

Notes referring to spreadsheet “NZTABS_DataSet_Sept22_2011.xlsx”

Heading Key:

Column A ‘UNIQUE.CODE’: Unique numeric identifier for the sample.

Column B ‘FIELD.ID’: Label given to sample in the field.

Column C ‘SORT.ID’: Tile and microhabitat from which the sample was collected.

Column D ‘GEOLOGY’: Dominant geology of the sample site.

Column E ‘ASPECT.DIRECTION’: Cardinal direction that the slope of the sample site faces.

Column F ‘TILE’: The tile from which the sample was collected.

Column G ‘MICROHABITAT.NUMBER’: Numeric designation of the microhabitat from which the sample was collected.

Column H ‘MICROHABITAT.CL1’: Description of the microhabitat.

Column I ‘LAT’: Latitude, in decimal degrees, of the sampling site.

Column J ‘LONG’: Longitude, in decimal degrees, of the sampling site.

Column K ‘LAT.GIS’: Latitude, in decimal degrees, of the location from which the GIS data (columns M through BT) was derived.

Column L ‘LONG.GIS’: Longitude, in decimal degrees, of the location from which the GIS data (columns M through BT) was derived.

Columns M through BT: See document ‘TABS GIS Spatial Joined Metadata.docx’

Column BU ‘N.PCT’: Percentage of nitrogen in the soil sample.

Column BV ‘C.PCT’: Percentage of carbon in the soil sample.

Column BW ‘CN’: Percentage of carbon in the soil divided by the percentage of nitrogen in the soil sample.

Column BX ‘AGG.SOIL.WATER’: Percentage of water in the soil sample (gram/gram).

Column BY ‘ACTIVE.WATER’: Active water in the soil sample.

Column BZ ‘AGG.CONDUCTIVITY’: Conductivity of the soil sample in millisiemens.

Column CA 'AGG.PH': pH of the soil sample.

Column CB 'VEG.SURVEY': Area covered in vegetation survey, expressed in square meters.

Column CC 'AGG.CYANO.COVERAGE': Units of cyanobacterial mat observed in a vegetation survey at the study site.

Column CD 'CYANO.PCT': Percentage of cyanobacterial mat coverage observed in vegetation survey at the study site.

Column CE 'AGG.MOSS.COVERAGE': Units of moss observed in a vegetation survey at the study site.

Column CF 'MOSS.PCT': Percentage of moss cover observed in vegetation survey at the study site.

Column CG 'AGG.LICHEN.COVERAGE': Units of lichen observed in a vegetation survey at the study site.

Column CH 'LICHEN.PCT': Percentage of lichen cover observed in vegetation survey at the study site.

Column CI 'AGG.SPRINGTAILS': Number of springtails observed in a 10 minute survey at the study site.

Column CJ 'AGG.MITES': Number of mites observed in a 10 minute survey at the study site.

Column CK 'AGG.ATP': Amount of ATP (adenosine triphosphate) measured in 0.2 g of the soil sample.

Column CL 'SOIL.RESP.RATE': Respiration rate measured in the soil sample, expressed in nanograms of carbon per gram of soil per day.

Column CM 'BIOMASS': Biomass in the soil sample, expressed in milligrams of carbon per gram of soil.

Column CN 'AGG.DNA': Amount of DNA recovered from the soil sample, expressed in nanograms of DNA per gram of soil.

Column CO 'BACT.S': Estimated number of bacterial phylotypes present in the soil sample based on bacterial community fingerprinting (ARISA).

Column CP 'BACT.J': Estimated evenness of bacterial phylotypes present in the soil sample based on bacterial community fingerprinting (ARISA). Calculated using the equation:

$$J = H/\log(S)$$

Where H is the Shannon index of bacterial diversity (see column CQ) and S is the estimated number of bacterial phylotypes (see column CO).

Column CQ 'BACT.SHANNON': Shannon's index of bacterial diversity in the soil sample, based on bacterial community fingerprinting (ARISA). Calculated using the equation:

$$H = -\sum(P_i \ln(P_i))$$

Where P_i is the proportion of the i th phylotype.

Column CR 'BACT.SIMPSON': Simpson's estimate of bacterial diversity in the soil sample, based on bacterial community fingerprinting (ARISA). Calculated using the equation:

$$D = 1/\sum(P_i^2)$$

Where P_i is the proportion of the i th phylotype.

Column CS 'FUNGAL.S': Estimated number of fungal phylotypes present in the soil sample based on fungal community fingerprinting (ARISA).

Column CT 'FUNGAL.J': Estimated evenness of fungal phylotypes present in the soil sample based on fungal community fingerprinting (ARISA). Calculated using the equation:

$$J = H/\log(S)$$

Where H is the Shannon index of fungal diversity (see column CU) and S is the estimated number of fungal phylotypes (see column CS).

Column CU 'FUNGAL.SHANNON': Shannon's index of fungal diversity in the soil sample, based on fungal community fingerprinting (ARISA). Calculated using the equation:

$$H = -\sum(P_i \ln(P_i))$$

Where P_i is the proportion of the i th phylotype.

Column CV 'FUNGAL.SIMPSON': Simpson's estimate of fungal diversity in the soil sample, based on fungal community fingerprinting (ARISA). Calculated using the equation:

$$D = 1/\sum(P_i^2)$$

Where P_i is the proportion of the i th phylotype.

Column CW 'CYANO.S': Estimated number of cyanobacterial phylotypes present in the soil sample based on cyanobacterial community fingerprinting (ARISA).

Column CX 'CYANO.J': Estimated evenness of cyanobacterial phylotypes present in the soil sample based on cyanobacterial community fingerprinting (ARISA). Calculated using the equation:

$$J = H/\log(S)$$

Where H is the Shannon index of cyanobacterial diversity (see column CY) and S is the estimated number of cyanobacterial phylotypes (see column CW).

Column CY 'CYANO.SHANNON': Shannon's index of cyanobacterial diversity in the soil sample, based on cyanobacterial community fingerprinting (ARISA). Calculated using the equation

$$H = -\sum(P_i \ln(P_i))$$

Where P_i is the proportion of the i th phylotype.

Column CZ 'CYANO.SIMPSON': Simpson's estimate of cyanobacterial diversity in the soil sample, based on cyanobacterial community fingerprinting (ARISA). Calculated using the equation

$$D = 1/\sum(P_i^2)$$

Where P_i is the proportion of the i th phylotype.

Column DA 'SCOTTNEMA.M.LIVE': Number of live Scottnema males observed per kilogram of dry soil sample.

Column DB 'SCOTTNEMA.M.DEAD': Number of dead Scottnema males observed per kilogram of dry soil sample.

Column DC 'SCOTTNEMA.F.LIVE': Number of live Scottnema females observed per kilogram of dry soil sample.

Column DD 'SCOTTNEMA.F.DEAD': Number of dead Scottnema females observed per kilogram of dry soil sample.

Column DE 'SCOTTNEMA.JUV.LIVE': Number of live Scottnema juveniles observed per kilogram of dry soil sample.

Column DF 'SCOTTNEMA.JUV.DEAD': Number of dead Scottnema juveniles observed per kilogram of dry soil sample.

Column DG 'SCOTTNEMA.ADULTS.LIVEANDDEAD': Number of live and dead Scottnema adults observed per kilogram of dry soil sample.

Column DH 'SCOTTNEMA.JUV.LIVEANDDEAD': Number of live Scottnema juveniles observed per kilogram of dry soil sample.

Column DI 'SCOTTNEMA.TOTAL.LIVE': Total number of live Scottnema observed per kilogram of dry soil sample.

Column DJ 'SCOTTNEMA.TOTAL.DEAD': Total number of dead Scottnema observed per kilogram of dry soil sample.

Column DK 'SCOTTNEMA.TOTAL.LIVEANDDEAD': Total number of live and dead Scottnema observed per kilogram of dry soil sample.

Column DL 'EUDORYLAIMUS.M.LIVE': Number of live Eudorylaimus males observed per kilogram of dry soil sample.

Column DM 'EUDORYLAIMUS.M.DEAD': Number of dead Eudorylaimus males observed per kilogram of dry soil sample.

Column DN 'EUDORYLAIMUS.F.LIVE': Number of live Eudorylaimus females observed per kilogram of dry soil sample.

Column DO 'EUDORYLAIMUS.F.DEAD': Number of dead Eudorylaimus females observed per kilogram of dry soil sample.

Column DP 'EUDORYLAIMUS.JUV.LIVE': Number of live Eudorylaimus juveniles observed per kilogram of dry soil sample.

Column DQ 'EUDORYLAIMUS.JUV.DEAD': Number of dead Eudorylaimus juveniles observed per kilogram of dry soil sample.

Column DR 'EUDORYLAIMUS.ADULT.LIVEANDDEAD': Number of live and dead Eudorylaimus adults observed per kilogram of dry soil sample.

Column DS 'EUDORYLAIMUS.JUV.LIVEANDDEAD': Number of live Eudorylaimus juveniles observed per kilogram of dry soil sample.

Column DT 'EUDORYLAIMUS.TOTAL.LIVE': Total number of live Eudorylaimus observed per kilogram of dry soil sample.

Column DU 'EUDORYLAIMUS.TOTAL.DEAD': Total number of dead Eudorylaimus observed per kilogram of dry soil sample.

Column DV 'EUDORYLAIMUS.TOTAL.LIVEANDDEAD': Total number of live and dead Eudorylaimus observed in a soil sample.

Column DW 'PLECTUS.F.LIVE': Number of live Plectus females observed per kilogram of dry soil sample.

Column DX 'PLECTUS.F.DEAD': Number of dead Plectus females observed per kilogram of dry soil sample.

Column DY 'PLECTUS.JUV.LIVE': Number of live Plectus juveniles observed per kilogram of dry soil sample.

Column DZ 'PLECTUS.JUV.DEAD': Number of dead Plectus juveniles observed per kilogram of dry soil sample.

Column EA 'PLECTUS.ADULTS.LIVEANDDEAD': Number of live and dead Plectus adults observed per kilogram of dry soil sample.

Column EB 'PLECTUS.JUV.LIVEANDDEAD': Number of live and dead Plectus juveniles observed per kilogram of dry soil sample.

Column EC 'PLECTUS.TOTAL.LIVE': Total number of live Plectus observed per kilogram of dry soil sample.

Column ED 'PLECTUS.TOTAL.DEAD': Total number of dead Plectus observed per kilogram of dry soil sample.

Column EE 'PLECTUS.TOTAL.LIVEANDDEAD': Total number of live and dead Plectus observed per kilogram of dry soil sample.

Column EF 'ROTIFERS.TOTAL': Total number of rotifers observed per kilogram of dry soil sample.

Column EG 'TARDIGRADES.TOTAL': Total number of tardigrades observed per kilogram of dry soil sample.

Column EH 'NEMATODES.LIVE': Total number of live nematodes observed per kilogram of dry soil sample.

Column EI 'NEMATODES.DEAD': Total number of dead nematodes observed per kilogram of dry soil sample.

Column EJ 'NEMATODES.LIVEANDDEAD': Total number of live and dead nematodes observed per kilogram of dry soil sample.

Column EK 'GEOLOGY.COMBINE': Dominant geology of the sample site with 9 categories, pooled as discussed in October 2011 TABS modeling meeting.

Column EL 'GEOMICRO.COMBINE': Dominant geology and microhabitat of the sample site with 9 categories, pooled as discussed in October 2011 TABS modeling meeting.

Column EM 'ROCK.OR.SEDIMENT': Categorization of sample site as rock or sediment.

Column EN 'X.CORD.M': Spatial X coordinate in metres.

Column EO 'Y.CORD.M': Spatial Y coordinate in metres

Column EP 'CTIGS10MAX.R30': 30 metre radius maximum filter run over the CTIGS10 GIS layer.

Column EQ 'CTIGS10MAX.R50': 50 metre radius maximum filter run over the CTIGS10 GIS layer.

Column ER 'TAXA.RICHNESS': Number of taxa present at the site from the following 9 categories: lithics, cyanobacterial mat, moss, lichen, springtails, mites, rotifers, tardigrades, and nematodes. Score ranges from 0-9.

Column ES 'AGG.LITHIC': Units of lithic communities observed in a vegetation survey at the study site.

Column ET 'LITHIC.PCT': Percentage of lithic community cover observed in vegetation survey at the study site.

General information regarding spreadsheet and formatting:

1. All sample rows that had only survey data, but no biological data were removed. This sheet does not include any data for samples with Unique Codes 408, 455, 480, 634, 812, 814, 815, 816, 819, 830, 835, 837, 839, 840, 844, 849, 853, 854 because the majority of data was missing for these samples.
2. 'NA': Any data that was not available and for which no estimates were made to fill in the missing data. Any cells with 'NA' are coloured yellow.
3. 'NOT': Cells for which a value could not be assigned because the result of the calculation is undefined. Any cells with 'NOT' are coloured orange.
4. Cells for which no data was available, but for which an estimate was made in order to fill in the missing data, are coloured green. Specifics of the calculations used to fill in each missing value are included as a comment in the corresponding cell in the spreadsheet.

General methodology for data cleaning and filling of missing values:

GIS Data (columns K through BT)

1. Refer to file 'TABS GIS Spatial Joined Metadata.docx' for the meaning of the headings and specific information about the data listed in columns K through BT.
2. If two samples have the same SORT_ID, and data was only available for one, then the data was duplicated for the sample with missing data (should be from the same geographic site and have very similar site data associated).
3. All data entered as '-9999' was replaced with 'NA' (40 instances).

Percent Carbon and Percent Nitrogen (columns BU to BV):

1. For duplicate samples, i) I duplicated the data if there was a value for one sample but not the other and ii) I averaged the values for the two samples if two sets of data were available (geometric mean).
2. For remaining missing data, I took the geometric mean of the values for all samples from the same geology, microhabitat and elevation interval (200 m interval used to stratify tiles). In some cases there were not enough samples to calculate a meaningful mean in terms of all three of these factors, so used one or two of them (see note associated with each missing data point for details).

Aggregate Soil Water (column BX)

1. Took geometric mean of all samples from sites with the same geology and microhabitat (only 2 were missing).

Active water (column BY)

1. Back calculated from Agg.Water.Content. Act water appears to increase logarithmically with Agg.Water.Cont to a max of 1. Split the Agg.Water.Content by geology and took the mean of the active water value associated with the 5 Agg.Water.Cont values on either side of the missing value. Exception was schist, where there was not 5 values on either side, so used mean of 5 closest values.

Conductivity and pH data (columns BZ to CA)

1. For duplicate samples, i) I duplicated the data if there was a value for one sample but not the other and ii) I averaged the values for the two samples if two sets of data were available (mean).
2. For remaining missing data, I took the geometric mean of the values for all samples from the same geology, microhabitat and elevation interval (200 m interval used to stratify tiles). In some cases there were not enough samples to calculate a meaningful mean in terms of all three of these factors, so used one or two of them (see note associated with each missing data point for details).

Vegetation survey and mite and springtail counts (columns CB to CK and ES-ET):

1. For duplicate samples, i) I duplicated the data if there was a value for one sample but not the other and ii) averaged the values for the two samples if two sets of data were available (geometric mean).
2. Values assigned zeros if no vegetation was noted in Nov'09 site surveys.

Biomass data (columns CL to CN)

1. For duplicate samples, i) I duplicated the data if there was a value for one sample but not the other and ii) averaged the values for the two samples if two sets of data were available (mean).
2. For remaining missing data, I took the geometric mean of the values for all samples from the same geology, microhabitat and elevation interval (200 m interval used to stratify tiles). In some cases there were not enough samples to

calculate a meaningful mean in terms of all three of these factors, so used one or two of them (see note associated with each missing data point for details).

Microbiological data (columns CO to CZ):

1. For duplicate samples, i) I duplicated the data if there was a value for one sample but not the other and ii) averaged the values for the two samples if two sets of data were available (mean).
2. Values assigned NA if not enough DNA recovered to proceed with the fingerprinting.

Invertebrate Data (all except mites and springtails) (columns DA to EJ)

1. All sample data deleted if there was no consistency in sites with similar characteristics.
2. Median values of variables from sites with similar characteristics.

Columns added after October 2011 TABS modeling meeting (EK to ET)

1. Note that all these columns are based on other pre-existing columns that may have had missing data, and this influence may not be noted in the new columns. This note does not apply to columns ES and ET (lithics), which are based on the vegetation surveys.
2. Columns ES and ET, missing data filled in as noted in 'Vegetation survey and mite and springtail counts' above.