

Geometric morphometrics

Geometric morphometric data structure

Manuel F. G. Weinkauff

Univerzita Karlova, Prague, Czech Republic

26–27 August 2022



Section 1

Introduction

The problem of morphological data

- Data from traditional morphometrics are easily stored in tabular form
- Geometric morphometric data include x - y -(z)-coordinate sets for several points (outline or landmark coordinates) per specimen

The problem of morphological data

An example

Traditional morphometrics

Specimen	Length	Width	...
Sp 1	12	4	...
Sp 2	14	5	...
...	...	...	...
Sp x	11	5	...

The problem of morphological data

An example

Traditional morphometrics

Specimen	Length	Width	...
Sp 1	12	4	...
Sp 2	14	5	...
...	...	...	...
Sp x	11	5	...

**All data fit easily within a 2 D
tabular structure**

The problem of morphological data

An example

Traditional morphometrics

Specimen	Length	Width	...
Sp 1	12	4	...
Sp 2	14	5	...
...	...	...	...
Sp x	11	5	...

All data fit easily within a 2 D tabular structure

Geometric morphometrics

Specimen 1

x	y
12	14
16	8
7	3
...	...

Specimen 2

x	y
12	14
16	8
7	3
...	...

The problem of morphological data

An example

Traditional morphometrics

Specimen	Length	Width	...
Sp 1	12	4	...
Sp 2	14	5	...
...	...	...	...
Sp x	11	5	...

All data fit easily within a 2 D tabular structure

Geometric morphometrics

Specimen 1

x	y
12	14
16	8
7	3
...	...

Specimen 2

x	y
12	14
16	8
7	3
...	...

Each specimen alone requires a 2 D tabular structure. Multiple specimens would require a 3 D array that is difficult to structure in reality.

Section 2

Major file formats for geometric morphometrics

.nts file format

Origins and intention

- Written for the FORTRAN program NTSYSpc by James Rohlf in 1966
- ASCII-encoded file, semi-human-readable
- Mainly used for outline data, but could be used for landmark data and other data types as well
- .nts file extension

.nts file structure

```
1 10L 10 0 dim=2
Sp1 Sp2 Sp3 Sp4 Sp5 Sp6 Sp7 Sp8 Sp9 Sp10
0.386 0.745 0.793 0.001 0.459 0.844 0.093 ...
0.846 0.279 0.904 0.627 0.702 0.910 0.393 ...
0.987 0.237 0.848 0.396 0.234 0.772 0.753 ...
0.550 0.789 0.557 0.432 0.809 0.129 0.381 ...
0.691 0.163 0.119 0.990 0.333 0.856 0.432 ...
0.266 0.021 0.434 0.427 0.443 0.573 0.979 ...
...
```

.nts file structure

```
1_10L_10_0_dim=2
```

```
Sp1_Sp2_Sp3_Sp4_Sp5_Sp6_Sp7_Sp8_Sp9_Sp10
```

```
0.386_0.745_0.793_0.001_0.459_0.844_0.093_...
```

```
0.846_0.279_0.904_0.627_0.702_0.910_0.393_...
```

```
0.987_0.237_0.848_0.396_0.234_0.772_0.753_...
```

```
0.550_0.789_0.557_0.432_0.809_0.129_0.381_...
```

```
0.691_0.163_0.119_0.990_0.333_0.856_0.432_...
```

```
0.266_0.021_0.434_0.427_0.443_0.573_0.979_...
```

```
...
```

.nts file structure

First row: Data structure

Data structure code

- 1st number: Matrix type
- 2nd and 3rd numbers: Number of rows and columns, followed by 'L' if row/column names are provided
- 4th number: Either 0 if no values are missing, or '1 XXX' if values are missing ('XXX' then encodes the value for 'NA')
- 'dim=2' or 'dim=3' for 2 D or 3 D data, respectively

```
1 10L 10 0 dim=2
```

```
Sp1 Sp2 Sp3 Sp4 Sp5 Sp6 Sp7 Sp8 Sp9 Sp10
0.386 0.745 0.793 0.001 0.459 0.844 0.093 ...
0.846 0.279 0.904 0.627 0.702 0.910 0.393 ...
...
```

.nts file structure

Matrix types

1	Rectangular data matrix
2	Symmetric dissimilarity matrix
3	Symmetric similarity matrix
4	Diagonal matrix
5	Tree matrix for dissimilarity data
6	Tree matrix for similarity data
7	Graph matrix for dissimilarity data
8	Graph matrix for similarity data

.nts file structure

Matrix types

1	Rectangular data matrix
2	Symmetric dissimilarity matrix
3	Symmetric similarity matrix
4	Diagonal matrix
5	Tree matrix for dissimilarity data
6	Tree matrix for similarity data
7	Graph matrix for dissimilarity data
8	Graph matrix for similarity data

.nts file structure

Optional second and/or third row: Row and column names

Row and column names

- An 'L' after the second or third number in row 1 indicates the presence of row and/or column names respectively
- These are given in rows 2 and/or 3

```
1 10L 10 0 dim=2
```

```
Sp1 Sp2 Sp3 Sp4 Sp5 Sp6 Sp7 Sp8 Sp9 Sp10
```

```
0.386 0.745 0.793 0.001 0.459 0.844 0.093 ...
```

```
0.846 0.279 0.904 0.627 0.702 0.910 0.393 ...
```

```
...
```

.nts file structure

Data block

Data block structure

- Each row is one specimen
- Within each row, the data are given as
x1 y1 (z1) x2 y2 (z2) x3 y3 (z3) ...

```
1 10L 10 0 dim=2
```

```
Sp1 Sp2 Sp3 Sp4 Sp5 Sp6 Sp7 Sp8 Sp9 Sp10
```

```
      x1      y1      x2      y2      x3      y3      x4      ...
```

```
Sp1 0.386 0.745 0.793 0.001 0.459 0.844 0.093 ...
```

```
Sp2 0.846 0.279 0.904 0.627 0.702 0.910 0.393 ...
```

```
...
```

.tps file format

Origins and intention

- Written for the tps-series of morphometric programs by James Rohlf, Fred Bookstein, and Bill Green in the 1990s
- ASCII-encoded file, human-readable
- Mostly used for landmark data; can also be used for outline data but will get very long
- .tps file extension

.tps file structure

LM=5

0.386 0.745

0.793 0.001

0.459 0.844

0.093 0.162

0.420 0.059

IMAGE=Image1.jpg

ID=Sp1

LM=5

0.846 0.279

...

IMAGE=Image2.jpg

ID=Sp2

...

.tps file structure

LM=5

0.386_0.745

0.793_0.001

0.459_0.844

0.093_0.162

0.420_0.059

IMAGE=Image1.jpg

ID=Sp1

LM=5

0.846_0.279

...

IMAGE=Image2.jpg

ID=Sp2

...

.tps file structure

Landmark data

Specimen landmark data structure

- Each new specimen starts with line 'LM=XXX', with 'XXX' as number of landmarks
- This is followed by a block of landmark coordinates
 - Each row is one landmark
 - Within each row, the data are given as 'x y (z)'

LM=5

0.386 0.745

...

0.420 0.059

IMAGE=Image1.jpg

ID=Sp1

...

.tps file structure

Metadata

Specimen metadata structure

- Each specimen can contain the following metadata
 - 'IMAGE=XXX', with 'XXX' as image name
 - 'ID=XXX', with 'XXX' as specimen ID
 - 'SCALE=XXX', with 'XXX' as scaling factor used for this specimen's landmark coordinates

LM=5

0.386 0.745

...

0.420 0.059

IMAGE=Image1.jpg

ID=Sp1

...

Section 3

Other file formats for geometric morphometrics

PAST morphological data

PAST morphological data structure

- Used by the program PAST
- Simplified NTS file format, containing only the data block
 - Each row is one specimen
 - Within each row, the data are given as
 x1 y1 (z1) x2 y2 (z2) x3 y3 (z3) ...

0.386	↯0.745	↯0.793	↯0.001 ...
0.846	↯0.279	↯0.904	↯0.627 ...
...			

IMP morphological data

IMP morphological data structure

- Used by the program IMP6
- Simplified NTS file format, containing only the data block and the centroid size as last column
 - Each row is one specimen
 - Within each row, the data are given as
x1 y1 x2 y2 x3 y3 ... cs

0.386_0.745_0.793_0.001_..._9

0.846_0.279_0.904_0.627_..._12

...

Section 4

File creation and conversion

Geometric morphometric files

- Most programs that you can use to generate morphometric data have either a default output format or let you choose an output format during export of your data
- A convenient tool to transform between different morphometric data file types is available on my GitHub repository: https://github.com/WeinkMFG/R/blob/master/MorphoFiles_Function.r