

**Development of an AUSRIVAS model for assessment of  
stream health in the Darwin-Daly region of the Northern  
Territory using family level data**

**FINAL REPORT**

Report to the National River Health Assessment  
Technical and Professional Advisory Committee

December 2005



AUSRIVAS Model Report

December 2005

Authors: Amara Barlow and Gisi Lamche

Project Staff:

Leader: Prof. Richard Norris

Project officer: Amara Barlow

Client contact: Dr. Gisi Lamche

Project Duration: August 2005 to November 2005

Inquires regarding this document should be addressed to:

Dr. Gisi Lamche

Aquatic Ecologist

Water Monitoring Branch, Department of Natural Resources, Environment & the Arts

PO Box 30, Palmerston NT 0831

Ph: + 61 8 8999 3629

Fax: + 61 8 8999 4403

Email: [Gisela.Lamche@nt.gov.au](mailto:Gisela.Lamche@nt.gov.au)

# TABLE OF CONTENTS

<b>1</b>	<b>Introduction</b> .....	<b>1</b>
<b>2</b>	<b>Analyses</b> .....	<b>2</b>
2.1	Data formatting and application of transformations .....	2
2.2	Rare taxa .....	2
2.3	Classification .....	2
2.4	Reference site groups .....	2
2.5	Discriminant Function Analysis.....	4
2.6	Validating reference sites .....	6
2.7	O/E Taxa Bands Representing Biological Health .....	6
<b>3</b>	<b>References</b> .....	<b>6</b>
<b>4</b>	<b>APPENDIX A</b> .....	<b>8</b>

# 1 Introduction

---

A regional AUSRIVAS model for the Darwin Daly region of the Northern Territory had been developed in 2004 for the following reasons: Firstly, it appeared that the early developed territory-wide models were not sensitive enough to detect possible change or impact. Secondly, the Darwin Daly region is of high economic interest within the Northern Territory, which supports a need for regional aquatic health monitoring. The Darwin region is the highest populated and developed part of the Northern Territory and therefore it is considered important to be able to detect changes in water quality by AUSRIVAS. The Daly region is at present in a state of low development, but there is ongoing discussion and proposals for intensive development. A moratorium on approval of development applications for subdivisions and native vegetation clearing is currently in place, however, with a change in policy, this can be overturned.

With the aim to sensitively detect changes in water quality the Darwin Daly model developed in 2004 used mostly genus level data and is hence called the 'Darwin Daly genus level AUSRIVAS model' (Dostine, 2004). For practical reasons it was decided to create a family level model for the Darwin Daly region as this would save resources due to shorter identification times and less expertise needed for identification on family level. A family level model would, however, only be justified, if it displayed a similar sensitivity for detection of impact comparable to the genus level model.

This report details the development of the Darwin Daly region family level model. Once the model is approved and available, the results obtained with both models for all data available will be compared to characterise the sensitivity of the family model in relation to the genus level model. This comparison is still outstanding and not part of this report.

The reference site dataset for the Darwin Daly family level model is based on the data used to create the AUSRIVAS NT Darwin Daly region genus level model (Dostine, 2004). This dataset was created combining samples collected in the early dry season between 1996 and 2002 through several programs and projects. Data stem from the Australian Rivers Assessment System (AUSRIVAS) program and the First National Assessment of River Health (FNARH) stages 1 and 2 of the Monitoring River Health Program in the Northern Territory. Further data were gathered during a Natural Heritage Trust funded project on the water quality in the Mary River catchment, and the ongoing water quality monitoring program of catchments draining into the Darwin Harbour.

Macroinvertebrate specimens mostly identified to genus level were merged at the family level to generate the NT Darwin Daly region family level model presented here. All sites were considered to be in good ecological condition by the absence of catchment disturbances through land clearing, pastoralism, mining impacts and suburban development.

The early dry season was defined from late May to mid August. Most samples were collected over a six week period between late June and early August. The samples were collected over seven years, with most of the data obtained from samples gathered between 1998 and 2000.

## 2 Analyses

---

### 2.1 Data formatting and application of transformations

The raw data was checked for missing values and inappropriate variables. There are a total of 117 sites and 68 macroinvertebrate taxa. The mites were originally at the family level, however, were combined in the one class (Acarina). Total site taxa richness ranges from 12 taxa (sites EE01DW45, EE00MA15 and EE00MC03) to 33 taxa (site EE96FN03). Site locations were converted to 8 character alpha-numeric codes because PC-ORD allows a maximum of 8 characters for row and column names. The location code is comprised of EE (Early dry season and edge habitat); year of sampling; and site code (basin plus site number). Data were transformed to presence/absence data because the AUSRIVAS models only use presence/absence data.

### 2.2 Rare taxa

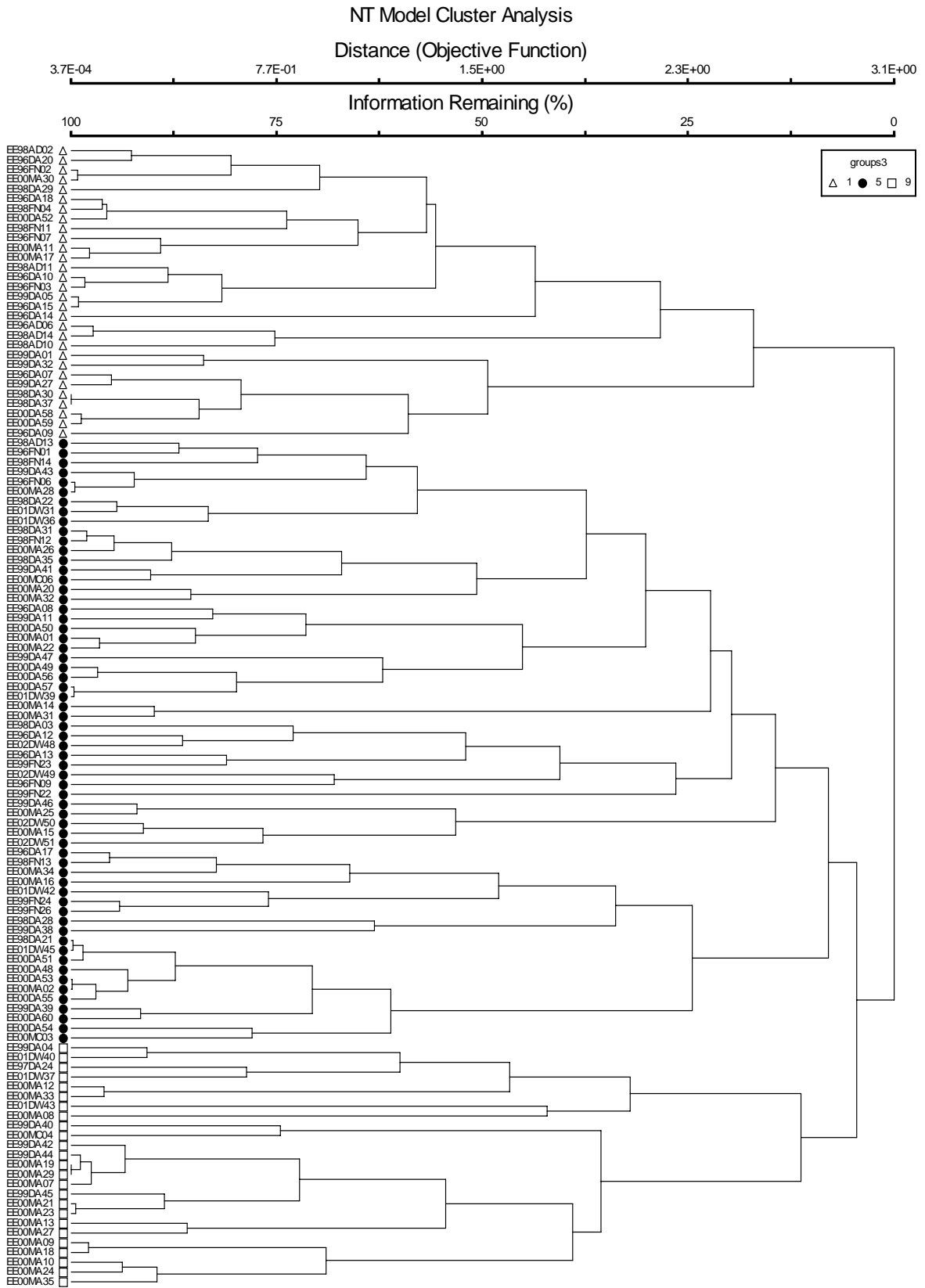
A total of 31 rare taxa were removed from the macroinvertebrate data set. For AUSRIVAS, rare taxa are considered those that occur at less than 10% of sites for data sets of 100 sites or less, or taxa that occur at less than 10 sites for data sets of more than 100 sites. There were no taxa that occurred at exactly 10 sites in the reference data set.

### 2.3 Classification

Sites that had similar macroinvertebrate compositions, based on family level presence/absence data, were determined using cluster analysis in the PC-ORD multivariate analysis package, version 4.20 (McCune and Mefford, 1999). The Bray-Curtis distance measure was used, because it is a robust measure of association for cluster analysis (Faith *et al.*, 1987). The agglomerative clustering technique, flexible Unweighted Pair-Group arithMetic Averaging (UPGMA) was the method used to form reference site groupings. The classifications were viewed as dendrograms allowing the fusion level, which divides sites into groups, to be selected. A beta value of -0.1 was used, as recommended for flexible UPGMA classifications (Belbin and McDonald, 1993).

### 2.4 Reference site groups

Based on the recommendations of Wright *et al.* (1993) groups should contain not less than 5 sites. Small classification groups should be either deleted from further analysis or these sites amalgamated with another group of appropriate reference sites. Three groups of sites were recognised based on macroinvertebrate family level data (Fig. 1).



**Figure 1. Classification of sites in the Darwin-Daly region of the Northern Territory. Three groups of sites were recognised based on macroinvertebrate family level data (presence/absence data).**

## 2.5 Discriminant Function Analysis

The AUSRIVAS models use habitat features (predictor variables) from a site to predict which taxa should occur at that site in the absence of environmental damage. Habitat variables that are commonly affected by human induced impacts should not be used as predictor variables. Variables such as turbidity, dissolved oxygen and phosphorus concentrations are often affected by human impacts and might provide spurious predictions if used to predict the membership of test sites to the reference site groups. In contrast, habitat features such as altitude, distance from source, latitude and longitude often make good predictor variables because they are rarely affected by human impacts.

The reference site groups from the UPGMA classification were entered as a variable into the reference habitat data set and a Stepwise Multiple Discriminant Function Analysis (MDFA) performed to select the predictor variables used in the AUSRIVAS model. The Stepwise MDFA was performed in the SAS statistical package version 6.12 (SAS Institute 1995). This procedure selected a subset of habitat variables that best discriminated among the groups of sites formed from the faunal classifications. The stepwise procedure included habitat variables one at a time, selecting at each step the variables that gave the best group discrimination. At each step of the analysis the significance of variables already included were checked and variables that were no longer significant were removed. The significance levels for variables to enter and be retained by the Stepwise MDFA both were set at 0.05.

The subsets of habitat variables from the Stepwise MDFA were then tested in a MDFA to predict the probabilities of group membership for a reference site. Biased discriminations were avoided by using the cross-validation option that predicts group membership of each site separately. A subset of habitat variables that produced the lowest error in predicting the group membership of reference sites was obtained from this procedure. However, the actual value of the misclassification error is not critical because AUSRIVAS uses all the probabilities of a site belonging to each group for site predictions, rather than the allocation to a single group as performed by the cross-validation procedure. Thus, sites with an affinity for two or more groups can be misclassified but still provide adequate predictions for a model.

The subsets of habitat variables obtained from the stepwise MDFA are used as predictor variables for the AUSRIVAS model under construction. The predictor variables and the reference site invertebrate-classification form the foundation of AUSRIVAS, allowing predictions of the taxa expected at new "test" sites. The list of predictor variables, their descriptions and the misclassification errors for the NT family model are provided in Tables 1 and 2.

**Table 1: Predictor variables and misclassification errors for the NT family model.**

<b>Cross-validation Misclassification Rate (%)</b>	<b>Resubstitution Misclassification Rate (%)</b>	<b>Predictor Variables</b>
40.19	37.20	AVGWID, LNGUDE, LATUDE, RIP500

**Table 2: AUSRIVAS codes, description and units of predictor variables used for the NT family model.**

Predictor variable code	Description and units
LATUDE*	WGS84Lat (decimal degrees)
LNGUDE*	WGS84Long (decimal degrees)
DAYNUM	Day of the year with 1st January being day 1
DSTSRC	Distance from source (kilometres)
CATARE	Catchment area (square kilometres)
LOGCAT	Log 10 catchment area
STRORD	Stream order
AVGWID*	Average of 3-6 measurements of stream width of the reach
ALKNTY	Alkalinity (Ca CO <sub>3</sub> derived in the laboratory)
AVGVEL	Average of 3 measurements of water velocity (metres / second)
SDVTOP	Standard Deviation DEM Window of 5X5 cell window (derived from GIS coverages - Standard deviation of elevations from 25 cells centred on collecting site derived from digital elevation model)
RNGTOP	Range DEM 5x5 Window (derived from GIS coverages - difference between min and max elevation of elevations from 25 cells centred on collecting site derived from digital elevation model)
ELVATN	Elevation (m) WGS84 derived from DEM (ArcInfo)
RIP100	Area of riparian rainforest in 100m radius of site
SPR100	Area of spring rainforest in 100m radius of site
TOT100	Total area of rainforest in 100m radius of site
RIP250	Area of riparian rainforest in 250m radius of site
SPR250	Area of spring rainforest in 250m radius of site
TOT250	Total area of rainforest in 250m radius of site
RIP500*	Area of riparian rainforest in 500m radius of site
SPR500	Area of spring rainforest in 500m radius of site
TOT500	Total area of rainforest in 500m radius of site
RIP1000	Area of riparian rainforest in 1000m radius of site
SPR1000	Area of spring rainforest in 1000m radius of site
TOT1000	Total area of rainforest in 1000m radius of site

\* Selected as predictor variables for the model

## 2.6 Validating reference sites

The full reference invertebrate and habitat data set were used to create the first version of the model. The same reference data set was then entered into the model as if they were test sites to validate reference site condition. Reference sites with observed to expected taxa ratios (O/E ratios) below 0.75 were reviewed by the NT Water Monitoring Branch. No sites were rejected as reference sites for model development. The O/E ratio output at this stage is considered to represent the distribution of ecological health for the population of reference sites.

## 2.7 O/E Taxa Bands Representing Biological Health

To simplify interpretation, AUSRIVAS presents the O/E taxa ratios as bands that represent different levels of biological condition. The widths of the bands are based on the distribution of the O/E taxa values for the reference sites of each particular model. Test sites that fall between the 10<sup>th</sup> and 90<sup>th</sup> percentiles are considered equivalent to reference condition (band A). Impaired test sites will fall into a band equivalent to the severity of impact that the site is experiencing. The next two bands (B and C), which represent increasing levels of impairment, are the same width as Band A. Band D will vary in width depending on the variability of the reference O/E taxa values used to create the model. Test sites with biological condition richer than reference condition (i.e. > 90<sup>th</sup> percentile) are placed in Band X. The bands used for the new AUSRIVAS model are listed in Table 3.

**Table 3. O/E taxa bands for the new AUSRIVAS Darwin Daly region Family model.**

Model	O/E Taxa Bands				
	X	A	B	C	D
NT Family	> 1.18	0.82 - 1.18	0.45 - 0.81	0.08 - 0.44	0 - 0.07

## 3 Conclusion

An AUSRIVAS family level model for the Darwin Daly region of the NT has been developed and validated and is presented in this report. The Water Monitoring Branch will use the model and compare results with the Darwin Daly region genus level model. If the new family level model is of comparable sensitivity to detect changes and impairment of water quality it will be a valuable tool for regular water monitoring and enable reduction of resources due to less time and expertise needed for macroinvertebrate identification.

## 4 References

Belbin, L. and McDonald, C. (1993). Comparing three classification strategies for use in ecology. *Journal of Vegetation Science*, **4**: 341-348.

Dostine P (2004): A predictive model for assessment of stream health in the Darwin Daly region of the Northern Territory using mostly genus level data. Report 42/2004D, Natural Resource

Management Division, Conservation and Natural Resources Group, Department of Infrastructure, Planning & Environment, Darwin NT, 27 p.

Faith, D.P., Minchin, P.R. and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, **69**: 57-68.

McCune, B. and M.J. Mefford. (1999). Multivariate Analysis of Ecological Data Version 4.20. MjM Software, Gleneden Beach, Oregon, U.S.A.

Wright, J.F., Furse, M.T., Armitage, P.D., and Moss, D. (1993). New procedures for identifying running-water sites subject to environmental stress and for evaluating sites for conservation, based on the macroinvertebrate fauna. *Arch. Hydrobiologia*, **127(3)**: 319-326.

## 5 APPENDIX A

---

**Sites used in the model creation and their group membership as determined by macroinvertebrate classification.**

<b>SITE CODE</b>	<b>GROUP</b>	<b>LATITUDE</b>	<b>LONGITUDE</b>
EE00DA48	2	-14.624	132.594
EE00DA49	2	-14.6686	131.6827
EE00DA50	2	-14.0767	131.8144
EE00DA51	2	-13.79	131.3524
EE00DA52	1	-13.7873	131.2971
EE00DA53	2	-13.7892	131.2983
EE00DA54	2	-14.186	131.3938
EE00DA55	2	-14.1377	131.4227
EE00DA56	2	-13.8095	131.3412
EE00DA57	2	-13.8128	131.3427
EE00DA58	1	-13.862	131.1855
EE00DA59	1	-13.9998	131.231
EE00DA60	2	-13.7808	131.3853
EE00MA01	2	-13.6794	131.9801
EE00MA02	2	-13.5973	132.2096
EE00MA07	3	-13.6352	131.9968
EE00MA08	3	-13.5287	132.0498
EE00MA09	3	-13.5271	132.0487
EE00MA10	3	-13.5279	132.05
EE00MA11	1	-13.4795	132.0201
EE00MA12	3	-13.6836	132.0213
EE00MA13	3	-13.5982	132.1326
EE00MA14	2	-13.5281	132.0621
EE00MA15	2	-13.5411	132.209
EE00MA16	2	-13.5307	132.0613
EE00MA17	1	-13.3383	131.9487
EE00MA18	3	-13.3461	131.9593
EE00MA19	3	-13.3477	131.9559
EE00MA20	2	-13.3327	131.9386
EE00MA21	3	-13.3265	131.9397
EE00MA22	2	-13.4832	132.0221
EE00MA23	3	-13.474	132.0213
EE00MA24	3	-13.5968	132.2081
EE00MA25	2	-13.6382	132.312

<b>SITE CODE</b>	<b>GROUP</b>	<b>LATITUDE</b>	<b>LONGITUDE</b>
EE00MA26	2	-13.6362	132.3004
EE00MA27	3	-13.6423	132.3087
EE00MA28	2	-13.5797	132.2465
EE00MA29	3	-12.9249	131.6868
EE00MA30	1	-13.2733	131.9114
EE00MA31	2	-13.2157	131.8984
EE00MA32	2	-13.1557	131.8716
EE00MA33	3	-12.7227	131.7423
EE00MA34	2	-12.8128	131.604
EE00MA35	3	-12.8052	131.6587
EE00MC03	2	-13.6762	131.7322
EE00MC04	3	-13.5321	131.7038
EE00MC06	2	-13.2653	131.7115
EE01DW31	2	-12.7094	131.0034
EE01DW36	2	-12.8005	130.9199
EE01DW37	3	-12.831	130.9087
EE01DW39	2	-12.6435	131.095
EE01DW40	3	-12.6133	131.0805
EE01DW42	2	-12.5554	131.1379
EE01DW43	3	-12.4623	131.0856
EE01DW45	2	-12.5266	131.1233
EE02DW48	2	-13.0766	130.699
EE02DW49	2	-13.1151	130.7862
EE02DW50	2	-13.0935	130.7833
EE02DW51	2	-13.1229	130.6667
EE96AD06	1	-13.0787	131.2507
EE96DA07	1	-13.8615	131.0741
EE96DA08	2	-13.8135	130.7148
EE96DA09	1	-14.0708	131.2517
EE96DA10	1	-13.7858	131.3628
EE96DA12	2	-14.1665	132.0725
EE96DA13	2	-13.9654	131.6953
EE96DA14	1	-13.744	131.5722
EE96DA15	1	-13.8403	131.1485
EE96DA17	2	-13.745	131.0972
EE96DA18	1	-14.2335	130.9142
EE96DA20	1	-14.1602	132.7092
EE96FN01	2	-13.2519	130.7427
EE96FN02	1	-13.5386	130.8669

<b>SITE CODE</b>	<b>GROUP</b>	<b>LATITUDE</b>	<b>LONGITUDE</b>
EE96FN03	1	-13.4823	130.805
EE96FN06	2	-13.4057	130.7883
EE96FN07	1	-13.281	130.7033
EE96FN09	2	-13.2297	130.9492
EE97DA24	3	-13.543	133.2282
EE98AD02	1	-13.4258	131.0865
EE98AD10	1	-13.2052	131.1547
EE98AD11	1	-13.0852	131.2403
EE98AD13	2	-13.3508	131.1355
EE98AD14	1	-13.184	131.21
EE98DA03	2	-14.3252	132.4128
EE98DA21	2	-13.7646	133.0852
EE98DA22	2	-13.6847	133.2085
EE98DA28	2	-14.4363	132.2768
EE98DA29	1	-14.462	132.2567
EE98DA30	1	-14.519	132.194
EE98DA31	2	-14.3747	132.3873
EE98DA35	2	-14.0112	132.7438
EE98DA37	1	-14.5482	132.1295
EE98FN04	1	-13.2805	130.7044
EE98FN11	1	-13.05	130.9947
EE98FN12	2	-12.913	130.715
EE98FN13	2	-13.022	130.9512
EE98FN14	2	-13.1636	130.6835
EE99DA01	1	-14.3627	131.5567
EE99DA04	3	-14.3007	132.4186
EE99DA05	1	-14.7573	131.5946
EE99DA11	2	-14.187	132.1718
EE99DA27	1	-14.1048	131.274
EE99DA32	1	-14.6525	132.0185
EE99DA38	2	-14.2942	131.8692
EE99DA39	2	-14.1119	131.7888
EE99DA40	3	-14.9261	131.8775
EE99DA41	2	-14.702	132.071
EE99DA42	3	-13.6758	133.2122
EE99DA43	2	-13.4157	133.1895
EE99DA44	3	-13.6482	132.9668
EE99DA45	3	-13.7197	132.9005
EE99DA46	2	-13.879	132.7812

---

<b>SITE CODE</b>	<b>GROUP</b>	<b>LATITUDE</b>	<b>LONGITUDE</b>
EE99DA47	2	-13.659	131.4003
EE99FN22	2	-13.0027	131.0297
EE99FN23	2	-12.9892	131.0265
EE99FN24	2	-12.9725	130.9525
EE99FN26	2	-12.9698	130.9502

---