a maximum of eight characters. It is also useful to keep the names of remeasured surveys the same as the original, as renaming surveys can cause confusion down the track. For example, the Anatoki 2002 survey (e.g. ANATOK02.REC) is a remeasure of the Anatoki 1992 survey (e.g. ANATOK92.REC).

### 2.2 What is the ‘text file’ associated with a NVS dataset, and what information should it contain?

Each survey has an associated text file named using the survey name recorded on the RECCE field sheets (e.g. ANATOK02.txt). This file contains any important information regarding the survey that is not captured in the data-files or metadata. Make sure the text file is well laid out and easy to follow. Avoid using abbreviations. Separate each section of the text file with clear headings, e.g. include a section containing general notes regarding the whole survey, and separate sections for notes associated with each plot.

A number of different things should go into the text file:

- Information regarding any deviations from standard methodologies.
- Names of people involved in data collection.
- Data corrections made during data entry, such as corrected species codes or duplicate tag corrections; along with the date and name of the person making the correction.
- Nomenclature followed for a survey (for each group of taxa recorded).
- Any notes regarding data quality (e.g. made during data collection, data checking or data entry).
- Any assumptions made during data entry.
- Notes recorded on field sheets that do not conform to standard fields allowable in the data-entry template or database.
- Any problems to address at the next plot measurement.
- Other notes as relevant.

### **2.3 How should error corrections to NVS data be documented?**

All errors found in NVS datasets should be documented. Details of corrections made can be recorded into a text file associated with the dataset (as described throughout this report). Alternatively, where a large number of corrections are made to a dataset because of similar types of problems, they may be more easily documented by listing the changes made in an MS-Excel spreadsheet. Contact the NVS database administrator (email: [nvs@landcareresearch.co.nz](mailto:nvs@landcareresearch.co.nz)) for further information.

### **2.4 What happens when a six-letter species code could denote more than one species?**

Each plant species recorded in NVS vegetation plot data is denoted with a six-letter code, which usually consists of the first three letters of the genus followed by the first three letters of the specific epithet of the botanical name. For example *Pseudopanax crassifolius* is denoted as PSEcra. Using this intuitive code system some six-letter codes can denote more than one species. For example, the intuitive code of both *Pseudopanax colensoi* and *Pseudowintera colorata* is PSEcol. To avoid such duplications non-intuitive codes are used for some species (e.g. the NVS code used for *Pseudowintera colorata* is PSEcol, while that used for *Pseudopanax colensoi* is NEOcol).

Be aware of this problem. See <http://nvs.landcareresearch.co.nz/html/NameSearchHelp.aspx> or contact the NVS database administrator (email: [nvs@landcareresearch.co.nz](mailto:nvs@landcareresearch.co.nz)) for an up-to-date list of NVS species codes. Field teams are not always aware of the correct code for a species, so where a code recorded could potentially represent either of two alternative species it is always prudent to check which species was intended, with the person who collected the data.

## **2.5 How can I check that the species codes I am entering are correct?**

Do a search against the common, scientific or six-letter NVS code for a species at <http://nvs.landcareresearch.co.nz/WebForms/SearchForm.aspx> to make sure that each six-letter code you enter is correct. If any anomalies are found query the people who measured the plot (asking for a full species name). Note any anomalies between what is recorded on the field sheets and what was entered, under an appropriate heading in the text file for the plot.

## **2.6 Why is my species not on the NVS list of species codes?**

In some cases a species may not yet be on the NVS six-letter code list. If you know the full name, you can check that this is a current name using the NZ Plant Names Database Ngā Tipu Aotearoa – New Zealand Plants (<http://www.landcareresearch.co.nz/databases/>). When you find the full name, contact the NVS database administrator who will add the species onto the NVS list, and will let you know the correct code to use. Do not assign a six-letter code to any new species without first checking with the NVS database administrator, as the code could conflict with a six-letter code already assigned to another species.

If the species name that has been used is not on the NZ Plant Names Database, but is formally recognised, use the feedback function on the NZ Plant Names Database website or contact the NVS database administrator.

## **2.7 What should be done if the value for any data is outside the expected range?**

Query the people who collected the data. If the data value remains inconceivable, do not enter it but note it in the text file for that plot.

## **2.8 Are there any data parameters that I should consider entering that are not used by the PC-RECCE, PC-DIAM, or PC-USTOREY packages?**

Data fields used by analyses included in the PC- packages are described in the relevant manuals (Hall 1992, 1994a, 1994b). In addition to the fields used by the PC- packages:

- It is useful to have a digital copy of the full GPS grid references of all plots in a RECCE or permanent-plot survey.
- For permanent-plot data it is recommended that the subplot of each tagged tree is entered, which can aid future plot remeasurements, and is essential for certain types of data analyses.
- For permanent-plot data the ‘stem state’ of each tagged tree should also be entered using the following four categories: alive, dead, not found, unknown. This provides an indication of data quality, e.g. ‘not found’ indicates that a field team looked for a missing tree, whereas if a large number of trees were in the ‘unknown’ category this might highlight that a dataset is of poor quality.

## **2.9 What should be done if data are illegible?**

If any data are illegible, the people who measured the plots should be queried. If the data can still not be deciphered then the item should not be entered, but should be brought to the attention of data users via a note in the text file for that plot.

### **2.10 Should corrections to data be made on the field sheets?**

Corrections to data should be noted in the text-file associated with a survey, which should be printed and stored with the original field sheets. In general, notes or corrections to data should NOT be made on original field sheets. This is because corrections made to the original field sheets can make the original data illegible, which will be a major problem if it is later found that the correction was actually incorrect! It is also impossible to determine whether corrections to original field sheets were made in the field at the plot location, or 10 years later by somebody in the office who decided that a particular change should be made.

An exception to this rule occurs when collected specimens have been properly identified. In this case the tag name written on field sheets should be corrected but the '©' should be retained so that people know that a specimen was collected.

---

## **3. RECCE data-entry guidelines**

---

### **3.1 What should I do if the field team forgot to collect data for any parameter on the RECCE site description?**

When data for any particular site parameter is unknown for any reason (e.g. slope of a plot was not recorded) DO NOT ENTER ANY DATA in that field during data entry. DO NOT ENTER ZERO or any other coding, as these could be misinterpreted. If clarification is needed on how the data written on the field sheet has been interpreted during data entry, write a note in the text file.

### **3.2 How should plot aspect be entered for flat plots?**

Plots that are flat or nearly flat should not have an aspect. Aspect for flat plots should be left blank during data entry. DO NOT ENTER ZERO for the aspect of flat plots, since this signifies a northerly aspect.

### **3.3 What should be done if there are two entries for a species on a RECCE description?**

If two entries are recorded for the same species on a RECCE description, create one record for the species by entering the greatest of the cover-class scores in each height tier. Make a note under an appropriate heading in the text file for the survey.

e.g. General Notes:

When a species was recorded more than once on a RECCE description, one record was created by entering the greatest of the cover-class scores in each height tier.



---

## 4. Stem diameter data-entry guidelines

---

### 4.1 What format should I use to enter tree tag numbers?

Some software packages make distinctions between tag numbers that are entered in different formats. Thus a tag entered as 0517 may be distinguished from a tag entered as 517. If entering data from previously measured plots, first check electronic copies of the data to see how tag numbers were previously entered. Enter tag data in the same format as was done previously to ensure that individual tags ‘match-up’ during data checking and analysis. Be aware that if using MS-Excel the format of the cells where tags numbers are entered may need to be changed to retain zeros in front of tag numbers, if required.

The NVS databank can now accept tags with >4 digits. For newly established permanent-plot surveys you may wish to enter the full tag number (however, be aware that the PC-DIAM package can only accept tags that are ≤4 digits).

### 4.2 What should be done when trees on the same plot have been recorded against the same tag number?

‘Duplicate tags’, where trees have been recorded against the same tag number in the same plot, can occur for various reasons (Newell & Baldwin 2000). If you encounter a duplicate tag while entering or checking data, first check that it really is a duplicate (not just misread). There have been various approaches taken to deal with duplicate tags; however, best practice is as follows:

If one of the individuals with the duplicate tag number is DEAD, then this individual should be ‘digitally retagged’ in the data file. Otherwise, when both individuals with the duplicate tag are alive, the tree with the SHORTEST HISTORY OF REMEASUREMENT should be the one digitally retagged. For example, if one of the trees was first tagged in 1975, and the other first tagged in 1985, and both were remeasured in 1995 and 2005, the tree first tagged in 1985 has a shorter measurement history so this individual should be the one digitally retagged.

#### *How to ‘digitally retag’ a tree*

In the data file assign the tree to be digitally retagged with a new tag number. This new tag number will begin with a \* followed by the last three digits of the duplicate tag number, e.g. 8095 becomes \*095. The tag number for the other tree is entered as normal, i.e. 8095.

Because the treatment of duplicate tag numbers may have been inconsistent between each measurement of any particular survey, digital files from the previous measurement(s) must be checked and the change backdated if necessary. If data from any of these previous measurements are lodged with NVS, please contact the NVS database administrator so changes can be made in the NVS database.

#### *Documentation required*

The change must be documented in the text file for the survey. Put the information into the appropriate text file under the Plot Name and then the heading ‘Data Corrections’.

e.g. Plot 54/1

Data Corrections:

Tag # W1563 has been recorded against two different stems in 2005, ULEeur 3.1 cm & ULEeur 6.6 cm. ULEeur 3.1 cm has been entered in the digital file with the tag # \*563 and the change has also been backdated in the 1985 and 1995 digital files.

On the list of problems to be resolved for the plot (e.g. in the text file), highlight the need to physically retag the digitally retagged individual the next time the plot is measured.

e.g. Plot 54/1

Data Problems:

Tag # W1563 has been recorded against two different stems in 2005, ULEeur 3.1 cm & ULEeur 6.6 cm. ULEeur 3.1 cm has been entered in the digital file with the tag # \*563. Please check and assign the correct/new tag no. to one of these.

#### **4.3 What should be done if a tree has been retagged with a different tag number during a plot remeasurement?**

Trees should NEVER be retagged with a different tag number; however, this may have sometimes occurred in the past. The tag number change must be backdated in all digital files from previous plot measurements and noted under an appropriate heading in the text file for that plot.

e.g. Plot 54/1

Data Corrections:

Tree AR1ser 6375 was retagged with the tag number 8217 in the 1995 survey. The tag number change for this tree has been backdated in both the 1975 and 1985 digital files.

If data from any of the previous measurements are lodged with NVS, please contact the NVS database administrator so changes can be made in the NVS database.

#### **4.4 What should be done about species identification inconsistencies for trees between surveys?**

Detailed descriptions on dealing with species identification anomalies between surveys are given in Newell & Baldwin (2000). Inconsistencies in species identification are often found during data-checking procedures. When species identification for a particular tree has been inconsistent between measurements, neither identity can be assumed correct unless both:

- pre-printed stem diameter sheets were used listing data from a previous measurement;
- AND
- the change in identification for a particular tag number was specifically noted as checked on the plot sheet.

#### *Solution and documentation required*

If the above criteria are met, the identification change must be backdated in all digital files and noted under an appropriate heading in the text file for that plot.

e.g. Plot 54/1

Data Corrections:

Tree #'s 6372, 6373, and 6374 were incorrectly identified as ARIsr in the 1995 survey. This identification was corrected to HOHglA in the 2005 survey. The species code for these trees has been changed in the 1995 digital file consistent with this correction.

If data from any of the previous measurements are lodged with NVS, please contact the NVS database administrator so changes can be made in the NVS database.

If the identity of the tree remains uncertain, no change to any digital files should be made, and the problem should be noted in the text file for that plot.

e.g. Plot 54/1

Data Problems:

Tree #6450 was identified as CARser in 1995 and as HOHglA in 2005. Confirm identity at next plot measurement.

#### **4.5 What should be done when two stems that were previously measured separately have fused together, and field staff remeasured them as one stem?**

This issue is best addressed at the data-analysis stage. Ensure that the stem state code for both stems is still 'alive' but do not enter a stem diameter. DO NOT GUESS at what its diameter would have been, as data analyses often rely on accurate diameter measurements. The problem must be noted under appropriate headings in the text file.

e.g. Plot 54/1

Data Problems:

CARser 6200 and CARser 6201 stems have fused together. These stems were measured separately in 1995, but together in the 2005 survey. The fused diameter recorded in 2005 was 35.4 cm. No diameter has been entered into the digital file for the 2005 survey.

#### **4.6 What should be done if a tree was missed at a previous plot measurement?**

This problem may be encountered during data checking and analysis. Trees thought to have been missed at a previous plot measurement (e.g. a newly tagged tree that is too large to be a new recruit, or a tagged tree was missed at a previous measurement) are best addressed at the data-analysis stage. DO NOT GUESS at what its diameter would have been, as data analyses often rely on accurate diameter measurements. The problem must be noted under appropriate headings in the text file.

e.g. Plot 54/1

Data problems:

CARser 6211 was measured in 1975 (dbh = 15.2 cm), and 1995 (dbh = 25.6 cm) but was missed in the 1985 survey. No diameter has been entered into the digital file for the 1985 survey.

#### **4.7 What should be done if field staff found two tags on a tree stem during a plot remeasurement?**

On rare occasions a field team may find two tags on an individual tree during plot remeasurement. The history of each tag should be traced to determine which tag has the longest measurement history, and this tag should be retained in the dataset. Alternatively, if one tag is placed at a position that deviates significantly from breast height (notes on the field sheet should indicate whether this is the case) then the other tag should be retained. The tag to be removed should be deleted from all digital files and the change documented in the text file, with a note indicating which tag should be removed by the field team next visiting the plot.

If data from any of the previous measurements are lodged with NVS, please contact the NVS database administrator so changes can be made in the NVS database.

#### **4.8 What should be done if the tree has been tagged but the species name is not recorded, or the tree is tagged and the species recorded but the diameter is missing?**

If a tree has been tagged and measured but the species is unknown, or if a tag and species are listed on the field sheet but no diameter recorded, enter the data along with a detailed note in the text file.

Make a note in the text file directing the next field team to check the situation in the field.

e.g. Plot 54/1

Data Problems:

Tag 6213 was recorded with a diameter of 15.3 cm in 2005, but no species code was recorded. Check the identity of this tree.

---

## 5. Seedling and sapling data-entry guidelines

---

### 5.1 What should be done if there are two entries for a species in a seedling subplot?

Combine the two entries into one record, i.e. sum the tallies of seedlings recorded for each height class, and make a note under an appropriate heading in the text file for the survey.

e.g. General Notes:

When a species was recorded twice on a seedling subplot the records were combined for data entry.

### 5.2 What should be done if there is no data recorded against a species name recorded on a seedling subplot?

Assume that the species was present in the <15-cm tier only. Make a note under an appropriate heading in the text file for the survey.

e.g. General Notes:

When a species was recorded on a seedling subplot but no tick or tally was recorded, the species was entered as a presence in the <15-cm tier.

### 5.3 How should the presence of non-woody species in seedling tiers be entered?

Plot measurement protocols require non-woody species to be recorded as present only (i.e. not counted) when they occur in seedling height-tiers >16-cm. NVS databank convention has previously been to enter the presence of non-woody species in seedling tiers >16-cm as '1', as the PC-USTOREY analysis package requires a numeric value in this field. This procedure should be retained until database changes have been made and new NVS analysis packages are developed. Make a general note in the text file associated with the survey.

e.g. General Notes:

Non-woody species recorded as 'present' in seedling subplots in tiers >16-cm were entered as a '1' in the appropriate tier.

### 5.4 What should be done if there are two entries for a species for a sapling subplot?

Combine the two entries into one record for the sapling subplot, i.e. sum both sapling counts, and make a note under an appropriate heading in the text file for the survey.

e.g. General Notes:

When a species was recorded twice on a sapling subplot the records were added together then entered.

---

## 6. References

---

- Hall GMJ 1992. PC-RECCE: vegetation inventory data analysis. Forest Research Institute Bulletin 182, Christchurch, New Zealand. 108 p.
- Hall GMJ 1994a. PC-DIAM: stem diameter data analysis. Christchurch, Landcare Research. 88 p.
- Hall GMJ 1994b. PC-USTOREY: seedling and sapling data analysis. Christchurch, Landcare Research. 44 p.
- Newell CL, Baldwin AB 2000. Data-standard guidelines for improving the quality of permanent plot data archived in the National Vegetation Survey Databank – first approximation. Landcare Research Contract Report LC9900/140 (unpublished), 40 p.
- Wiser SK, Bellingham PJ, Coomes DA, Burrows LE, Gordon RFS 1999. An assessment of the quality of data stored in the National Vegetation Survey Database, with recommendations for minimising errors. Landcare Research Contract Report LC9899/139 (unpublished), 32 p.