Appendix 8: Statistical details

8.1 The specification of a simple method for using an isotope profile to exclude locations from those that may be a source of the sample that produced the profile, see final report section 4.5.2.3

"Database >95%"

The database has 2816 rows. Each row provides the following data about a district-level postcode.

1   Unlabelled row number

2   OBJECTID: a code

3   District: district postcode

4,5   X, Y: easting and northing

6,7   Longitude, Latitude

8   EXPC1: expected value of PC1 at this location

9   EXPC2: expected value of PC2 at this location

10   EXPC3: expected value of PC3 at this location

11   EXPC4: expected value of PC4 at this location

12   H: Expected H isotope ratio at this location

13   C: Expected C isotope ratio at this location

14   N: Expected N isotope ratio at this location

15   S: Expected S isotope ratio at this location

16   sePC1: Estimated variation expressed as a standard deviation around EXPC1

17   sePC2: Estimated variation expressed as a standard deviation around EXPC2

18   sePC3: Estimated variation expressed as a standard deviation around EXPC3

19   sePC4: Estimated variation expressed as a standard deviation around EXPC4

"Database 80"

This has the same structure as "Database >95%". Estimated variations around expected principal component scores are based on the median rather than 95 percentile of prediction standard errors, observed during model fitting.

Constants used in process

mH -99.27395
mC= -25.54976
mN= 7.291326
mS= 6.161561

vH= 4.514928
vC= 1.619716
vN= 1.063242
vS= 2.755421

$$V = \begin{pmatrix} -0.5728024 & 0.3159510 & 0.4872710 & 0.5784802 \\ -0.5707021 & -0.4072466 & 0.3383457 & -0.6276715 \\ -0.3434787 & 0.7431636 & -0.4170670 & -0.3946964 \\ 0.4777227 & 0.4266532 & 0.6885814 & -0.3400057 \end{pmatrix}$$

CRIT95= 9.487729   this is the 95th percentile of the Chi-squared distribution with 4 degrees of freedom

CRIT80= 5.988617   this is the 80th percentile of the Chi-squared distribution with 4 degrees of freedom

User input

Choose between confidence=">95%" and confidence="80%", enter HCNS isotope ratio results.

Process

The process has five steps.

1   Get results: H, C, N, and S for the sample

2   Calculate the standardised values

$$Hs = (H - mH)/vH$$

$$Cs = (C - mC)/vC$$

$$Ns = (N - mN)/vN$$

$$Ss = (S - mS)/vS$$

3   Calculate the principle components scores using the matrix multiplication

$$(PC1 \quad PC2 \quad PC3 \quad PC4) = (Hs \quad Cs \quad Ns \quad Ss).V$$

If matrix multiplication is not available, the principle components can be calculated manually using each element of **V**: V[row, column].

$$PC1 = Hs.V[1,1] + Cs.V[2,1] + Ns.V[3,1] + Ss.V[4,1]$$

$$PC2 = Hs.V[1,2] + Cs.V[2,2] + Ns.V[3,2] + Ss.V[4,2]$$

$$PC3 = Hs.V[1,3] + Cs.V[2,3] + Ns.V[3,3] + Ss.V[4,3]$$

$$PC4 = Hs.V[1,4] + Cs.V[2,4] + Ns.V[3,4] + Ss.V[4,4]$$

4   Then for each line i of **beefraster with extrapolated points.csv if confidence=">95" or beefraster with extrapolated points 80.csv if confidence="80"],** calculate a score, score[i].

$$score[i] = \left(\frac{PC1 - EXPC1[i]}{sePC1[i]}\right)^2 + \left(\frac{PC2 - EXPC2[i]}{sePC2[i]}\right)^2 + \left(\frac{PC3 - EXPC3[i]}{sePC3[i]}\right)^2 + \left(\frac{PC3 - EXPC3[i]}{sePC3[i]}\right)^2$$

5   Return a list of locations from the database: District[i], X[i], Y[i], Longitude[i], Latitude[i] which meet the condition.

If confidence = ">95"

$$score[i] > CRIT95$$

If confidence = "80"

$$score[i] > CRIT80$$

The sample is assessed as NOT being from two-letter postcodes for which all districts meet the condition $score[i] > CRITxx$

8.2 Details of hydrogen (a) and carbon 9b) equation, see final report section 4.5.2.5

<u>Details of (a) – hydrogen equation</u>

Formula: y ~ K + A * sin(2 * pi * (308/365) * xt + X)

Parameters:
```
        Estimate Std. Error t value Pr(>|t|)
        K -0.9953    0.1888  -5.270 2.30e-07 ***
        A -0.9996    0.2538  -3.938 9.79e-05 ***
        X  0.3055    0.2512   1.216   0.225
```

<u>Details of (a) – carbon equation</u>

Formula: y ~ K + A * sin(2 * pi * (308/365) * xt + X)

Parameters:
```
        Estimate Std. Error t value Pr(>|t|)
        K  0.14994   0.08039  1.865  0.0629 .
        A  0.42186   0.10687  3.947 9.42e-05 ***
        X  1.70981   0.25622  6.673 8.96e-11 ***
```