

## RESEARCH ARTICLE

## Realising the Promise of Large Data and Complex Models

# Constructing a multiple-part morphospace using a multiblock method

Daniel B. Thomas<sup>1</sup>  | Aaron M. T. Harmer<sup>1</sup>  | Simone Giovanardi<sup>1</sup>  |  
 Emma J. Holvast<sup>1</sup>  | Cushla M. McGoverin<sup>2,3</sup>  | Arthur Tenenhaus<sup>4</sup> 

<sup>1</sup>School of Natural Sciences, Massey University, Auckland, New Zealand

<sup>2</sup>Department of Physics, University of Auckland, Auckland, New Zealand

<sup>3</sup>The Dodd-Walls Centre for Photonic and Quantum Technologies, Auckland, New Zealand

<sup>4</sup>Laboratoire des Signaux et Systèmes, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

**Correspondence**

Daniel Thomas

Email: [d.b.thomas@massey.ac.nz](mailto:d.b.thomas@massey.ac.nz)

**Funding information**

Massey University

**Handling Editor:** Laura Graham

**Abstract**

1. Popular current methods for quantifying variation in biological shape are well-suited to analyses of isolated parts (e.g. the same bone from the skeletons of many individuals). An analytical challenge exists for quantifying variation between the shapes of multiple-part objects where each part has a different position, rotation or scale (e.g. partial or whole articulated skeletons). We investigated regularised consensus principal component analysis (RCPCA) as a multiblock method for quantifying variation in the shape of multiple-part objects. Multiblock methods are routinely used in other big data research fields such as bioinformatics/medicine, marketing and food research, but have not been widely embraced for evolutionary biology research.
2. We have created the new package `MORPHOBLOCKS` for the R programming language to make RCPCA more accessible for shape evolution research. `MORPHOBLOCKS` provides a complete workflow for formatting, analysing and visualising the variation between multiple-part objects by integrating functions from a diverse range of other packages. In particular, global components produced by RCPCA provide a consensus space that we present here as a morphospace for multiple-part objects.
3. `MORPHOBLOCKS` is demonstrated with a case study of manually placed landmarks and automatically placed pseudolandmarks from the partial wing skeletons of 15 extant penguin species and five fossil penguin species. Our case study provides quantitative support for a historical hypothesis about the magnitude and mode of morphological change across the evolutionary history of penguins.
4. RCPCA can be used to analyse two- or three-dimensional datasets with 10s of landmarks, or 100s to 1,000s of semilandmarks or pseudolandmarks, from 10s to 100s of specimens comprised of two or more parts. We use `MORPHOBLOCKS` on a small three-bone case study and provide a framework for applying this method to much larger studies investigating the ecological or evolutionary significance of multiple-part objects.

[Corrections added on 29 December 2022, after first online publication: The copyright line has been changed.]

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

## KEYWORDS

bone, generalised Procrustes surface analysis, geometric morphometrics, MORPHOBLOCKS, penguin, regularised consensus principal component analysis, regularised generalised canonical correlation analysis, shape

## 1 | INTRODUCTION

Evolution of shape is a major research theme across biology (Felice & Goswami, 2018; Kysela et al., 2016; Sherratt et al., 2017). Shape is often analysed by recording positions of  $p$  points in  $k$  dimensions across  $n$  biological parts, where points represent corresponding structures across the set of objects (Bookstein, 1978). Points identify homologous anatomical features (i.e. landmarks) or are interpolated between homologous anatomical features (e.g. semi- or pseudolandmarks; Pomidor et al., 2016; Webster & Sheets, 2010). The  $kp$  measures of landmarks or semilandmarks from each object are superimposed, rotated and rescaled to minimise the sum of squared distances between corresponding points in each of the  $n$  objects (Procrustes transformation; see Rohlf & Slice, 1990 and references therein). Shape evolution research produces big  $n \times kp$  datasets:  $k = 2$  or 3 dimensions for  $p = 10$ s of landmarks or 100s to 1,000s of semi- or pseudolandmarks across  $n = 10$ s to 100s of objects (Goswami et al., 2019). Shape differences are therefore commonly studied by reducing  $kp$  to fewer dimensions that explain major sources of variation in the dataset (Kendall, 1984; Mitteroecker & Huttegger, 2009).

Geometric morphometrics (i.e. shape analyses) typically focus on the landmark configuration from one part per individual (e.g. the left femora of 256 people; Cavaignac et al., 2016), where Procrustes transformation is applied to a single  $n \times kp$  block to give all objects the same Cartesian origin (Mitteroecker & Huttegger, 2009). Consider that the shape of femora from  $n$  individuals is easily compared as a single block of data, but the shape of the lower limb skeleton must account for the shape of each leg bone and also the rotations of bones around the knee and other joints. Each individual bone in the lower leg would have a different Cartesian origin and is therefore a separate  $n \times kp$  block. Here we investigate a method for analysing the shapes of objects that contain multiple parts (i.e. skeletons instead of single bones).

Established methods for analysing multiple-part objects either (1) make a priori decisions about the positions and rotations of each part relative to one another and then fix the parts into those positions and rotations (e.g. SHAPEROTATOR in R; Vidal-García et al., 2018), or (2) superimpose the parts into a new amalgamated configuration about the same Cartesian origin and then analyse the amalgam (e.g. Adams, 1999). For (2), the landmark subsets from each fixed part (i.e. each block) are transformed using generalised Procrustes analysis and scaled (Adams, 1999; Collyer et al., 2020; Profico et al., 2019). A scaling method that normalises landmark configurations using centroid size is described by Adams (1999) and Collyer et al. (2020), and has recently been introduced into the GEOMORPH package for the R programming language (see geomorph::combine.subsets; Adams & Otárola-Castillo, 2013). The scaled and transformed landmark configurations from the fixed parts are then combined and analysed as a single set of landmarks per individual (i.e. Adams, 1999; Collyer

et al., 2020). Downstream analyses of landmark configurations presented as a single block often use principal component analysis (PCA), where components of variation describe distances between samples with respect to the greatest sources of variation in the dataset (although see Bookstein, 2017 for discussion). For both (1) and (2), the a priori decisions about the scaling, positions and rotations of each part in the multiple-part object prior to fusion can introduce artificial sources of variation to shape analyses.

We investigated regularised consensus principal component analysis (RCPCA) as a method for comparing the shapes of multiple-part objects (Janné et al., 2001; Tenenhaus et al., 2017; Westerhuis et al., 1998; Wold et al., 1996). Regularised consensus principal component analysis does not require the positions or rotations of parts in a multiple-part object to be arbitrarily fixed relative to one another, but does require consideration about scaling parts relative to one another. RCPCA (Hanafi et al., 2011; Tenenhaus et al., 2017; Westerhuis et al., 1998) is a multiblock method. It is a particular instantiation of regularised generalised canonical correlation analysis (RGCCA; Tenenhaus & Guillemot, 2017; Tenenhaus et al., 2017).

Regularised RGCCA is a very general statistical framework for multiblock component analysis. RGCCA includes consensus principal component analysis as a special case. The goal of RCPCA is to find jointly a set of components per block (block component) and a set of global components. The block components are designed to (a) explain their own block well (in terms of variance criteria) and (b) be correlated with the global components. The global components are obtained as a linear combination of the block components. Block components, global components and the weights of the linear combination that link the block components to the global components are determined by maximising a covariance-based optimisation problem (see Tenenhaus et al., 2017 for a detailed discussion on the RCPCA optimisation problem and the links between block components and the global components).

Regularised consensus principal component analysis does not have any limitations that are not also limitations to PCA, except that RCPCA is only relevant when variables are naturally structured in blocks. For RGCCA and thus RCPCA the number of data points must be the same for each block, but the nature of the variables can differ between blocks. In general, and especially for the covariance-based criterion, the data might be pre-processed to ensure comparability between variables and blocks. To make variables comparable, standardisation is applied (zero mean and unit variance). To make blocks comparable, a strategy is to divide each block by the square root of its number of variables. This two-step procedure leads to  $\text{trace}(\mathbf{X}_i^T \mathbf{X}_i) = n$  for each block (i.e. the sum of the eigenvalues of the covariance matrix of block  $\mathbf{X}_i$  is equal to one for each block). Such a pre-processing step is applied by default when using RGCCA (see Tenenhaus & Tenenhaus, 2014; Garali et al., 2018 for a detailed discussion of these pre-processing steps). The

RCPCA algorithm is a very efficient global convergent algorithm and supports blocks with large numbers of observations. In terms of optimisation, it is possible to include 10s or 100s of blocks, but data interpretation could become more difficult with larger numbers of blocks.

(Standardised) principal component analysis assumes that each variable has the same importance in the analysis. The user is not easily able to weight the contribution of blocks during the analysis (i.e. when performing PCA on a concatenated superblock). Instead, block weighting can only be determined a priori, for example, by using normalised centroid size weighting as a data pre-processing step (discussed in Collyer et al., 2020). In contrast, consensus principal component analysis weights the contribution of each block to global components. Before RCPCA, users might still adjust the weight of each block a priori as a data pre-processing step. Normalised centroid size weighting is applied by default as a pre-processing step in `morphoBlocks::combineBlocks`, meaning that blocks are given similar weighting before RCPCA. One consequence of this a priori weighting is that PCA of the superblock and RCPCA of the individual blocks can give similar results for many datasets. However, there may be analytical reasons why normalised centroid size weighting or other scaling steps are not appropriate for particular analyses (discussed in Collyer et al., 2020). If block weighting is not applied as a pre-processing step (i.e. `morphoBlocks::analyseBlocks`, `option = 'cent.scale = FALSE'`), then RCPCA will still determine the weighting that each block should have to the global components, whereas PCA of the superblock will be applied to the unweighted data. In general, PCA reveals the contribution of variables to the construction of principal components, whereas RCPCA reveals the contribution of blocks to the construction of global components.

In the framework of RCPCA, the  $n \times kp$  block is separated into two or more blocks along  $p$ . Consider for example  $p$  landmarks distributed across multiple bones in a skeleton, and then repeated for skeletons of  $n$  individuals, where data for the same bone in all  $n$  individuals are sectioned off into a separate matrix (i.e. an individual data block). This process is repeated for each other biologically meaningful group of  $p$  (e.g. one data block for each type of bone). As with PCA performed on a single block, RCPCA reduces the  $kp$  dimensions in each individual block to fewer dimensions that explain major sources of variation (i.e. principal components, PCs), and calculates the amount of variance explained along these new dimensions (Tenenhaus et al., 2017; Westerhuis et al., 1998). Moreover, RCPCA produces global components resulting in the so-called consensus that can be viewed as a compromise space that integrates all the individual blocks. This compromise space enables visualising all  $n$  individuals and eases the interpretation of the results. For a full description of RCPCA and RGCCA see Tenenhaus et al. (2017). For biological shape analyses, this consensus space could be used to construct a morphospace for multiple-part objects (i.e. a skeleton morphospace).

Exploring rates of morphological change within a clade is one potential application for a multiple-part morphospace. Projecting a phylogeny into a morphospace is now a well-established method for incorporating evolutionary history into analyses of morphological change (Sidlauskas, 2008), and a morphospace informed by multiple

skeletal elements is a new opportunity to map changes in a limb or whole skeleton through evolutionary time. Hypothetical phylomorphospaces for whole organisms were the basis for Simpson's famous 'grid diagram' imagery that explored the existence of adaptive zones and their influence on lineage evolution (McGhee, 2006; Simpson, 1944, 1953). Here we demonstrate the utility of RCPCA for calculating a multiple-part phylomorphospace and discuss the potential for investigating the evolutionary rate of morphological change in a body region. We present a workflow for constructing a multiple-part morphospace with RCPCA using both traditional landmarks and pseudolandmarks.

## 2 | MATERIALS AND METHODS

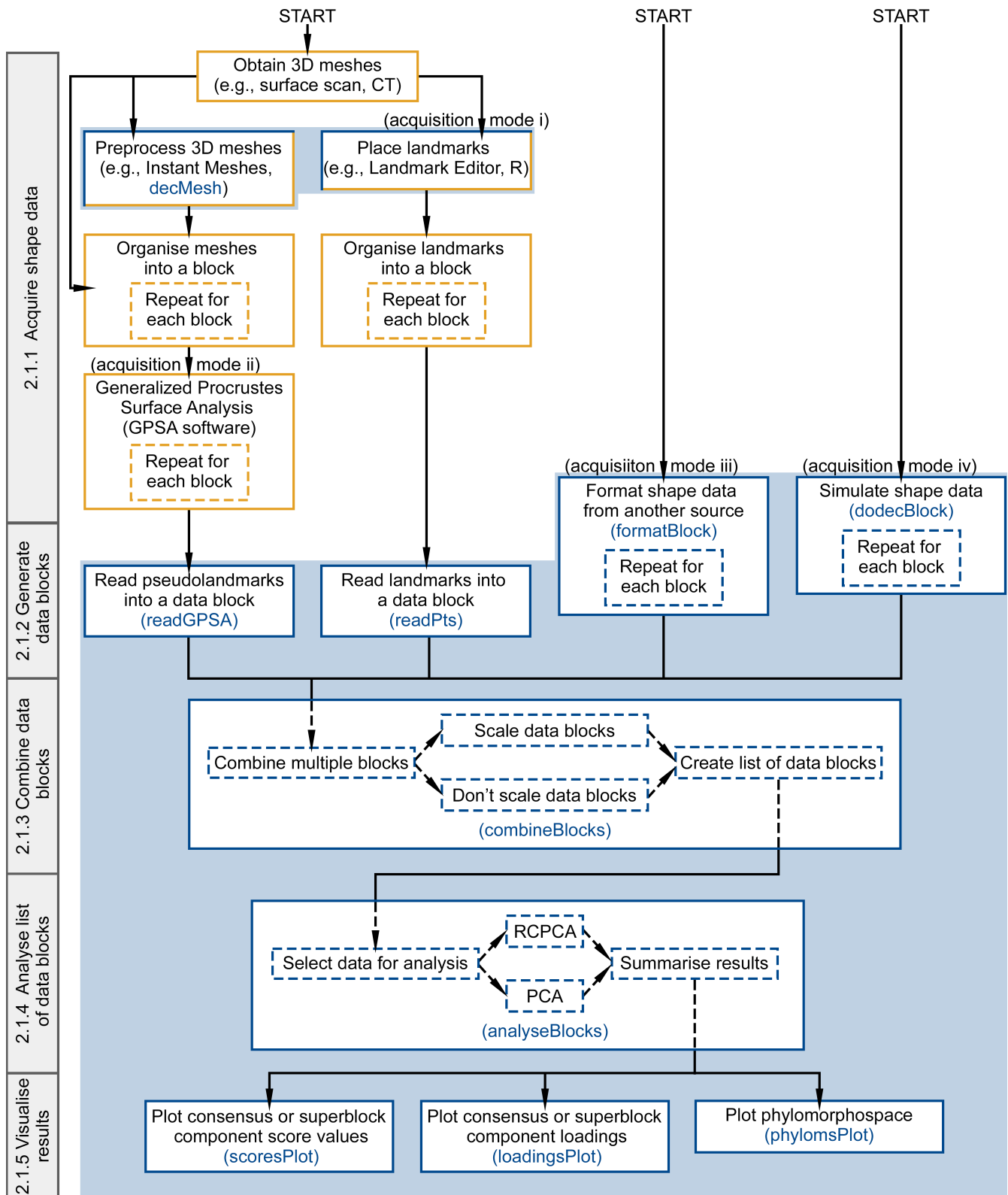
### 2.1 | Multiple element workflow

`MORPHOBLOCKS` (Harmer & Thomas, 2021; <https://aharmer.github.io/morphoBlocks/>) facilitates the processing, analysis and visualisation of variation between shape configurations from objects comprised of multiple parts (see Figure 1 for workflow). Most functions in `MORPHOBLOCKS` are wrappers to conveniently collate functions from a diverse range of other packages (e.g. `APE`, `ADEPHYLO`, `GEOMORPH`, `MORPHO`, `PHYTOOLS` and `RGCCA`; Adams & Otárola-Castillo, 2013; Jombart et al., 2010; Paradis & Schliep, 2019; Revell, 2012; Schlager, 2017; Tenenhaus & Guillemot, 2017). Although `MORPHOBLOCKS` provides a complete workflow within the R environment, some data acquisition steps may require additional software or other resources (Figure 1). The workflow begins with acquisition of shape configurations from a dataset of samples that each contain two or more parts.

#### 2.1.1 | Acquire shape data

Shape data are acquired or generated using methods that are often study specific. Here we describe four data acquisition modes that should represent most evolutionary or ecological studies of biological shape. Acquisition mode i—shape configurations are acquired by placing landmarks or semilandmarks on 3D meshes using, for example, Landmark Editor (IDAV, Wiley et al., 2005), 3D Slicer (Kikinis et al., 2014) or `geomorph::digitsurface` (Adams & Otárola-Castillo, 2013).

Acquisition mode ii—shape configurations are acquired as automatically placed pseudolandmarks from generalised Procrustes surface analysis (GPSA, Pomidor et al., 2016). Three-dimensional meshes for corresponding parts across all samples are included in the same GPSA to produce a single homologised points file for that part. `morphoBlocks::decMesh` decimates and evenly distributes vertices and may be used as a preprocessing step before GPSA for meshes with very large vertex counts, or for datasets with highly disparate vertex counts between meshes. Note that the centroid sizes used for configuration scaling are calculated from the pseudolandmark configurations and not from the original meshes before transformation with GPSA.



**FIGURE 1** Workflow for analysing multiple-part objects using functions in MORPHOBLOCKS. Blue boxes and blue background indicate operations performed in R (R Core Team, 2019)

Acquisition mode iii—shape configurations are generated as landmarks placed on two-dimensional images (e.g. Profico et al., 2019), spatial coordinates acquired using a Microscribe (e.g. Badawi-Fayad

& Cabanis, 2007), from previously prepared matrices of shape data, or are acquired from other sources. These data are available for analysis using `morphoBlocks::formatBlock`. The user reads shape

configurations from the same part in each sample into R and organises the configurations as a matrix with each row representing a different sample. The user also provides additional descriptors about the shape configurations including an optional matrix and an optional vector for describing curves and surfaces (`morphoBlocks::formatBlock` calls `geomorph::gpgagen`; Adams & Otárola-Castillo, 2013). Other supplied information may include a vector of centroid sizes for the configurations (required for configuration scaling, e.g. Collyer et al., 2020), the number of dimensions for each point in the shape configurations and a logical statement to determine if generalised Procrustes analysis should be performed. The shape configurations and ancillary information are used by `morphoBlocks::formatBlock` to generate a *block*-class data block object (see Section 2.1.2).

Acquisition mode iv—shape data are simulated. `morphoBlocks::dodecBlock` generates a data block of two or more dodecahedra where the vertices of each dodecahedron represent a landmark configuration with 20 points ( $p$ ) in three dimensions ( $k$ ) ( $n$  is user specified). `morphoBlocks::dodecBlock` allows the user to change the size, position and rotation of individual configurations within the data block, and to alter the position of one vertex, and therefore introduce a known amount of variation into a dataset. The data block of simulated configurations is transformed using generalised Procrustes analysis and formatted as a *block*-class data block object (see Section 2.1.2).

## 2.1.2 | Generate data blocks

Shape configurations from the corresponding part in each sample are organised into a data block. For acquisition mode i, landmarks in .pts format from the same part in each sample are read into the R environment using `morphoBlocks::readPts` and will represent the data block for that part (e.g. 10 .pts files from the humerus in each of 10 right forelimbs). For acquisition mode ii, the homologised points file from the corresponding parts across all samples is read into R using `morphoBlocks::readGPSA` and will represent the data block for that part. The `morphoBlocks::formatBlock` and `morphoBlocks::dodecBlock` functions of acquisition modes iii and iv likewise produce a data block. The data block from each function is structured as a *block*-class object to facilitate downstream functions. The *block*-class object includes the shape configurations before any downstream transformations (@raw), the configurations after Procrustes transformation organised into 2D and 3D arrays (@gpa.2D and @gpa.3D respectively), the centroid sizes of the raw configurations (@centroid) and the number of points (@p), dimensions (@k) and samples (@n) in the data block.

## 2.1.3 | Combine data blocks

Each data block represents shape configurations for a different part in the multiple-part dataset. Two or more data blocks of shape configurations (i.e. *block*-class objects) are assembled into a list using `morphoBlocks::combineBlocks`. The shape configurations within

each data block are scaled using the normalised centroid size method described by Profico et al. (2019) and Collyer et al. (2020) and described in the supporting information (see also `geomorph::combine.subsets`; Adams & Otárola-Castillo, 2013). The scaling is summarised here in four steps. Step 1—the centroid size of each configuration is divided by the square root of  $p$  for that data block. Step 2—all relativised centroid sizes from step 1 are squared, and then the values from the same sample are summed (i.e. the squared relativised centroid size from each data block for sample one are added together). Step 3—each squared relativised centroid size from each block is divided by the sum of these values for all blocks from the same sample and squared, producing a normalised centroid size. Step 4—each shape configuration is scaled by multiplying by the normalised centroid size calculated in step 3. Normalised centroid size scaling occurs by default in `morphoBlocks::combineBlocks` but can be prevented if the study requires data that are not scaled. Future revision of `morphoBlocks::combineBlocks` might consider introducing additional scaling methods (e.g. unweighted centroid scaling; Collyer et al., 2020). Finally, a column-wise concatenation of each block is added as a final ‘superblock’ to the list produced by `morphoBlocks::combineBlocks`.

## 2.1.4 | Analyse list of data blocks

Variation between the multiple-part samples is calculated using `morphoBlocks::analyseBlocks`. `morphoBlocks::analyseBlocks` is a wrapper for PCA performed with `stats::prcomp`, and for RCPCA with `RGCCA::rgcca` (version 2.1.2; Tenenhaus & Guillemot, 2017; Tenenhaus et al., 2017).

‘pca’ option—a principal component analysis is performed on the superblock (e.g. Adams, 1999). The result from `stats::prcomp` is retained and the component scores and loadings are formatted for downstream analysis (see ‘`morphoBlocksMethods`’ vignette in `MORPHOBLOCKS` for example).

‘rcpca’ option—RCPCA is performed on the blocks and superblock using the default design matrix and  $g(x) = x^2$  (see supporting documentation for the `RGCCA` package; Tenenhaus & Guillemot, 2017). Default standardisation from `RGCCA::rgcca` is applied: each block is standardised and divided by the square root of its number of variables. The complete result from `RGCCA::rgcca` is retained including (1) explained variance for each individual block and for the consensus space (see `RGCCA` package documentation by Tenenhaus & Guillemot, 2017 for definition), and (2) global component scores and loadings associated with the consensus space. (2) Global loadings reveal the contribution of each landmark or pseudolandmark to the variation explained in each principal component.

## 2.1.5 | Visualise the distribution of variation

Component scores from the superblock (`morphoBlocks::analyseBlocks`, option = ‘pca’), or global component scores from the consensus space (`morphoBlocks::analyseBlocks`, option = ‘rcpca’), provide a morphospace

for multiple-part samples (visualised with `morphoBlocks::scoresPlot`). Likewise, component loadings from the superblock or global component loadings from the consensus space reveal the contribution that each region of the shape configurations make to the variation explained by each component (visualised with `morphoBlocks::loadingsPlot`). If the multiple-part samples being analysed are represented in a phylogenetic tree (*phylo-class*) then the component scores may be presented as a phylomorphospace using `morphoBlocks::phylomsPlot`. `morphoBlocks::phylomsPlot` is a wrapper for `phytools::phylomorphospace` (Revell, 2012), which is in turn based on the projection of a phylogenetic tree into a morphospace by Sidlauskas (2008).

## 2.2 | Case study

A humerus, radius and ulna from 15 extant species of penguin and also from five fossil species of penguin contributed to a dataset of 60 wing bones (Table 1). A 3D digital replica of each bone was produced using structured light 3D surface scanning (HP SLS-3 scanner; HP Inc.). Data collection and 3D model production used HP 3D Scan 5 software (version 5.6.0.2037; HP Inc.). The minimum resolvable surface feature of the 3D digital replicas was approximately 0.2 mm. Right wing bones were selected when available. Left wing elements were 3D scanned and digitally mirrored in Blender 2.8.1 when right elements were not available (Blender Online Community, 2019). Two analyses are described to show how `MORPHOBLOCKS` can be used for shape configurations acquired by different methods. The case study described in Section 2.2.1 is available as the 'penguinWings' vignette in `MORPHOBLOCKS`.

### 2.2.1 | Manually placed landmarks

The dataset of 60 digital bone replicas was replicated to produce three identical datasets. A copy of the digital bone dataset was provided to three authors for landmark placement. Landmark schemes were chosen to emphasise muscle attachment sites, articulation and general proportions (see Supporting Information including Figure S1). The humerus scheme featured three single point landmarks, four curves (each divided into 10 semilandmarks) and three patches (each divided into 25 semilandmarks). The radius scheme featured one single point landmark, four curves and one patch, and the ulna scheme featured two single point landmarks, four curves and one patch. Landmarks were placed using Landmark Editor 3.0 (Institute of Data Analysis and Visualization IDAV, University California Davis, USA; Wiley et al., 2005).

Landmark configurations were separated by author and block into nine directories (i.e. 20 humerus .pts files from author DBT in one directory, 20 humerus configurations from author EJM in a separate directory, etc.) before processing with functions from `MORPHOBLOCKS`. `morphoBlocks::readPts`—.pts files in each directory were separately read into R version 4.0.2 (R Core Team, 2019) with `gpa = FALSE` (i.e. generalised Procrustes transformation not performed). Landmark configurations from each block replicate were extracted from the `@raw` term of their respective *block-class* objects and averaged to control for

variation in landmark placement. `morphoBlocks::formatBlock`—the averaged landmarks were formatted into a data block using the `@curves` term from one of the replicates and with `gpa = TRUE`. Generalised Procrustes transformation was performed using `geomorph::gpagen` (version 3.0.6; Adams & Otárola-Castillo, 2013), which is called by `morphoBlocks::formatBlock`. `morphoBlocks::combineBlocks`—the three data blocks of Procrustes-transformed configurations (humerus, radius and ulna) were organised into a list of data blocks and scaled using the normalised weighted centroid size method from Collyer et al. (2020). `morphoBlocks::analyseBlocks`—scaled data blocks were analysed with RCPA using `RGCCA::rgcca` (Tenenhaus & Guillemot, 2017), which is called by `morphoBlocks::analyseBlocks` when option = 'rcpca'.

### 2.2.2 | Automatically placed pseudolandmarks

The dataset of 60 digital bone replicas was separated into three blocks of 20 bones according to bone type (i.e. humerus, radius and ulna) and the following process was applied to each block independently. Pseudolandmark positions and correspondences were determined for bones in each block using GPSA version 20160308a (Pomidor et al., 2016). Here each humerus, radius and ulna was represented by 404,513, 198,947 and 306,841 pseudolandmarks respectively. For full GPSA method details see Pomidor et al. (2016). Briefly, one 3D digital bone replica is chosen as a reference target for setting the Cartesian origin of all other digitised bones in the same block. Each digital replica is scaled and superimposed on the reference model using an iterative closest point process. Vertices in each aligned digital replica are paired with the nearest vertex in the reference bone replica. Vertices in the aligned digital replicas and the reference replica are then remapped using the averaged position of the vertex pairings. The remapped vertices are used to determine correspondences between vertices in each aligned replica and the reference model. Remapped vertices are stored as a homologised points file (.dat) and are here treated as pseudolandmarks that correspond between digital bone replicas. `morphoBlocks::readGPSA`—the homologised points files from the humeri, radii and ulnae were separately read into R. `morphoBlocks::combineBlocks`—blocks of pseudolandmarks (humeri, radii and ulnae) were organised into a list of data blocks and scaled using the normalised weighted centroid size method from Collyer et al. (2020). `morphoBlocks::analyseBlocks`—scaled data blocks were analysed with RCPA using `RGCCA::rgcca` (Tenenhaus & Guillemot, 2017) called by `morphoBlocks::analyseBlocks` option = 'rcpca'.

## 3 | CASE STUDY RESULTS

### 3.1 | Consensus space

Regularised consensus principal component analysis applied to datasets of manually placed landmarks and automatically placed pseudolandmarks produced broadly similar results (Figures 2 and 3). Approximately a quarter or more of the variation in the dataset was explained along component one (C1) of the individual blocks, and

**TABLE 1** 3D digital replicas were produced from bones of modern and fossil penguins. Institution abbreviations: CM, Canterbury Museum, Christchurch, New Zealand; NMNZ, Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand; OM, Otago Museum, Dunedin, New Zealand; UC, University of Canterbury, Christchurch, New Zealand (specimen held at OU, Geology Museum, University of Otago, Dunedin, New Zealand). Specimen ages from or compiled by Slack et al. (2006), Ksepka and Ando (2011) and Ksepka et al. (2012) and literature reviewed therein

Specimen				Element			
Accession	Genus	Species	Age	Humerus	Radius	Ulna	Institution
AV19569	<i>Aptenodytes</i>	<i>forsteri</i>	Extant	Right	Left	Right	CM
2013-1-257	<i>Aptenodytes</i>	<i>patagonicus</i>	Extant	Left	Right	Right	CM
OR.019303	<i>Eudyptes</i>	<i>filholi</i>	Extant	Right	Right	Right	NMNZ
OR.017176	<i>Eudyptes</i>	<i>pachyrhynchus</i>	Extant	Right	Left	Left	NMNZ
OR.030196	<i>Eudyptes</i>	<i>robustus</i>	Extant	Right	Right	Right	NMNZ
OR.027787	<i>Eudyptes</i>	<i>schlegeli</i>	Extant	Right	Right	Right	NMNZ
OR.025560	<i>Eudyptes</i>	<i>sclateri</i>	Extant	Right	Right	Right	NMNZ
OR.018965	<i>Eudyptula</i>	<i>minor</i>	Extant	Left	Left	Left	NMNZ
OR.026794	<i>Megadyptes</i>	<i>antipodes</i>	Extant	Right	Right	Right	NMNZ
AV38751	<i>Pygoscelis</i>	<i>adeliae</i>	Extant	Left	Left	Left	CM
AV32995	<i>Pygoscelis</i>	<i>antarctica</i>	Extant	Right	Right	Right	CM
AV38694	<i>Pygoscelis</i>	<i>papua</i>	Extant	Right	Right	Right	CM
OR.022800	<i>Spheniscus</i>	<i>demersus</i>	Extant	Right	Right	Right	NMNZ
CU2138	<i>Spheniscus</i>	<i>humboldti</i>	Extant	Right	Right	Right	UC
OR.027329	<i>Spheniscus</i>	<i>magellanicus</i>	Extant	Right	Right	Right	NMNZ
OR.01451	<i>Platydyptes</i>	<i>novaezeelandiae</i>	27.3–21.7 Ma	Left	Right	Left	NMNZ
GL407	<i>Archaeospheniscus</i>	<i>lowei</i>	29.8–25.2 Ma	Right	Right	Right	OM
OU22065	<i>Kairuku</i>	<i>grebneffi</i>	29.8–25.20 Ma	Right	Right	Right	OU
CM 2016-6-1	<i>Sequiwaimanu</i>	<i>rosieae</i>	61 Ma	Left	Left	Right	CM
OU12651	<i>Muriwaimanu</i>	<i>tuatahi</i>	61.3–56.0 Ma	Right	Left	Left	OU

along global component one (GC1) of the consensus (Table 2). The radius accounted for the most variation of the three individual blocks (0.37 and 0.56 average variance explained along C1 for landmarks and pseudolandmarks respectively). The relative positions of scores along C1 and C2 in each individual block and GC1 and GC2 in the consensus were similar where specimens tended to cluster by age (Figures 2 and 3). The oldest specimens in the analyses (Palaeocene) were separated from all other specimens along C1 for all three individual blocks and GC1 of the consensus space, and this separation was greatest for the radius block and for the consensus space (Figures 2 and 3).

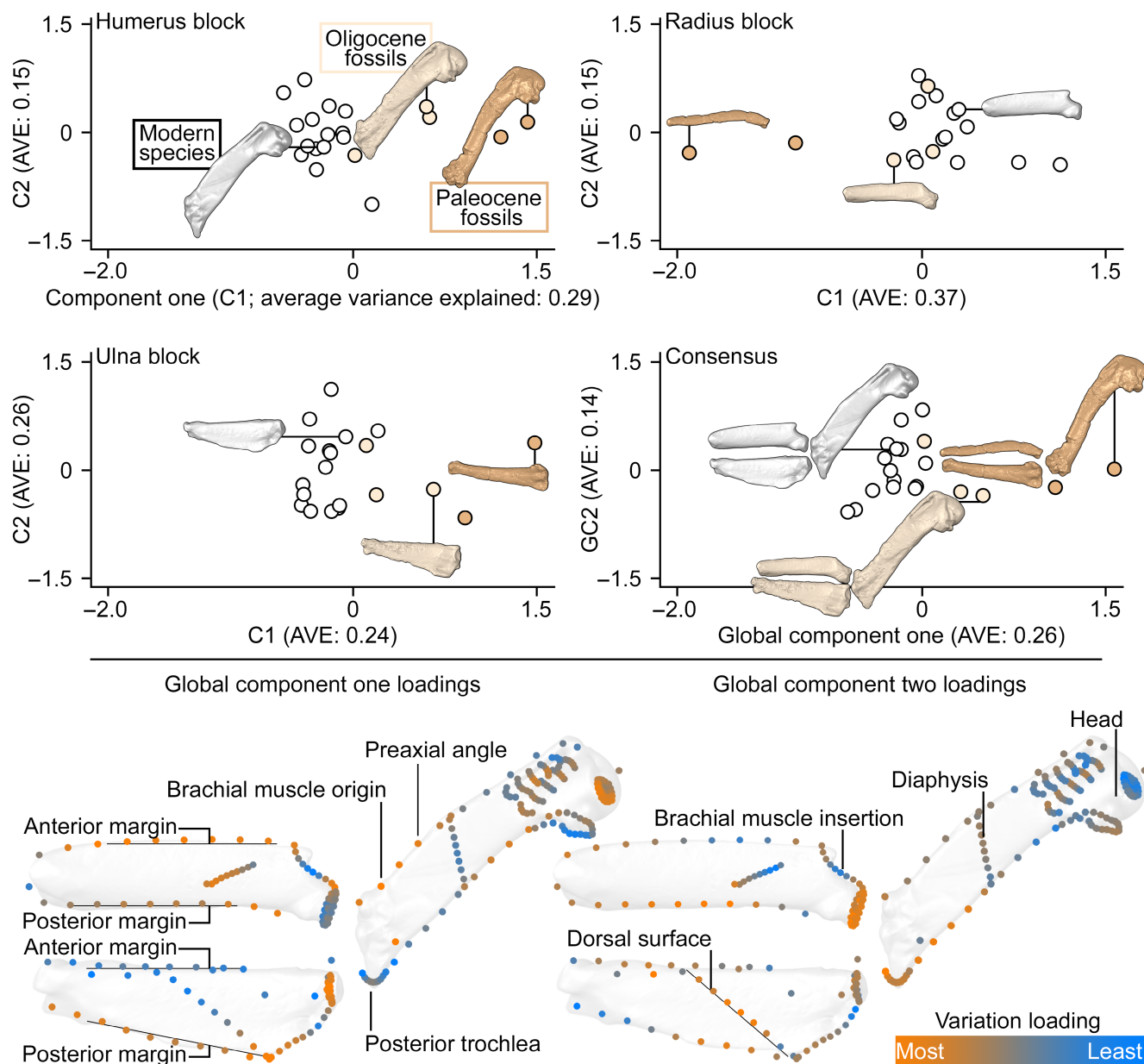
### 3.2 | Distribution of shape variation across blocks

Shape variation was unevenly distributed across the data blocks as revealed by the global component loadings in both the RCPA of the manually placed landmark data and the RCPA of the automatically placed pseudolandmark data (Figures 2 and 3). For the manually placed landmark data, the highest variation for GC1 was along the anterior and posterior margins of the radius (Figure 2). The posterior margin of the ulna and the more distal region of the humerus (near the preaxial angle) were highly variable along GC1. Variation along GC2 was loaded on the scar for the insertion of

the brachial muscle on the radius, the dorsal surface of the ulna and the head and other proximal regions of the humerus. For the automatically placed pseudolandmark data, the greatest variation along GC1 was again observed for the anterior and posterior margins of the radius, the anterior and posterior margins of the ulna, as well as several regions on the humerus: preaxial angle, scar for the origin of the brachial muscle, and the posterior trochlea (Figure 3). The ventral surface of the humeral diaphysis and the proximal surface of the humeral head were major sources of variation for GC2, as were the proximal articular surface of the radius, the distal articular surface of the ulna and the dorsal and ventral surfaces of the ulna.

### 3.3 | Phylomorphospace

The first global component (GC1) from the pseudolandmark RCPA was plotted against time to visualise shape evolution as a phylogenetic morphospace (Revell, 2012; Sidlauskas, 2008) using a phylogeny adapted from Thomas et al. (2020) (see Figure 4 and Table 1). The phylogeny was pruned to only include taxa from Table 1. *Archaeospheniscus lopdelli* in the original phylogeny was substituted with *Archaeospheniscus lowei*, and *Platydyptes novaeseelandiae* was

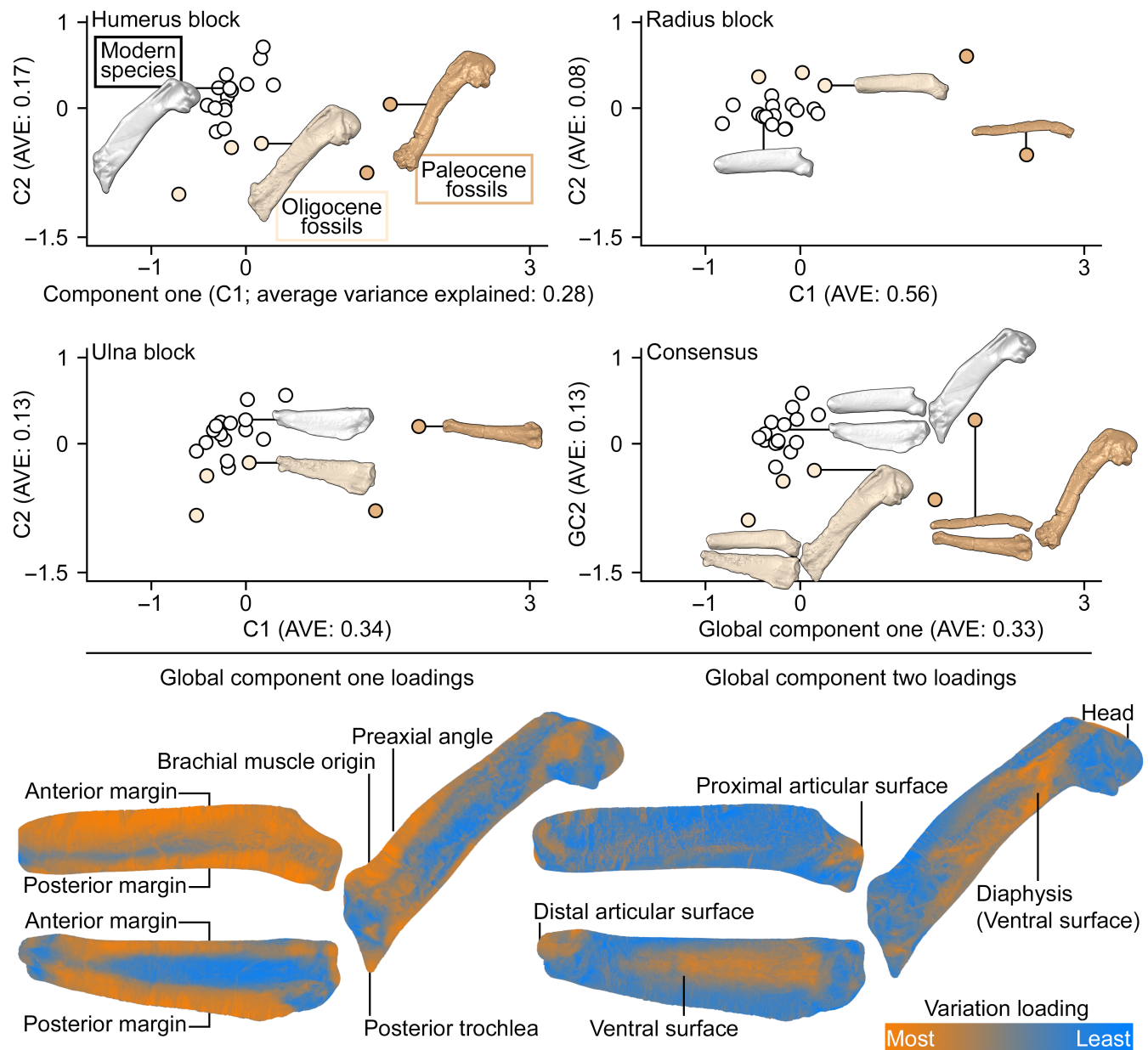


**FIGURE 2** Shape variation in the upper wing of fossil and modern penguins as an example of a multiple-part object. Landmarks manually placed on three wing bones (humerus, radius and ulna) constituted three data blocks analysed with regularised consensus principal component analysis. Shown here are component one (C1) and C2 scores from the individual blocks, and global component one (GC1) and GC2 scores from the consensus space (Tenenhaus & Guillemot, 2017; Tenenhaus et al., 2017). Scores are coloured according to specimen age: early stem-lineage penguins from the Palaeocene (brown), stem-lineage penguins from the Oligocene (light brown), and extant penguins (white). Global component loadings are visualised by colouring the mean position of each landmark in the superbloc. Stronger orange colours represent landmarks with larger loadings (i.e. greater amounts of variation), and stronger blue colours represent landmarks with weaker loadings. Low-opacity images of penguin bones have been placed behind the loadings for spatial context only

grafted onto the tree one node down from the most recent common ancestor to crown penguins. We acknowledge that this case study is only demonstrative and that a more robust analysis would include more specimens and additional bones.

The proportion of variation explained by GC1 of the automatically placed pseudolandmarks was 0.33 for the 20 specimens of upper wing skeletons we analysed (Table 2). Here we demonstrate a

substantial difference in shape for the oldest stem-lineage penguins and comparatively less shape variation between the younger fossil penguins and extant penguins (Figure 4a). Our small dataset echoes an hypothesis by Simpson (1946) in which a theoretical phylomorphospace for penguins shows more rapid morphological change early in the history of Sphenisciformes, and then a reduced rate of morphological change in more recent penguins (Figure 4b).



**FIGURE 3** Shape variation in the upper wing of fossil and modern penguins. Pseudolandmarks automatically placed on three wing bones (humerus, radius and ulna) using generalised Procrustes surface analysis software from Pomidor et al. (2016) constituted three data blocks analysed with regularised consensus principal component analysis. Shown here are component one (C1) and C2 scores from the individual blocks, and global component one (GC1) and GC2 scores from the consensus space (Tenenhaus et al., 2017). See Figure 2 caption for colour key and loadings description

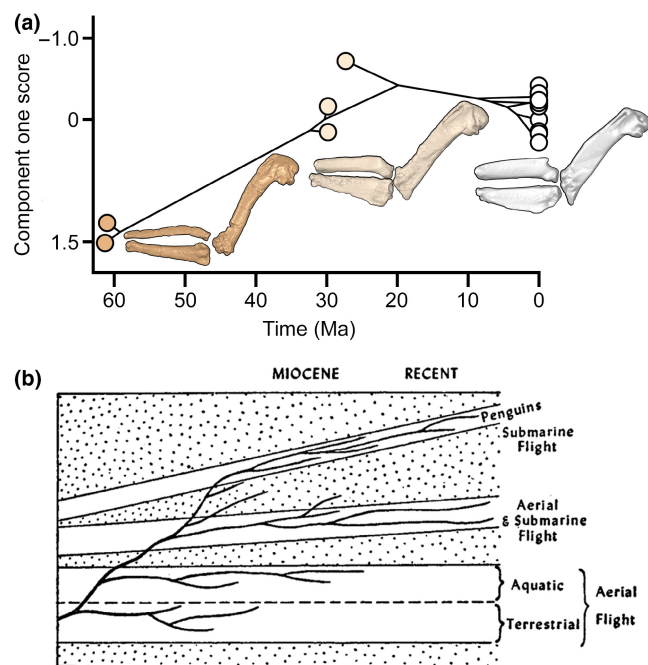
## 4 | DISCUSSION

Regularised consensus principal component analysis is useful for constructing a morphospace for objects comprised of multiple parts. In our three-bone case study we calculated a consensus space (i.e. skeleton morphospace) with variation distributed in a similar pattern to the ordination spaces of each of the individual bones in those skeletons. This was because all three bones showed similar morphological trends through evolutionary time (see Section 4.1 below for

further discussion). However, one could consider a different scenario in which bones in a set of skeletons show different levels of morphological conservation, and thus produce a consensus space for a full skeleton that differs substantially from the morphospace of the most morphologically conserved bone in that skeleton. Given the integration of variation across the skeleton, a consensus space has the potential to allow for emergent patterns of inter-specimen variation that might not be predicted from the morphospace of any one bone.

**TABLE 2** Average explained variance expressed as a proportion of variation in the dataset explained by components from regularised consensus principal component analyses of manually placed landmarks and automatically placed pseudolandmarks. Values for the first 10 components (C1–C10) from all three blocks (humerus, radius and ulna), and values for the first 10 global components from the consensus

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Manually placed landmarks										
Humerus	0.29	0.15	0.17	0.1	0.05	0.06	0.06	0.02	0.03	0.02
Radius	0.37	0.15	0.12	0.04	0.09	0.06	0.05	0.03	0.02	0.02
Ulna	0.24	0.26	0.14	0.05	0.08	0.06	0.04	0.02	0.03	0.03
Consensus	0.26	0.14	0.13	0.08	0.07	0.06	0.04	0.04	0.03	0.03
Automatically placed pseudolandmarks										
Humerus	0.28	0.17	0.08	0.08	0.06	0.05	0.04	0.04	0.03	0.03
Radius	0.56	0.08	0.07	0.06	0.02	0.02	0.03	0.02	0.02	0.02
Ulna	0.34	0.13	0.08	0.07	0.05	0.05	0.04	0.04	0.02	0.02
Consensus	0.33	0.13	0.08	0.07	0.05	0.05	0.05	0.04	0.03	0.03



**FIGURE 4** Shape variation in the upper wing of penguins through evolutionary time. (a) Global component one scores from consensus principal component analysis (Tenenhaus & Guillemot, 2017; Tenenhaus et al., 2017) of automatically placed pseudolandmarks presented as a phylomorphospace (Revell, 2012; Sidlauskas, 2008) against specimen age (Table 1). Scores are coloured according to specimen age: early stem-lineage penguins from the Palaeocene (brown), stem-lineage penguins from the Oligocene (light brown), and extant penguins (white). (b) Grid diagram of the theory of penguin evolution from Simpson (1946). Figure 31 from Simpson (1946) used with permission from American Museum of Natural History

#### 4.1 | Case study

Much of the shape variation in the penguin upper wing was explained by coeval dorsoventral flattening of all three limb bones in the early history

of penguins, resulting in an increased distance between the leading and trailing edges of the wing. Simultaneous broadening of the humerus, radius and ulna over evolutionary time implies a common evolutionary driver being applied to the shape of all three bones (i.e. viscosity difference of water compared with air). Our skeleton morphospace provides quantitative support for the proposed mode of morphological evolution in penguins hypothesised by Simpson (1946): ‘... there would be very strong selection pressure all in one direction, the direction of greatest aquatic adaptation, which would be toward the structure actually shown by the penguins. The tendency would be for a rapid shift in adaptive relationships and in morphology ...’ (Simpson, 1946, pp. 86–87).

#### 4.2 | Data source

The dataset of manually placed landmarks generated in this study emphasised osteological correlates to functional regions of the bones (e.g. muscle origins), and the dataset of automatically placed pseudolandmarks more-broadly described the shape of the bones. The distribution of scores in the consensus spaces from both datasets was generally informed by similar loadings and thus had a similar structure. The choice of using manually placed landmarks or automatically placed pseudolandmarks is study dependent (e.g. Pomidor et al., 2016; Watanabe, 2018) and here we show the utility of RCPCA for both options. Studies seeking to include more taxa or bones to expand the case study we present here are recommended to use automatically placed pseudolandmarks, at least initially, because of the dramatically reduced processing time compared with manually placing landmarks. However, care must be taken when interpreting RCPCA results from a dataset of automatically placed pseudolandmarks that includes damaged or otherwise incomplete bones, as non-homologous shape artefacts will be introduced as a source of variation into the dataset. Our case study includes some damaged fossils (e.g. CM 2016-6-1 *Sequiwaimanu rosieae*; see Mayr et al., 2017) but we note that the damaged regions of these fossils are not highly loaded along C1 in the individual blocks or GC1 in the consensus.

## 5 | CONCLUSIONS

Analyses of the biological shape from multiple-part objects can be achieved using RCPA using RGCCA:rgcca (Tenenhaus & Guillemot, 2017), which is here accessed with the new MORPHOBLOCKS package. Regularised consensus principal component analysis can be used to analyse two- or three-dimensional ( $k$ ) datasets with 10s of landmarks or 100s to 1,000s of semilandmarks or pseudolandmarks ( $p$ ), from 10s to 100s of specimens ( $n$ ) comprised of two or more parts along  $p$ . The resulting analysis is a summary of shape variation within each  $n \times kp$  part (e.g. bone) and across the full set of data blocks (e.g. skeleton). The major advantage of RCPA for analysing skeleton shape is that positions and orientations of each block are independent of one another. Regularised consensus principal component analysis can be applied to manually placed landmarks and automatically placed pseudolandmarks, and can be used to track shape variation through evolutionary time for a small three-element case study, and potentially scale to much larger studies involving more complex datasets like complete skeletons.

### ACKNOWLEDGEMENTS

We thank the following for access to collections: RE Fordyce & M Richards at Geology Museum, University of Otago, Dunedin, New Zealand; A Tennyson & T Schultz at Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand; RP Scofield at Canterbury Museum, Christchurch, New Zealand; E Burns and K Fleury at Otago Museum. We thank A Ellison, T Guillerme and an anonymous reviewer for their time and for their helpful comments. Massey University Research Funding for supporting the project. SG was supported by a Massey University Doctoral Scholarship.

### CONFLICT OF INTEREST

None to declare.

### AUTHORS' CONTRIBUTIONS

D.B.T. and C.M.M. conceived the study; D.B.T., E.J.H. and S.G. collected the data; D.B.T. analysed the data; A.T., D.B.T. and A.M.T.H. prepared programming code; D.B.T. and A.T. wrote the initial draft. All authors read, edited and approved the final draft. All authors contributed critically to the drafts and gave final approval for publication.

### PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13781>.

### DATA AVAILABILITY STATEMENT

MORPHOBLOCKS package available from Github (<https://aharm.github.io/morphoBlocks/>) and archived with Zenodo <https://doi.org/10.5281/zenodo.5717384> (Harmer & Thomas, 2021). Dataset of manually placed landmarks included as a dataset in MORPHOBLOCKS.

### ORCID

Daniel B. Thomas  <https://orcid.org/0000-0001-7679-0991>  
 Aaron M. T. Harmer  <https://orcid.org/0000-0002-2244-5129>  
 Simone Giovanardi  <https://orcid.org/0000-0002-0949-1585>  
 Emma J. Holvast  <https://orcid.org/0000-0002-7865-8115>  
 Cushla M. McGovern  <https://orcid.org/0000-0003-3187-9743>  
 Arthur Tenenhaus  <https://orcid.org/0000-0003-3459-1518>

### REFERENCES

- Adams, D. C. (1999). Methods for shape analysis of landmark data from articulated structures. *Evolutionary Ecology Research*, 1, 959–970.
- Adams, D. C., & Otárola-Castillo, E. (2013). Geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4, 393–399. <https://doi.org/10.1111/2041-210X.12035>
- Badawi-Fayad, J., & Cabanis, E.-A. (2007). Three-dimensional procrustes analysis of modern human craniofacial form. *The Anatomical Record*, 290, 268–276. <https://doi.org/10.1002/ar.20442>
- Blender Online Community. (2019). *Blender—A 3D modelling and rendering package*. Retrieved from <https://www.blender.org/>
- Bookstein, F. L. (1978). *The measurement of biological shape and shape change*. Springer.
- Bookstein, F. L. (2017). A method of factor analysis for shape coordinates. *American Journal of Physical Anthropology*, 164, 221–245. <https://doi.org/10.1002/ajpa.23277>
- Cavaignac, E., Savall, F., Faruch, M., Reina, N., Chiron, P., & Telmon, N. (2016). Geometric morphometric analysis reveals sexual dimorphism in the distal femur. *Forensic Science International*, 259, 246.e241–246.e245. <https://doi.org/10.1016/j.forsciint.2015.12.010>
- Collyer, M. L., Davis, M. A., & Adams, D. C. (2020). Making heads or tails of combined landmark configurations in geometric morphometric data. *Evolutionary Biology*, 47, 193–205. <https://doi.org/10.1007/s11692-020-09503-z>
- Felice, R. N., & Goswami, A. (2018). Developmental origins of mosaic evolution in the avian cranium. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 555. <https://doi.org/10.1073/pnas.1716437115>
- Garali, I., Adanyeguh, I. M., Ichou, F., Perlberg, V., Seyer, A., Colsch, B., Moszer, I., Guillemot, V., Durr, A., Mochel, F., & Tenenhaus, A. (2018). A strategy for multimodal data integration: Application to biomarkers identification in spinocerebellar ataxia. *Briefings in Bioinformatics*, 19, 1356–1369. <https://doi.org/10.1093/bib/bbx060>
- Goswami, A., Watanabe, A., Felice, R. N., Bardua, C., Fabre, A.-C., & Polly, P. D. (2019). High-density morphometric analysis of shape and integration: The good, the bad, and the not-really-a-problem. *Integrative and Comparative Biology*, 59, 669–683. <https://doi.org/10.1093/icb/icz120>
- Hanafi, M., Kohler, A., & Qannari, E.-M. (2011). Connections between multiple co-inertia analysis and consensus principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 106, 37–40. <https://doi.org/10.1016/j.chemolab.2010.05.010>
- Harmer, A. M. T., & Thomas, D. B. (2021). aharmer/morphoBlocks: v0.1.1 (v0.1.1). *Zenodo*. <https://doi.org/10.5281/zenodo.5717384>
- Janné, K., Pettersen, J., Lindberg, N. O., & Lundstedt, T. (2001). Hierarchical principal component analysis (PCA) and projection to latent structure (PLS) technique on spectroscopic data as a data pretreatment for calibration. *Journal of Chemometrics*, 15, 203–213. <https://doi.org/10.1002/cem.677>
- Jombart, T., Balloux, F., & Dray, S. (2010). adephylo: New tools for investigating the phylogenetic signal in biological traits. *Bioinformatics*, 26, 1907–1909. <https://doi.org/10.1093/bioinformatics/btq292>
- Kendall, D. (1984). Shape manifolds: Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16, 81–121. <https://doi.org/10.1112/blms/16.2.81>

- Kikinis, R., Pieper, S. D., & Vosburgh, K. G. (2014). 3D Slicer: A platform for subject-specific image analysis, visualization, and clinical support. In F. A. Jolesz (Ed.), *Intraoperative imaging and image-guided therapy* (pp. 277–289). Springer.
- Ksepka, D. T., & Ando, T. (2011). Penguins past, present, and future: Trends in the evolution of the Sphenisciformes. In G. Dyke & G. Kaiser (Eds.), *Living dinosaurs: The evolutionary history of modern birds* (1st ed., pp. 155–186). John Wiley & Sons.
- Ksepka, D. T., Fordyce, R. E., Ando, T., & Jones, C. M. (2012). New fossil penguins (Aves, Sphenisciformes) from the Oligocene of New Zealand reveal the skeletal plan of stem penguins. *Journal of Vertebrate Paleontology*, 32, 235–254. <https://doi.org/10.1080/02724634.2012.652051>
- Kysela, D. T., Randich, A. M., Caccamo, P. D., & Brun, Y. V. (2016). Diversity takes shape: Understanding the mechanistic and adaptive basis of bacterial morphology. *PLOS Biology*, 14, e1002565. <https://doi.org/10.1371/journal.pbio.1002565>
- Mayr, G., De Pietri, V. L., Love, L., Mannering, A. A., & Scofield, R. P. (2017). A well-preserved new mid-paleocene penguin (Aves, Sphenisciformes) from the Waipara Greensand in New Zealand. *Journal of Vertebrate Paleontology*, 37, e1398169. <https://doi.org/10.1080/02724634.2017.1398169>
- McGhee, G. R. (2006). *The geometry of evolution*. Cambridge University Press.
- Mitteroecker, P., & Huttegger, S. M. (2009). The concept of morphospaces in evolutionary and developmental biology: Mathematics and metaphors. *Biological Theory*, 4, 54–67. <https://doi.org/10.1162/biot.2009.4.1.54>
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Pomidor, B. J., Makedonska, J., & Slice, D. E. (2016). A landmark-free method for three-dimensional shape analysis. *PLoS ONE*, 11, e0150368. <https://doi.org/10.1371/journal.pone.0150368>
- Profico, A., Piras, P., Buzi, C., Del Bove, A., Melchionna, M., Senczuk, G., Varano, V., Veneziano, A., Raia, P., & Manzi, G. (2019). Seeing the wood through the trees. Combining shape information from different landmark configurations. *Hystrix, the Italian Journal of Mammalogy*, 30, 157–165. <https://doi.org/10.4404/hystrix-00206-2019>
- R Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved from <https://www.r-project.org/>
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223. <https://doi.org/10.1111/j.2041-210x.2011.00169.x>
- Rohlf, F. J., & Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39, 40–59. <https://doi.org/10.2307/2992207>
- Schlager, S. (2017). Morpho and Rvcg—Shape analysis in R. In G. Zheng, S. Li, & G. Székely (Eds.), *Statistical shape and deformation analysis* (pp. 217–256). Academic Press.
- Sherratt, E., Serb, J. M., & Adams, D. C. (2017). Rates of morphological evolution, asymmetry and morphological integration of shell shape in scallops. *BMC Evolutionary Biology*, 17, 248. <https://doi.org/10.1186/s12862-017-1098-5>
- Sidlauskas, B. (2008). Continuous and arrested morphological diversification in sister clades of characiform fishes: A phylo-morphospace approach. *Evolution*, 62, 3135–3156. <https://doi.org/10.1111/j.1558-5646.2008.00519.x>
- Simpson, G. G. (1944). *Tempo and mode in evolution*. Columbia University Press.
- Simpson, G. G. (1946). Fossil penguins. *Bulletin of the American Museum of Natural History*, 87, 1–100.
- Simpson, G. G. (1953). *The major features of evolution*. Columbia University Press.
- Slack, K. E., Jones, C. M., Ando, T., Harrison, G. L., Fordyce, R. E., Arnason, U., & Penny, D. (2006). Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution*, 23, 1144–1155. <https://doi.org/10.1093/molbev/msj124>
- Tenenhaus, A., & Guillemot, V. (2017). RGCCA: Regularized and sparse generalized canonical correlation analysis for multiblock data 2.1.2. Retrieved from <https://CRAN.R-project.org/package=RGCCA>
- Tenenhaus, A., & Tenenhaus, M. (2014). Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of Operational Research*, 238(2), 391–403. <https://doi.org/10.1016/j.ejor.2014.01.008>
- Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. F. (2017). Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods. *Psychometrika*, 82, 737–777. <https://doi.org/10.1007/s11336-017-9573-x>
- Thomas, D. B., Tennyson, A. J. D., Scofield, R. P., Heath, T. A., Pett, W., & Ksepka, D. T. (2020). Ancient crested penguin constrains timing of recruitment into seabird hotspot. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20201497. <https://doi.org/10.1098/rspb.2020.1497>
- Vidal-García, M., Bandara, L., & Keogh, J. S. (2018). ShapeRotator: An R tool for standardized rigid rotations of articulated three-dimensional structures with application for geometric morphometrics. *Ecology and Evolution*, 8, 4669–4675. <https://doi.org/10.1002/ece3.4018>
- Watanabe, A. (2018). How many landmarks are enough to characterize shape and size variation? *PLoS ONE*, 13, e0198341. <https://doi.org/10.1371/journal.pone.0198341>
- Webster, M., & Sheets, H. D. (2010). A practical introduction to landmark-based geometric morphometrics. In J. Alroy & G. Hunt (Eds.), *Quantitative methods in paleobiology. The Paleontological Society Papers* (pp. 163–188). Cambridge University Press.
- Westerhuis, J. A., Kourti, T., & MacGregor, J. F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, 12, 301–321. [https://doi.org/10.1002/\(sici\)1099-128x\(199809/10\)12:5<301:Aid-cem515>3.0.Co;2-s](https://doi.org/10.1002/(sici)1099-128x(199809/10)12:5<301:Aid-cem515>3.0.Co;2-s)
- Wiley, D. F., Amenta, N., Alcantara, D. A., Ghosh, D., Kil, Y. J., Delson, E., Harcourt-Smith, W., St. John, K., Rohlf, F. J., & Hamann, B. (2005). Evolutionary morphing. In *IEEE visualization 2005—(VIS'05)* (pp. 431–438). IEEE.
- Wold, S., Kettaneh, N., & Tjessem, K. (1996). Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *Journal of Chemometrics*, 10, 463–482. [https://doi.org/10.1002/\(sici\)1099-128x\(199609\)10:5:6<463:Aid-cem445>3.0.Co;2-l](https://doi.org/10.1002/(sici)1099-128x(199609)10:5:6<463:Aid-cem445>3.0.Co;2-l)

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Thomas, D. B., Harmer, A. M. T., Giovanardi, S., Holvast, E. J., McGoverin, C. M., & Tenenhaus, A. (2023). Constructing a multiple-part morphospace using a multiblock method. *Methods in Ecology and Evolution*, 14, 65–76. <https://doi.org/10.1111/2041-210X.13781>