Geometric morphometrics

Morphometric data quality assessment

Manuel F. G. Weinkauf

Univerzita Karlova, Prague, Czech Republic

26–27 August 2022

Section 1

Introduction

Sources of error

- Greatest error is human factor
  - Landmarks are positioned manually → error in entire configuration
  - For outlines the starting point is chosen manually → error may propagate through all semi-landmarks
  - Different measurers and their measurement environment lead to deviations

Sources of error

- Greatest error is human factor
  - Landmarks are positioned manually → error in entire configuration
  - For outlines the starting point is chosen manually → error may propagate through all semi-landmarks
  - Different measurers and their measurement environment lead to deviations
- Measurement devices have precision, and images have resolution limiting precision

Sources of error

- Greatest error is human factor
  - Landmarks are positioned manually → error in entire configuration
  - For outlines the starting point is chosen manually → error may propagate through all semi-landmarks
  - Different measurers and their measurement environment lead to deviations
- Measurement devices have precision, and images have resolution limiting precision
- Object orientation is difficult and small deviances can accumulate

Sources of error

- Greatest error is human factor
  - Landmarks are positioned manually → error in entire configuration
  - For outlines the starting point is chosen manually → error may propagate through all semi-landmarks
  - Different measurers and their measurement environment lead to deviations
- Measurement devices have precision, and images have resolution limiting precision
- Object orientation is difficult and small deviances can accumulate
- Different magnifications lead to variations in precision of data extraction

Replication as error quantification

- Replication of measurements for each error factor allows to quantify the error associated with this factor
  1. **Landmark positions/outline starting point position**: Extract the morphometric data for each specimen at least twice, better three times

Replication as error quantification

- Replication of measurements for each error factor allows to quantify the error associated with this factor
  1. **Landmark positions/outline starting point position**: Extract the morphometric data for each specimen at least twice, better three times
  2. **Difference between measurers**: Each measurer extracts morphometric data for each specimen

Replication as error quantification

- Replication of measurements for each error factor allows to quantify the error associated with this factor
  1. **Landmark positions/outline starting point position**: Extract the morphometric data for each specimen at least twice, better three times
  2. **Difference between measurers**: Each measurer extracts morphometric data for each specimen
  3. **Object orientation**: Reorient each specimen at least once and repeat morphometric data extraction

Replication as error quantification

- Replication of measurements for each error factor allows to quantify the error associated with this factor
  1. **Landmark positions/outline starting point position**: Extract the morphometric data for each specimen at least twice, better three times
  2. **Difference between measurers**: Each measurer extracts morphometric data for each specimen
  3. **Object orientation**: Reorient each specimen at least once and repeat morphometric data extraction
  4. **Magnification/object size differences**: Repeat measurements for different magnifications

Replication as error quantification

- In practice, it is unfeasible to replicate all measurements on all error-levels due to the huge time effort that would require
- **Solution**: Replicate randomly across all error-factors
  1. **Landmark positions/outline starting point position**: Repeat data extraction in a randomly chosen subset of the specimens

Replication as error quantification

- In practice, it is unfeasible to replicate all measurements on all error-levels due to the huge time effort that would require
- **Solution**: Replicate randomly across all error-factors
  1. **Landmark positions/outline starting point position**: Repeat data extraction in a randomly chosen subset of the specimens
  2. **Difference between measurers**: Each measurer extracts morphometric data for same specimens in a randomly chosen subset from the sample

Replication as error quantification

- In practice, it is unfeasible to replicate all measurements on all error-levels due to the huge time effort that would require
- **Solution**: Replicate randomly across all error-factors
  1. **Landmark positions/outline starting point position**: Repeat data extraction in a randomly chosen subset of the specimens
  2. **Difference between measurers**: Each measurer extracts morphometric data for same specimens in a randomly chosen subset from the sample
  3. **Object orientation**: Reorient each specimen from a randomly chosen subset of the sample at least once and repeat morphometric data extraction

Replication as error quantification

- In practice, it is unfeasible to replicate all measurements on all error-levels due to the huge time effort that would require
- **Solution**: Replicate randomly across all error-factors
  1. **Landmark positions/outline starting point position**: Repeat data extraction in a randomly chosen subset of the specimens
  2. **Difference between measurers**: Each measurer extracts morphometric data for same specimens in a randomly chosen subset from the sample
  3. **Object orientation**: Reorient each specimen from a randomly chosen subset of the sample at least once and repeat morphometric data extraction
  4. **Magnification/object size differences**: Replicate measurements explicitly in sample with overall smallest and largest specimens and test if errors differ significantly

Section 2

Relative error in morphometrics

The ANOVA design for percent error estimation

- Calculates the relative error associated with any error-factor for which replication was conducted ⇒ allows error comparison between studies that used entirely different data
- Based on an **AN**alysis **O**f **VA**riances
- Developed by Yezerinac, Loogheed, and Handford (1992) Measurement error and morphometric studies: Statistical power and observer experience. *Syst. Biol.* 41: 471–82

Within- and among-measurement component

- Based on the mean squares (MSS) of a one-way ANOVA

Within-measurement component

$$s^2_{\text{within}} = \text{MSS}_{\text{within}}$$

Among-measurement component

$$s^2_{\text{among}} = \frac{\text{MSS}_{\text{among}} - \text{MSS}_{\text{within}}}{m}$$

with m as number of replications

Percent measurement error

- Percent measurement error ($\%ME$) is calculated based on individual components

$$\%ME = \frac{s^2_{\text{within}}}{s^2_{\text{within}} + s^2_{\text{among}}}$$

Example

- ANOVA of 2 replications from 20 specimens

	df	MSS
Replication	19	18.32
Residuals	20	2.73

$$\mathrm{MSS}_{\mathrm{within}} = s^2_{\mathrm{within}} = 2.73$$

$$\mathrm{MSS}_{\mathrm{among}} = 18.32$$

$$s^2_{\mathrm{among}} = \frac{18.32 - 2.73}{2} = 7.795$$

$$\%ME = \frac{2.73}{2.73 + 7.795} \times 100 = \mathbf{25.93824}\,\%$$

R

Example of error analysis

For a look at morphometric error analysis in R, we move on to exercise № 6

Open the exercise sheet for instructions and code examples