

Assessment of Aberrant Levels



EPHC

Environment Protection & Heritage Council

This paper was presented
at the Fifth National
Workshop on the Assessment
of Site Contamination

Proceedings of the Fifth National Workshop on the Assessment of Site Contamination

Editors: Langley A, Gilbey M and Kennedy B

The editors may be contacted through the NEPC Service Corporation for which contact details are provided below.



DISCLAIMER: This document has been prepared in good faith exercising due care and attention. However, no representation or warranty, express or implied, is made as to the relevance, accuracy, completeness or fitness for purpose of this document in respect of any particular user's circumstances. Users of this document should satisfy themselves concerning its application to, and where necessary seek expert advice about, their situation. The Environment Protection and Heritage Council, the National Environment Protection Council, the NEPC Service Corporation, Environment Australia or enHealth shall not be liable to the purchaser or any other person or entity with respect to liability, loss or damage caused or alleged to have been caused directly or indirectly by this publication.

Suggested policy directions and health and environment values presented in papers comprising these proceedings have not been endorsed by the Environment Protection and Heritage Council, the National Environment Protection Council, Environment Australia nor enHealth.

Further copies of these proceedings can be purchased from:

NEPC Service Corporation
Level 5, 81 Flinders Street
ADELAIDE SA 5000

Phone: (08) 8419 1200
Facsimile: (08) 8224 0912
Email: exec@ephc.gov.au

© National Environment Protection Council Service Corporation 2003

Printed version ISBN 0-642-32355-0 Electronic (web) ISBN 0-642-32371-2

This work is copyright. It has been produced by the National Environment Protection Council (NEPC). Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior permission from the NEPC, available from the NEPC Service Corporation. Requests and enquiries concerning reproduction and rights should be addressed to the Executive Officer, NEPC Service Corporation, Level 5, 81 Flinders Street, ADELAIDE SA 5000.

Printed on environmentally-friendly recycled content paper.

Assessment of Aberrant Levels

Costa Diomides¹, Ray Correll² and Ravi Naidu²

¹Victoria University, ²CSIRO Land and Water

1 INTRODUCTION

This paper does not try to be a review of methods and definitions used in the study of background values. It does however attempt to bring some novel concepts to the definition of background levels. In some cases, we attempt to define background levels in a direct sense, but in other cases we seek to demonstrate where there are non-background values in a data set that purports to be purely background values.

Background levels differ from site to site, and a suitable definition of background levels must be chosen to be fit for purpose. Such a purpose may be to define safe levels, to detect contamination or perhaps to detect naturally elevated concentration levels of one or more chemical constituent(s).

The definition of background must also include the measurement of interest. Often this measurement will be the concentration of a heavy metal, but there are many other possibilities. The list of these potential contaminants includes

- heavy metals
- hydrocarbons
- anthropogenic pesticides
- pH
- nutrients
- radioactivity and radionucleotides
- microbes.

In some cases (for example pesticides) the compounds are anthropogenic, and the natural background level should be zero. A definition of zero, however, is not feasible as that can never be validated – it may be necessary to define background in such a case as being less than some very low concentration.

In many cases, it is easier to say concentrations differ from (typically exceeding) background levels, rather than saying a soil is at background level. Naturally occurring concentrations will depend *inter alia* on the underlying geology. These changes may vary locally with changes in soil type. The changes may be gradual or quite abrupt. At times the concentration of various metals will be strongly correlated; this is particularly so in mineralised areas. Correll and Taylor (1973) give a description of such a site.

The comments made in this paper about background levels and the detection of outliers will be illustrated using heavy metal concentrations from two data sets. The first data represent some ostensibly background soils collected from the western suburbs of Melbourne by the first author of this paper. A subset of the data is given in Appendix A: the subset includes all soils where there are complete data for Cr, Cu, Pb, Ni and Zn for both A and B horizons.

The paper describes several ways to determine background (or more often what is not background) and then discusses these methods. The methodology is intended to assist in determining whether concentrations of a substance are approximately homogenous, and is not intended to be used to 'explain away' situations where a uniformly high level of contamination has occurred. The paper thus seeks to promote debate rather than to give a clear definition of background concentrations.

2 DETERMINATION OF BACKGROUND LEVEL

There are several methods of estimating background concentrations. These are described below.

2.1 EXCEEDENCE OF A DEFINED LIMIT

Perhaps the simplest method of defining background levels is by reference to some standard. Such a method is prescriptive and easily applied. The problem arises in setting such a limit. One method could be to collect a large number of samples from apparently undisturbed sites, and to select some percentile (say 95%) as the upper limit of background.

Several problems arise with this approach. Some of these problems are:

- The location of undisturbed sites; many sites that appear to meet this criterion have received some degree of contamination (e.g. dust from soil blown from fertilised fields). An example of this is described by Jones *et al.* (1986).
- Background concentrations vary with the soil type. A single criterion for background concentrations cannot be usefully applied across the full range of soil concentrations.
- Local mineralization will result in locally elevated concentrations of some elements. Depending on the definition of background used, some allowance should be made for local mineralization.

Application of the approach detailed in this paper will require extensive databases.

Some background concentrations values are available, for example, in the *Marine Pollution Bulletin*.

2.2 MINIMUM VALUE METHOD

A simple method of determining is to use the lowest mean concentration that has been observed in some arbitrary set of boundaries. The method fails to take into account the natural variability within a study area leading to the adoption of a background level that is relevant to only a small part of the study area. The minimum value method is subject to a statistical bias, and its reliability is difficult to quantify.

2.3 REMOTE SITE METHODS

A common approach used in the study of a contaminated area is to use a remote site as a reference. This is a simple approach but does not address the problems described above. In some cases the remote site data may be quite misleading. An example is in the study of the contamination of the marine sediments near Port Pirie (Ward and Correll, 1992). In that study, Edithburgh was used as a remote reference site. However, some Cd levels found at the remote site far exceeded those at some of the study sites.

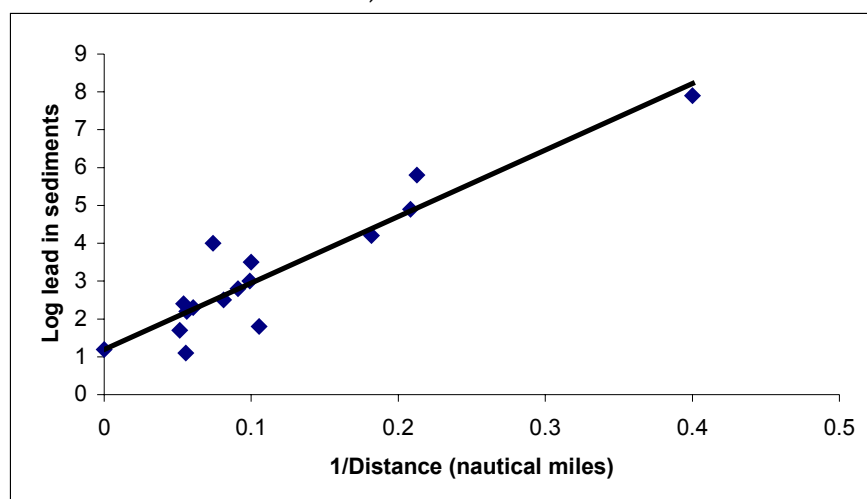
2.4 EXTRAPOLATION TO ZERO

The extrapolation to zero method is described in Ward and Correll (1992). The method is applicable where contamination is from a point source and no pre-contamination data are available. In that case, a linear function based on the reciprocal of the distance from the contaminant source (or some other suitable function) can be used to describe the concentrations of the contaminant. In that case, the intercept represents a location infinitely distanced from the contamination. Such a location can be considered as background.

The method is illustrated in Figure 1, where the intercept corresponds to the background concentration in the sediment.

Approximate confidence intervals are available for the intercept using standard statistical theory. These can be readily translated into confidence intervals for the background concentration.

Figure 1. Estimation of background concentrations using the extrapolation to zero (approximate data from Ward and Correll 1992)



3 DETECTION OF DEPARTURE FROM BACKGROUND

In this section it is implicitly assumed that the background concentrations can be described by some distribution. Outliers from this distribution then represent a departure from the background concentration.

3.1 UNIVARIATE DETECTION OF OUTLIERS

The presence of contamination is often manifested as an extreme value of the concentration of a particular element. This is demonstrated in Figure 2, where the ranked concentrations of lead in the surface soil are plotted against the normal expected deviates. In that case there is approximately a linear relationship for most of the data, suggesting that the lead concentrations have an approximately normal distribution. There is one clear outlier (soil 13), perhaps indicating some contamination of that soil.

The outlier is less obvious when the data are plotted on a log scale as shown in Figure 3. The method does have limitations. For example, Figure 4 shows the corresponding plot of the lead concentrations from a heavily contaminated site (Correll, 2001). Those data

showed an approximately log-normal distribution apart from three outliers. However, after discarding the 3 obvious outliers, the last point that was on the line had a lead concentration of 2100 mg/kg. Testing for an outlier would not have detected that value as differing from background. An alternative way of viewing the data was that background lead for that site included values as high as 2100 mg/kg.

Figure 2. Concentration of lead plotted against the normal expected deviate (data given in Appendix A)

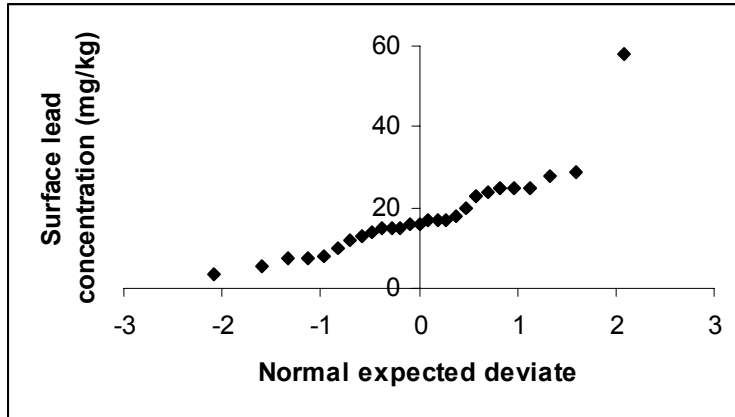


Figure 3. Concentration of lead (using a log scale) plotted against the normal expected deviate (data given in Appendix A)

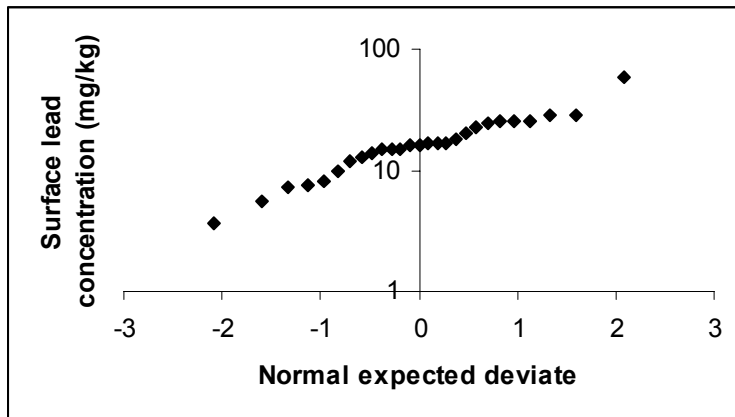
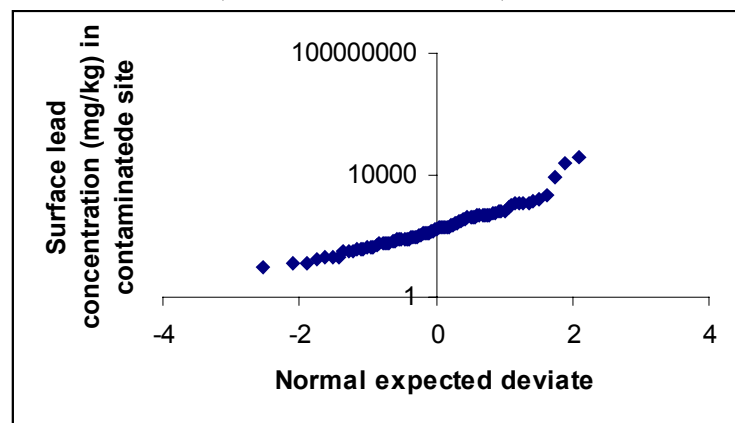


Figure 4. Concentration of lead (using a log scale) plotted against the normal expected deviate from a contaminated site (data from Correll, 2001)



3.2 MULTIVARIATE DETECTION OF OUTLIERS

A common method of describing soil properties is principal component analysis (PCA). PCA displays as much variation among the series of soil samples as is possible in a small number of components. It does this by choosing linear combinations of the soil concentrations to create the PCA scores. In its simplest case, the soil can be described using only a single component.

We applied this technique to the surface values test data set. The first PCA explained 56% of the total variation among the soils. The loadings are given in Table 1. What is of interest is that the loadings all have the same sign. In this case the first principal component score is a weighted average of all the metal concentrations. Note that the data were standardized as part of the PCA analysis to have zero mean and unit variance, and the analysis was performed on the correlation matrix. This is standard theory and can be found in texts such as Mardia, Kent and Bibby (1979).

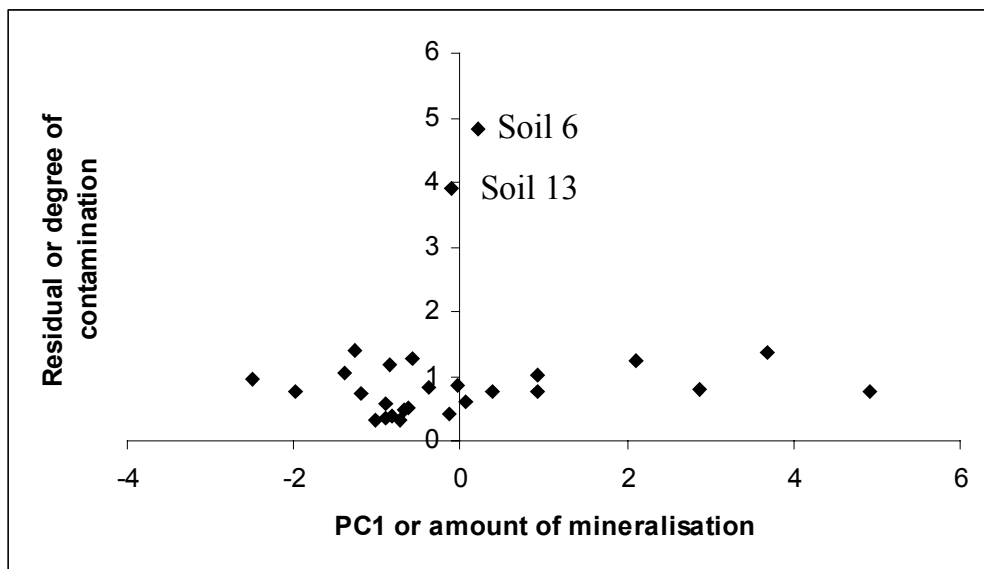
The PCA analysis can also be used to detect outliers. If the mineralisation can be represented by a small number of principal components, then deviations from that description (ie the residuals) represent some anomaly associated with that point. These anomalies could indicate soil contamination. The converse, namely a lack of outliers, does not mean that the site is not contaminated.

In the example given (Figure 5), there were two clear outliers. Soil 6 had an extreme value of zinc of 240 compared to the next highest value of 60 mg/kg in soil 27.

Table 1. Loadings for first principal component for surface soil in test data set

Metal	Loading
Chromium	0.56007
Copper	0.56828
Lead	0.17448
Nickel	0.54042
Zinc	0.20221

Figure 5. Display of the first principal component scores and the corresponding residual on the surface soil data from the test data set.



3.3 COMPARISON OF A AND B HORIZONS

Typically there is a strong correlation between the concentrations of elements in the surface and B horizons. This will not always be so – especially in sedimentary soils. The correlation does, however, offer a useful tool for detecting outliers in many situations.

Figure 6 illustrates the relationship between surface and B horizon Cr concentrations. The relationship is strong ($R^2 = 0.568$, $P < 0.001$) and there are no clear outliers. This supports the notion that there is no Cr contamination in the area represented by the data given in Appendix A.

Figure 6 Comparison of surface and B horizon Cr concentrations

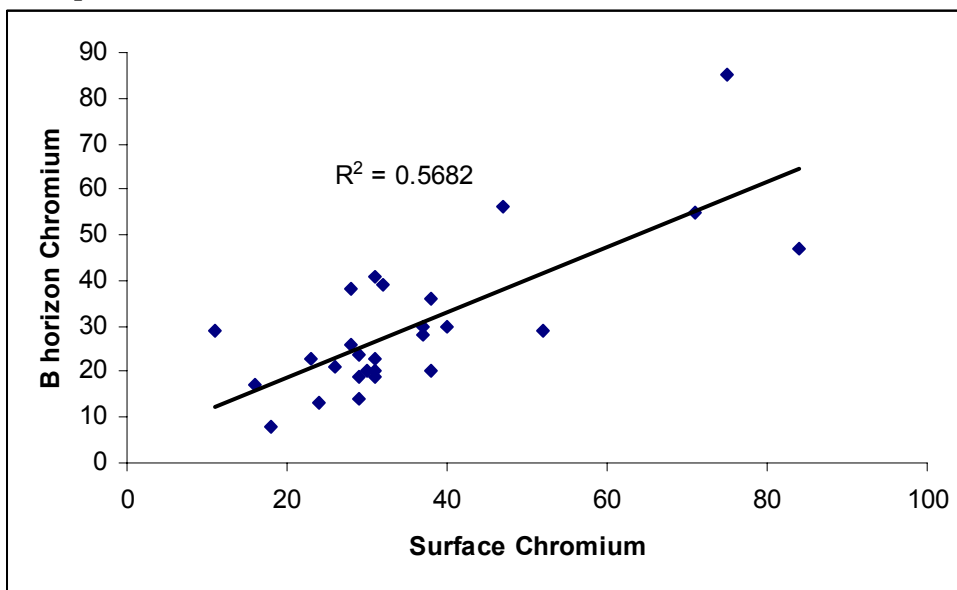


Figure 7 Comparison of surface and B horizon Zn concentrations

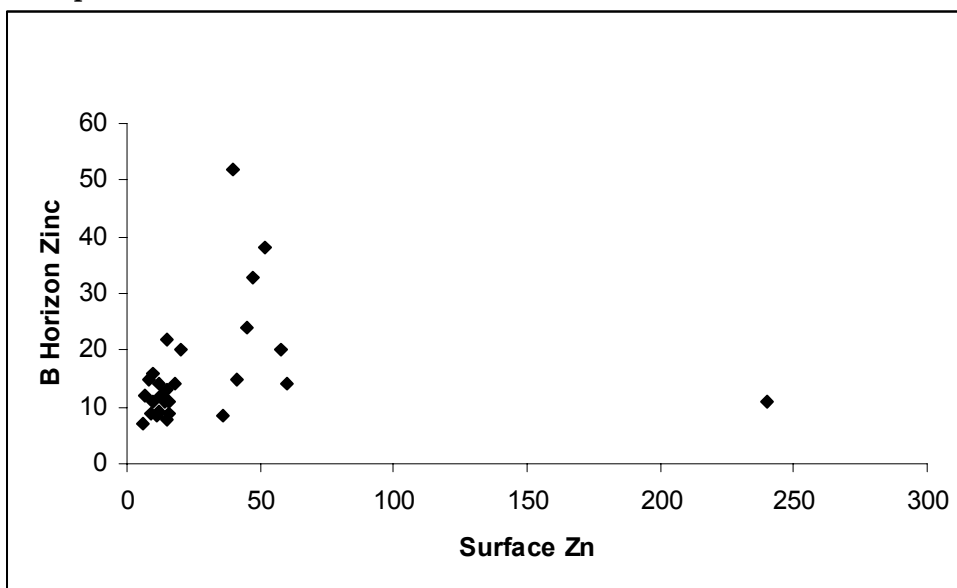
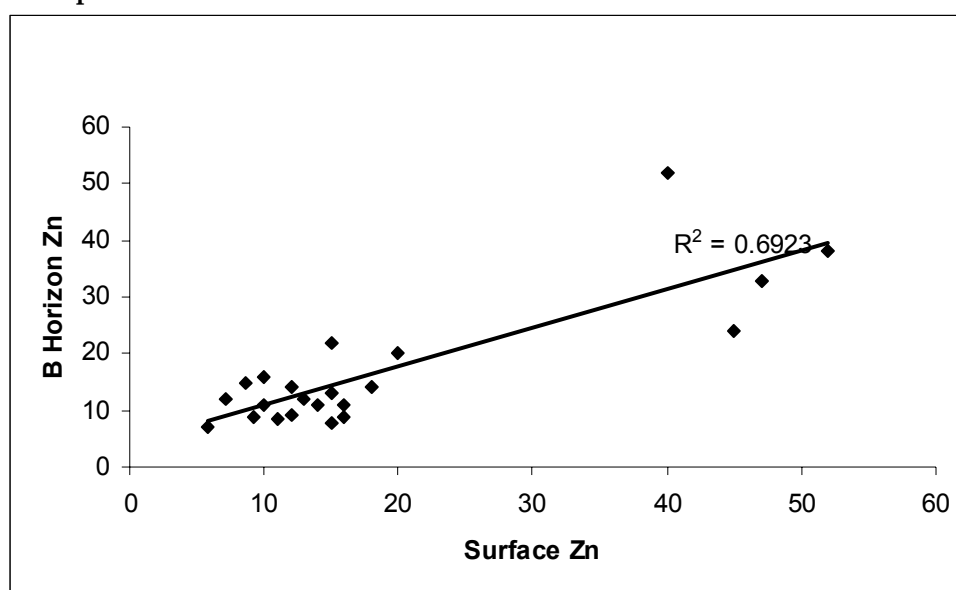


Figure 7 shows clearly that one of the surface values of Zn far exceeds the B horizon concentration. This outlier is soil 6. There are several other values where the concentration in the surface soil far exceeds the concentration in the B horizon. A total of 5 points had to be removed before the relationship displayed in Figure 8 was achieved. This approach suggests that 5 of the 27 soils described in Appendix A may have been contaminated on the surface with Zn.

We note in passing that a multivariate generalisation of the comparison of the surface and B horizons could be undertaken. This would involve describing the surface and B horizons with separate PCA, and comparing them using a Procrustes rotation (Gower, 1985). Large residuals from the Procrustes rotation would indicate different patterns between the samples taken at the two horizons at a single sampling point. Such a difference may indicate some contamination.

Figure 8. Comparison of surface and B horizon Zn concentrations after the removal of 5 outliers



4 CONCLUSIONS

Background concentrations may be defined directly using

- Reference to some standard
- Minimum value method
- Quantile from a local data base
- Method of extrapolation to zero
- Multivariate methods (for example principal component analysis) are available to give a more flexible approach to the definition of background.

In many cases it is easier to define points that are not background. Methods available for this include

- Normal plots
- Deviations from principal component representations of the data
- Contrasts between surface and B horizon concentrations.

ACKNOWLEDGEMENTS

Data for the Western suburbs of Melbourne were compiled from Environmental Audit reports (Contaminated land) prepared by independent environmental auditors. Access to these reports was obtained from the Environment Protection Authority - Land and Groundwater unit. The authors thank Mr Warren Müller for his constructive comments.

REFERENCES

- Correll, R.L. (2001). The use of composite sampling in contaminated sites - a case study. *Environmental and Ecological Statistics* 8 (3) 185-200
- Correll, R. L., and Taylor, R. G. (1973). An occurrence of *Polycarpeae spirostylis* at the Daisy Bell Copper Mine, Emuford, Queensland, Australia. *Trans. I.M.M.* 83, B30-B34.
- Gower, J. C. (1985). Multivariate analysis: ordination, multidimensional scaling and allied topics. In: Handbook of applied mathematics (ed. W. Ledermann), Statistics Vol VIB (ed. E.Lloyd). Wiley, Chichester.
- Jones, K. C.; Davies, B. E.; Peterson, P. J. Silver in Welsh Soils: Physical and Chemical Distribution Studies. *Geoderma* **1986**, 37, 157-174.
- Mardia, K.V., Kent, J.T. and Bibby, J.N. (1979). Multivariate analysis. Academic Press, London.
- Ward, T. J. and Correll, R. L. (1992) Estimating background levels of heavy metals in the marine environment. in *Proceedings of Bioaccumulation Workshop*¹, Sydney 1990. A.G. Miskiewicz ed. Sydney Water Board.

APPENDIX A TEST DATA SET

Sample ID	Surface					B Horizon				
	Chromium	Copper	Lead	Nickel	Zinc	Chromium	Copper	Lead	Nickel	Zinc
1	29	12	13	31	20	23.5	11	12	25.5	20
2	31	9.2	14	17	10	20	7.1	13	9.4	11
3	37	11	15	21	18	30	12	18	22	14
4	31	13	25	0	16	19	7.8	12	11	8.9
5	11	4.4	3.7	0	7.1	29	10	13	16	12
6	31	11	25	0	240	23	8.1	12	12	11
7	40	13	15	19	13	30	12	9.9	24	12
8	31	7.6	20	12	9.2	41	13	13	32	8.8
9	28	9	15	22	15	26	9.4	12	25	7.8
10	29	12	24	10	36	19	7.6	14	23	8.3
11	26	7	23	7.2	15	21	8	12	24	13
12	30	10	17	12	12	20	3.3	9.7	5.2	9.1
13	23	8.6	58	18	41	23	9.1	15	33	15
14	24	8.6	16	14	11	13	6.7	12	15	8.5
15	28	9.9	17	12	8.6	38	11	13	24	15
16	38	16	17	20	14	20	5	12	9.1	11
17	29	9.2	18	9.5	12	14	4.1	9.7	3.8	14
18	71	20	28	42	58	55	13	10	38	20
19	75	24	12	64	40	85	28	9.9	92	52
20	47	17	16	21	47	56	19	11	45	33
21	38	4.6	5.6	9.7	10	36	9	12	22	16
22	37	6.2	7.4	18	16	28	6	9	15	11
23	18	4.7	7.2	6.7	5.8	8	4.7	6.3	7.5	7.1
24	16	7	8	20	15	17	10	11	28	22
25	52	20	10	50	45	29	8.4	5	24	24
26	32	17	25	32	52	39	14	6	31	38
27	84	25	29	80	60	47	5.9	0	20	14

