An Exploration of Adolescent Facial Shape Changes with Age via Multilevel Partial Least Squares Regression

DJJ Farnell[1*], S Richmond[1], J Galloway[1], AI Zhurov[1], P Pirttiniemi[2,3], T Heikkinen[2,3], V Harila[2,3], H Matthews[4,5,6,7], and P Claes[4,5,8]

[1] *School of Dentistry, Cardiff University, Heath Park, Cardiff CF14 4XY, United Kingdom*

[2] *Research Unit of Oral Health Sciences, Faculty of Medicine, University of Oulu, Oulu, Finland*

[3] *Medical Research Center Oulu (MRC Oulu), Oulu University Hospital, Oulu, Finland*

[4] *Medical Imaging Research Center, UZ Leuven, 3000 Leuven, Belgium*

[5] *Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium*

[6] *Facial Sciences Research Group, Murdoch Children's Research Institute, Melbourne*

[7] *Department of Paediatrics, University of Melbourne, Melbourne, Australia*

[8] *Department of Electrical Engineering, ESAT/PSI, KU Leuven, 3000 Leuven, Belgium*

[*] Corresponding author: Dr Damian JJ Farnell, farnelld@cardiff.ac.uk

Abstract

Background and Objectives: Multilevel statistical models represent the existence of hierarchies or clustering within populations of subjects (or shapes in this work). This is a distinct advantage over single-level methods that do not. Multilevel partial-least squares regression (mPLSR) is used here to study facial shape changes with age during adolescence in Welsh and Finnish samples comprising males and females.

Methods: 3D facial images were obtained for Welsh and Finnish male and female subjects at multiple ages from 12 to 17 years old. 1000 3D points were defined regularly for each shape by using "meshmonk" software. A three-level model was used here, including level 1 (sex/ethnicity); level 2, all "subject" variations excluding sex, ethnicity, and age; and level 3, age. The mathematical formalism of mPLSR is given in an Appendix.

Results: Differences in facial shape between the ages of 12 and 17 predicted by mPLSR agree well with previous results of multilevel principal components analysis (mPCA); buccal fat is reduced with increasing age and features such as the nose, brow, and chin become larger and more distinct. Differences due to ethnicity and sex are also observed. Plausible simulated faces are predicted from the model for different ages, sexes and ethnicities. Our models provide good representations of the shape data by consideration of appropriate measures of model fit (RMSE and R^2).

Conclusions: Repeat measures in our dataset for the same subject at different ages can only be modelled indirectly at the lowest level of the model at discrete ages via mPCA. By contrast, mPLSR models age explicitly as a continuous covariate, which is a strong advantage of mPLSR over mPCA. These investigations demonstrate that multivariate multilevel methods such as mPLSR can be used to describe such age-related changes for dense 3D point data. mPLSR

might be of much use in future for the prediction of facial shapes for missing persons at specific ages or for simulating shapes for syndromes that affect facial shape in new subject populations.

Keywords

Multilevel partial least squares; 3D facial shape; adolescence

1 Introduction

3D facial imaging [1] can be used to explore facial shape changes during adolescence. There are also many other applications of such 3D imaging, e.g., in orthodontics [2-7] or in investigating various syndromes [8-12] that can affect facial shape. Facial simulation is of much interest for human–computer interfaces [13–15] and 3D (computed tomography) images have been used in facial reconstructions [15,16]. There are many different factors that can affect facial shape, including both genetic [18-25] and environmental factors [26-31]. In order to understand the effects that such influences can have on facial shapes, sophisticated computer and statistical methods must be employed to analyse the 3D facials scans and / or any extracted Cartesian landmark point data. In particular, and as noted in [32]: "Two general classes of morphometric methods can be used to analyse landmark coordinate data: coordinate-based methods and coordinate-free methods." Coordinate-based approaches use facial angles, distances between specific landmark points, and ratios of such distances in order to explore age-related changes in facial shape [32-37]. Multivariate statistics [38], multilevel univariate methods [39], and multivariate Bayesian-type methods [40] have been used to analyse such data. Incremental and positional changes over short- and long-term time periods have been explored [41].

Prediction of changes in facial shapes with age are useful for orthodontics, as well as for forensic purposes. Applications include facial reconstruction and synthetic growth or ageing. Facial shape changes over time in syndromic patients is also an important area of research. Age-related changes in facial size and shape have been investigated by multivariate modelling techniques, applied to coordinate data directly. These approaches were: simple "centring" and / or averaging techniques of facial shape at each age followed by a simple parametric model of growth [42], multivariate or kernel regression [43,44], principal components analysis (PCA) based methods [45–47], clustering and discriminant function analysis [48], and autoregressive moving averaging methods [49].

Multilevel principal components analysis (mPCA) [50-57] provides another multivariate method to fit multilevel models to shape coordinate data. mPCA has been used to investigate facial shape changes by ethnicity and sex [51,52], the act of smiling [53,54], maternal smoking and alcohol intake [55], and facial shape changes during adolescence [56,57]. Group centroids were found in [56,57] at (integer) ages 12 to 17 (i.e., 6 groups) in order to explore component scores at an appropriate level of the model. A limitation of this approach was that age was treated indirectly as a discrete rather continuous quantity.

Here we extend and improve upon these mPCA calculations by using a technique called multilevel partial-least squares regression (mPLSR) [58]. Unlike mPCA, mPLSR treats age explicitly as a continuous covariate. We present a discussion of multi-level regression, although the specifics of the various models are deferred to Appendices. We use mPLSR to investigate changes in shape between the ages of 12 and 17 and compare the results with mPCA. We then consider results of mPLSR for four different populations (both male and females for Welsh and Finnish subjects) at ages 12 to 17 years old. Finally, we present results for measures of model fit for single-level PLSR [59] compared to mPLSR.

2 Materials and Methods

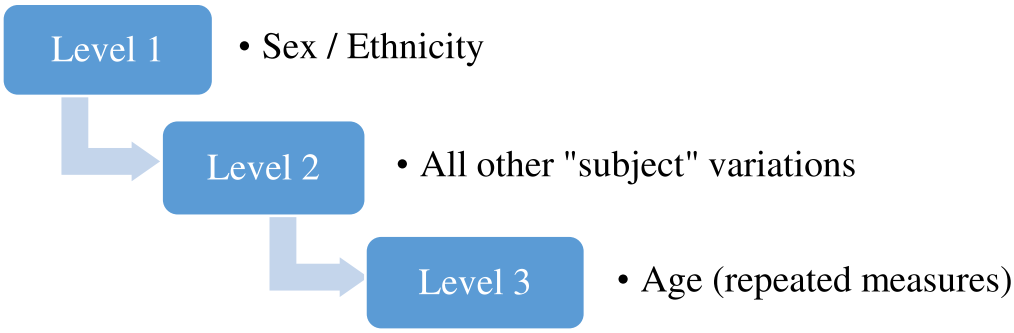

Figure 1. Schematic of a multilevel model of the effects of sex and age on facial shape.

729 3D facial images were collected for subjects aged 12 to 17 years old (5 to 10 shapes per subject at different ages). Sample sizes were: Welsh male = 28, Welsh female = 22, Finnish male = 25, Finnish female = 22." "Meshmonk" software [60] generated 1000 3D points regularly on each shape by using a non-rigid template. All subjects were healthy and of European descent. The data is not publicly available. All shapes were subsequently superimposed and scaled with respect to the mean shape using a generalized Procrustes analysis. Ethical approval for this study was obtained from the director of education, head teachers, school committees, and the relevant ethics committees of Bro Taf and Cardiff University (reference 04/WSE/109) and the City of Oulu (reference 7728/2006).

The mathematical formalisms of both mPCA and mPLSR are presented in Appendices A and B. However, underlying models used by mPCA and mPLSR are illustrated by Fig. 1. Sex and ethnicity are at level 1, all other differences between subjects except age, ethnicity and sex at level 2, and age at level 3. Note that no software exists specifically for mPCA and mPLSR applied to shapes and so an implementation in MATLAB was created by us. We calculate measures of model fit for traditional single-level PLSR and also mPLSR. The root-mean-squared error (*RSME*) is given by

$$RMSE = \sqrt{\frac{\sum_i^n \sum_k^d (z_{ik} - z_{ik}^{\text{model}})^2}{nd}} \quad , \tag{1}$$

where i indicates a specific shape of n (= 729 here), k indicates an element of the shape vector of length d (= 3000 here). The coefficient of multiple determination is given by

$$R^2 = 1 - \frac{\sum_i^n \sum_k^d (z_{ik} - z_{ik}^{\text{model}})^2}{\sum_i^n \sum_k^d (z_{ik} - \bar{z}_{ik})^2} \quad , \tag{2}$$

where $\bar{z}$ again indicates the mean shape vector over all shapes. A value of R^2 near to zero indicates poor model fit and a R^2 near to 1 indicates excellent to perfect model fit. Note finally quadratic models with age were tested here, although they did not provide very much additional accuracy in terms of model fit compared to a linear model. Results for the linear model with age that also contains terms for sex and ethnicity are therefore presented here only. Results for model fit for traditional single-level PLSR are presented here as a comparison to those results of mPLSR. Again, details for all of the models used here are presented in the Appendices.

3 Results

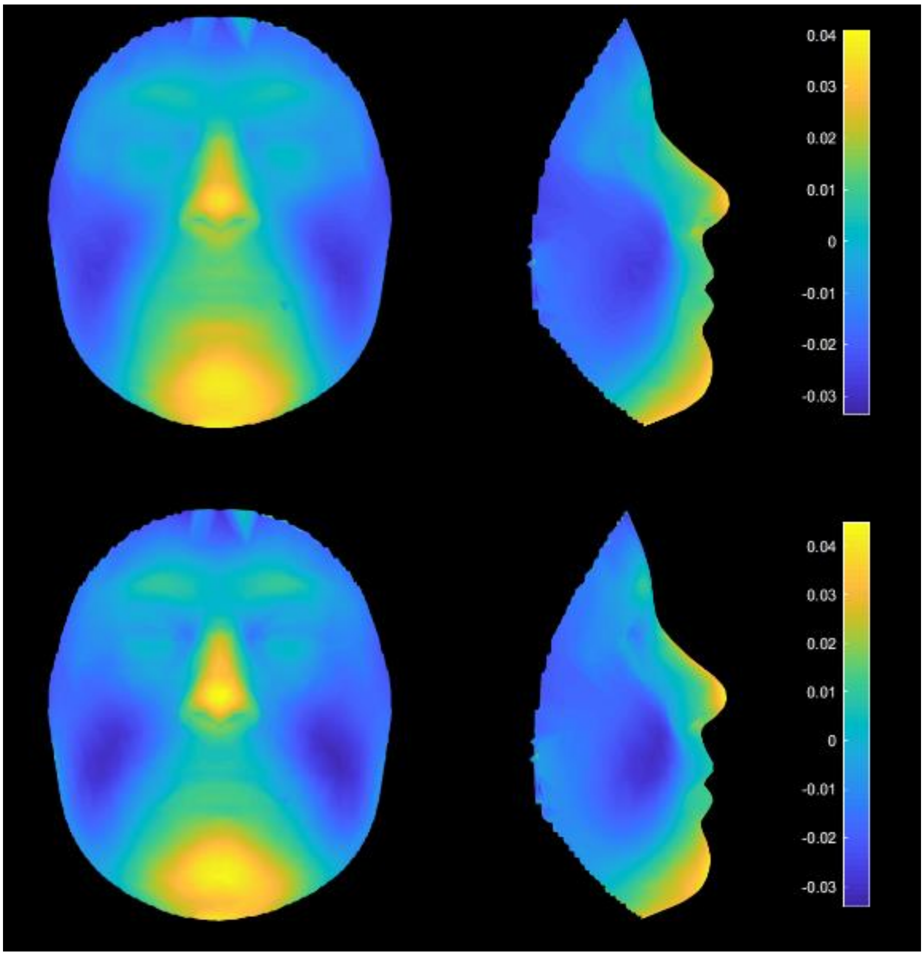

Fig. 2: Colour map of the projection representing differences in shapes normal to the surface at each point (broadly: blue indicates inward changes in mm and yellow indicates outward changes in mm) going from ages 12 to 17 using mPCA [57] (upper figure) and mPLSR (lower figure). Excellent correspondence is seen between the two approaches, although changes via mPLSR appear to be slightly more distinct.

Results of single-level PCA and mPCA (using the model shown in Fig. 1) for this dataset were presented previously in [57] and so we give a brief recap here only. For mPCA eigenvalues at level 1 (ethnicity/sex), level 2 (between-subjects) and level 3 (age) accounted for 7.9%, 71.5%, and 20.6% of shape variation in our dataset, respectively. Furthermore, results for modes of variation via mPCA demonstrated (correctly) changes in facial shape due to ethnicity and sex at level 1 and changes due to age at level 3. Evidence of clustering by ethnicity and sex and (separately) by age group was seen in the standardised component scores at levels 1 and 3, respectively, of the mPCA model (and also in scores for single-level PCA). It was also noted that the "trajectories" of centroids of standardised component scores for males and females appeared to be different at level 3 of the mPCA model and via single-level PCA. No obvious evidence of differences in trajectories between Welsh and Finnish subjects with age was observed. All of these results were good tests that mPCA provided reasonable results.

Facial shapes at ages 12 and 17 (independent of sex, ethnicity, and between subject variation) can be found via mPCA by using Eq. (A3) and then setting coefficients $\{a_l^1\}$ and $\{a_l^2\}$ to be zero at levels 1 and 2. Ref. [57] states that the centroid of component scores at level 3 (age) component 1 occurred at -$\sqrt{\lambda_1^3}$ for age 12 and at +$\sqrt{\lambda_1^3}$ for age 17. The difference perpendicular to the surface of the predicted shapes going from ages 12 to 17 via mPCA is shown in Fig. 2. As noted in Ref. [57], this mode of variation corresponds to an increase in the prominence of the chin and the size of the nose with increasing age, such that features become more strongly defined. Furthermore, the forehead also becomes slightly less prominent and cheeks also become less rounded with increasing age [57]. The difference perpendicular to the surface of the predicted shapes going from ages 12 to 17 via mPLSR of (e.g.) Eq. (B8) is also shown in Fig. 2. We see that results of mPCA and mPLSR are very similar quantitatively, which is an excellent first test of mPLSR. Unlike mPCA, mPLSR can be used to provide estimates of shape at any age readily because age is modelled as a continuous covariate.

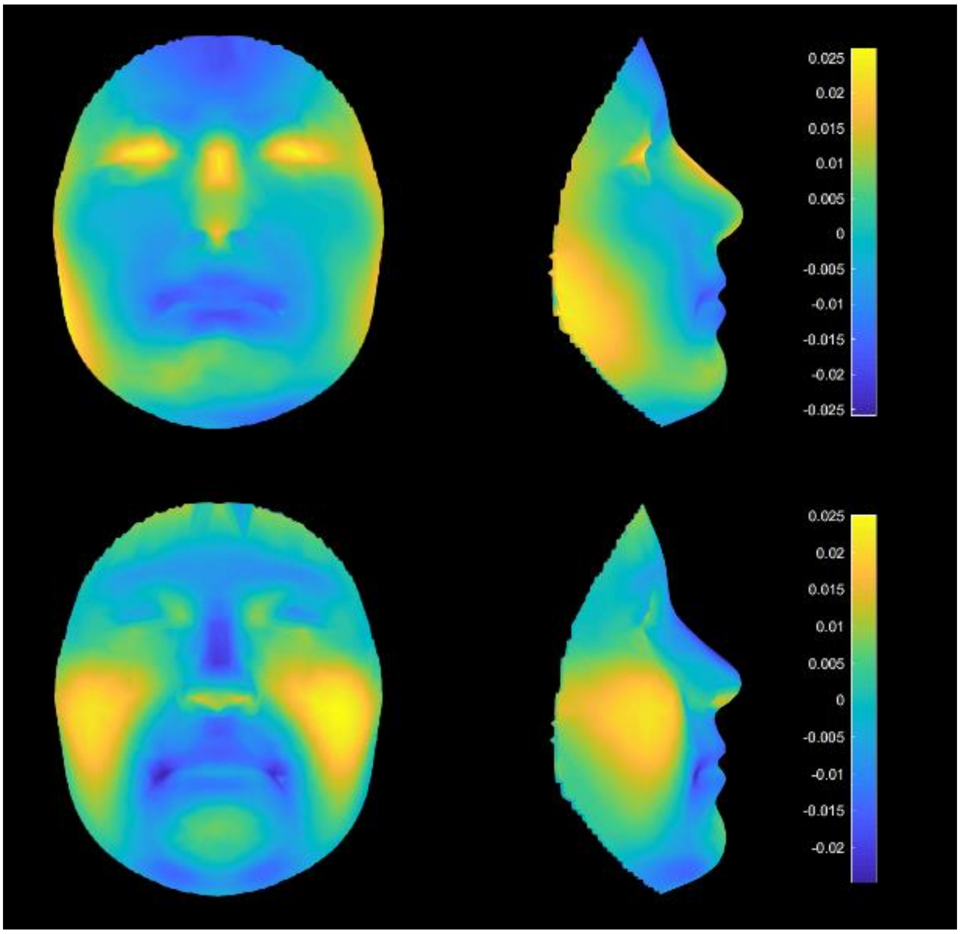

Fig. 3: Colour map of the projection perpendicular to the surface at each point (broadly: blue indicates inward changes in mm and yellow indicates outward changes in mm) represents differences in shapes going from Welsh to Finnish (upper figure) and going from male to female (lower figure) via mPLSR.

Differences in shapes between the two sexes and ethnicities predicted from mPLSR are shown as a colour map of the projections of the difference onto the surface normals, thus indicating change the locally inward-outward direction Fig. 3. Fig. 3 indicates that changes occur in the nose, chin, and cheeks between Finnish and Welsh subjects, whereas changes between sexes occur most prominently in the cheeks. Ref. [57] presents results for mean ± 2SDs via mPCA, and so the figures in Ref. [57] are "extreme" versions of shapes for the groupings by sex (male / female) and ethnicity (Welsh / Finnish). Thus, the magnitude of changes due to sex and ethnicity is smaller in Fig. 3 than in similar figures presented in Ref. [57], although the overall patterns are very similar, as expected.

Colour maps are useful in visualising the location and magnitude of changes. In addition we visualise the changes as facial shapes directly in Fig. 4. We see again in Fig. 4 that the nose and chin become more prominent and features more distinct with increasing age. These differences are more noticeable in the profile rather than in the frontal images. We note that mPLSR treats age as a continuous variable, so it is possible to produce shapes at any ages. This was not the case in Ref. [57], where age was grouped according to specific (integer) years. Interpolation to intermediate ages is more difficult for mPCA than mPLSR. Again, this is a strong advantage of mPLSR.

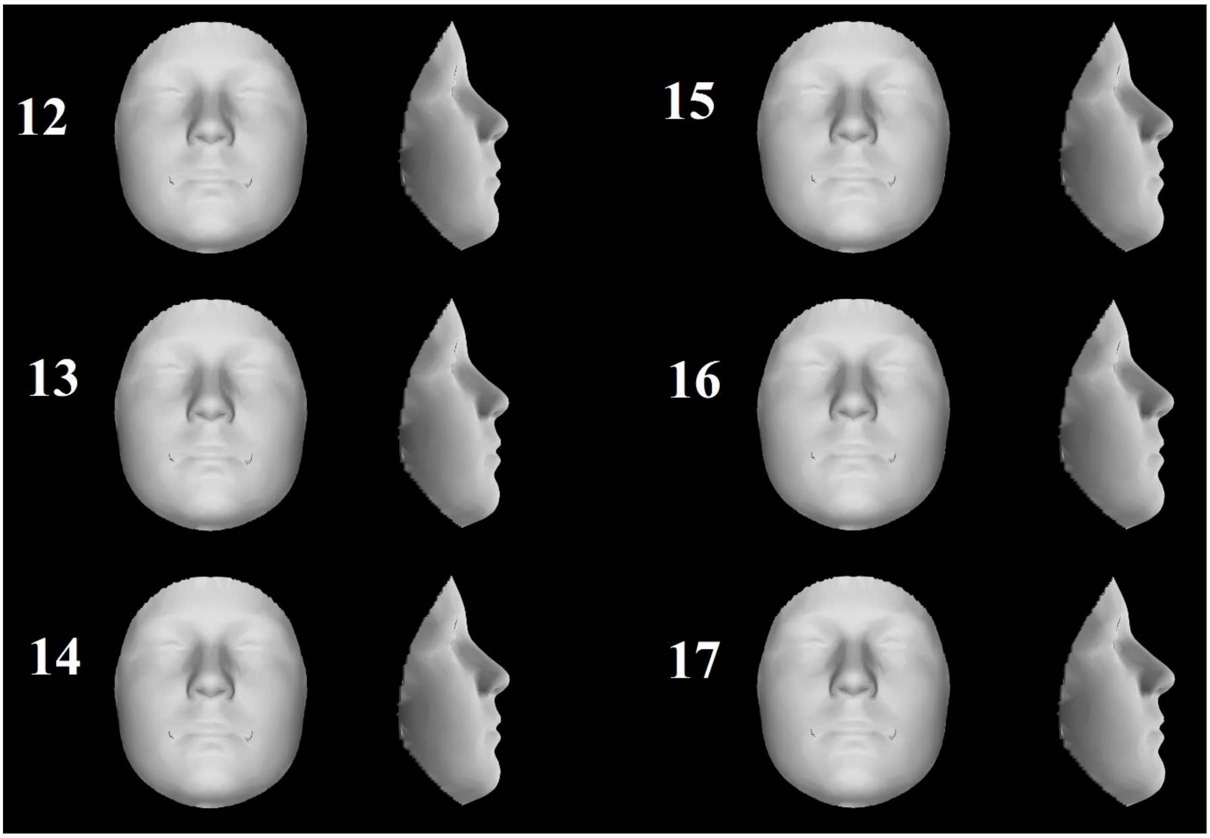

Fig. 4: Predicted faces of the PLSR 2 model of Eq. (B7) at all ages from 12 to 17. Facial features become more distinct, and the nose, forehead and chin become more prominent, with increasing age.

Results of the PLSR 3 model of Eq. (B9) are shown in Fig. 5. The effects of between-subject variation at level 2 can be illustrated by choosing coefficient $\{a_l^2\}$ randomly, where here we use: $a_l^2 = N(0, \sqrt{\lambda_l^2})$. Although changes with age still contribute to these facial shapes, the influence of shape variation at level 2 is clearly a much stronger influence. Indeed, we would expect level 2 variations to dominate because mPCA suggests that level 2 contributes 71.5% of shape variation, whereas level 3 contributes only 20.6%. Furthermore, all potential facial shapes are more varied than in Fig. 4, although all of them are still plausible. This is an excellent qualitative test of the model in Eq. (B9).

Results for measures of model fit for the case where the test and training datasets are the same are shown in Table 1. Furthermore, generalizability of our findings was explored by using cross-validation, where the data was split randomly into training (80%) and test sets (20%), and the results are also shown in Table 1. The simplest models of single-level PLSR 1 of Eq. (B5) and mPLSR 1 of Eq. (B7) contain age only and so it is not surprisingly that these approximations demonstrate the largest *RMSE* and the lowest values for R^2. This is improved somewhat by including sex and ethnicity in the model for single-level PLSR 2 of Eq. (B6). mPLSR 2 of Eq. (B8) gives very similar results to those of PLSR 2 of Eq. (B6), which is also not surprising as only age, sex, and ethnicity are included only in both models (albeit coded in slightly different ways). Clearly, results for *RMSE* and R^2 in Tables 1and 2 indicate that PLSR 4 of Eq. (B10) models the data accurately. Clearly, we wish to avoid both over- and under-fitting in our models and we note again that cross-validation is carried out here. Measures of model fit can also be examined with respect to the number of components, p_2, retained at level 2 of the model; Table 2 shows that these measures of model fit improve with increasing values of p_2. Results for measures of model fit were also found for models that included additional quadratic terms with age, although these analyses led to marginal improvements only in model fit and so they are not presented here.

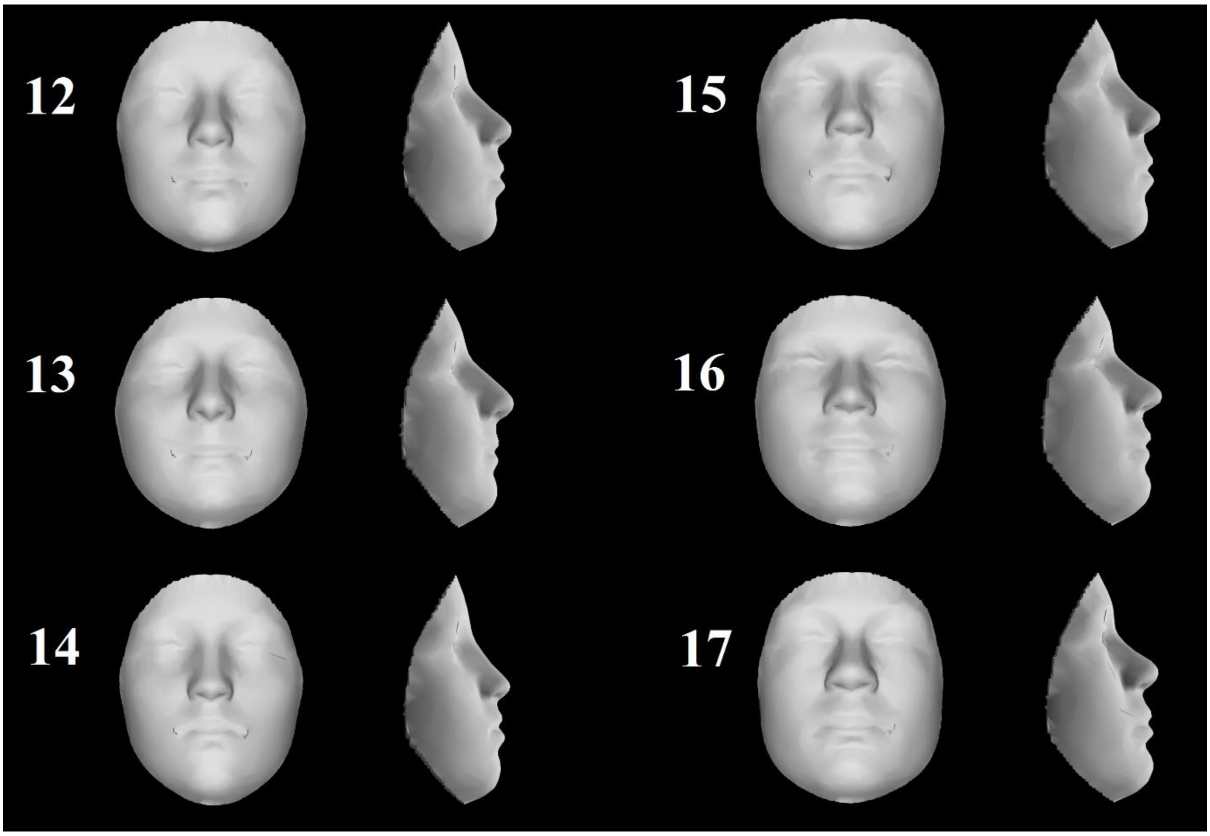

Fig. 5: Results of the PLSR 3 model of Eq. (B9) at all ages from 12 to 17. Contributions at level 2 sampled randomly for all shapes are used to give an idea how this level affects facial shape. All predicted facial results are plausible, which is a good qualitative test of this model.

Model		PLSR 1 (age only)	PLSR 2 (age, sex, ethnicity)	mPLSR 1 (age only)	mPLSR 2 (age, sex, ethnicity)	mPLSR 4
Data not split	*RMSE*	72.53	70.30	72.56	70.15	12.56
	R^2	0.036	0.094	0.035	0.098	0.971
Data Split: 80% training and 20% testing	*RMSE*	73.14	72.59	73.16	70.84	13.79
	R^2	0.032	0.046	0.031	0.092	0.966

Table 1: Results for measures of model fit for the *RMSE* and R^2 for both single-level PLSR and mPLSR. As expected, model fits improve with increased model complexity. (Note that p_2 = 50 components are retained at level 2 of this model for mPLSR 4.) Results are presented for the cases where data was not split (i.e., the test and training datasets were one-and-the-same) and where the data was split into training (80% selected randomly; n = 583) and test sets (remaining 20%; n = 146). (R^2 and *RMSE* are evaluated for the testing set only.)

p_2		1	2	5	10	50
Data not split	*RMSE*	60.88	53.18	43.47	33.20	12.56
	R^2	0.320	0.481	0.654	0.798	0.971
Data Split: 80% training and 20% testing	*RMSE*	61.98	54.44	44.24	33.78	13.79
	R^2	0.305	0.464	0.646	0.794	0.966

Table 2: Results for measures of model fit for the *RMSE* and R^2 for mPLSR 4 given as function of the number of components, p_2, retained at level 2 of the model. Results are presented for the cases where data was not split (i.e., the test and training datasets were one-and-the-same) and where the data was split into training (80% selected randomly; $n = 583$) and test sets (remaining 20%; $n = 146$). (R^2 and *RMSE* are evaluated for the testing set only.)

4 Discussion

We have shown here that multilevel multivariate regression methods (mPLSR here) can be used to study age-related changes in adolescents in different populations (i.e., Finnish and Welsh populations containing both males and females). Excellent quantitative agreement was observed between mPLSR and mPCA for predicted differences in shape between the ages of 12 and 17 (see Fig. 2). A detailed exploration of mPCA results for facial changes with age in adolescents is presented in the discussion section of [57]. However, we note here that results of mPLSR (and mPCA) also agree with the literature in this field, e.g. as stated in [61] that: “A chubby face turns into an elongated and more distinct face during the growth” and that “Facial features, such as the nose and the eyes, become more prominent and evolve to the main characteristics of a face.” Figs. 4 and 5 demonstrated that reasonable results were seen for the four sex / ethnicity groups, albeit by assuming a common ageing trajectory. However, the two sexes are known to exhibit different growth patterns during adolescence, e.g., that most changes occurred in the lower third of the face for males [26], and that surface changes are greater (with more changes occurring later) for boys compared girls [27]. In principle, multilevel multivariate regression models ought to be able to include such differences in growth trajectories by sex (etc.) and this will be carried out in future studies.

A distinct advantage of direct regression models such as mPLSR compared to, say, mPCA is that age can be modelled explicitly as a continuous covariate. Note that ages must be “binned” together in some way for mPCA (to the nearest integer in [57]). Thus, interpolating shapes via mPCA at other intermediate ages is more difficult because we must carry out curve or surface fitting of component scores as a function of age in order to carry out this interpolation. Furthermore, models that reflect the structure of subject populations ought to perform better than those that do not and some early evidence for Monte Carlo simulated data [51] indicate that this is indeed the case for mPCA compared to single-level PCA. Furthermore, eigenvectors obtained at different levels of the model in mPCA (or indeed mPLSR) do not necessarily have to be orthogonal, although they must be orthogonal within a given level. In principle, such approaches provide a route to overcoming the "orthogonality problem" of single-level PCA-type methods that can mix different effects together. By explicitly modelling (e.g.) sex and ethnicity at specific levels of the model, we hope to improve generalizability compared to those methods that do not model the effects of these factors explicitly. Prediction of facial shapes to new scenarios is considered below. However, we note that a strong advantage of PCA-type methods in particular is that one can constrain the expansion coefficients so that the model is constrained to realistic solutions, within the modelled limits of variation within the training data when fitting new cases to the model (e.g., $|a_l^2| \leq 3\sqrt{\lambda_l^2}$ at level 2 of the model presented here).

The number of groups at a given level is often quite small (e.g., males & females at level 1 here) and this will limit the complexity of the underlying model at this level. In some cases, this might lead to overly simplistic models. Furthermore, the rank of covariance matrices and so the number of non-zero eigenvalues is limited to the *number of groups* – 1 at that level. In turn, this will limit the number of non-zero eigenvalues at this level in PCA-type methods. Furthermore, another criticism that has been levelled against between-groups PCA [62] (a form of two-level mPCA) is that it can overestimate differences between groups because "between-group" variation (level 1 in our terminology) is represented well by differences between mean shapes, whereas within-group variation (level 2) can be underestimated. The author [62] makes some excellent cautionary remarks about the validity of such techniques in the high *p/n* limit, where the number of parameters in the model far exceeds the sample size.

mPCA [57] suggests that only 7.9% and 20.6% of variation is explained by sex / ethnicity (level 1) and age (level 3), respectively, which helps us to explain why values for R^2 for regression models that include sex, ethnicity, and age are relatively modest only in Tab. 1 (i.e., R^2 = 0.094 and 0.098). mPCA provides a complementary analysis to mPLSR therefore. Higher values of R^2 were obtained in Tab. 1 by including contributions in the regression model at level 2 ("between subject" variation). Caution should be exercised when interpreting R^2 in this case though as this is somewhat of a false comparison: deviations at level 2 are modelled by a complete set of orthogonal vectors and so (e.g.) we would expect the model fit to be perfect if all possible (3000 here) eigenvectors were included in the model. Furthermore, here we wished only to observe how model fits changed with increasing complexity of model in these initial exploration of method. Despite these points, it is still encouraging that excellent model fits can be obtained for relatively modest numbers of eigenvectors at level 2, i.e., p_2= 50 for this high dimensional data, as shown in Tab. 1.

Indeed, we see from Fig. 5 that plausible shapes can be obtained via Eq. (B9) when contributions are level 2 ("between subjects") are sampled randomly, which is another excellent (qualitative) test of this approach. A strength of such multilevel multivariate methods (illustrated schematically by Fig. 1) compared to single-level methods is that they are extremely flexible. For example, we can readily include such "between-subjects" variations within our multilevel model. In these initial calculations, we wish to explore how model fit changes in broad terms as we include different types of terms within the basic model.

By fitting the model for mPLSR 4 of Eq. (B10) to a new facial shape, we can find the level 2 (between-subject) parameters $\{a_l^2\}$. We can then fix these parameters $\{a_l^2\}$ in order to extrapolate this shape to, say, older ages, given (e.g.) an appropriate model that takes into account "plateauing" after the adolescent growth spurt. Clearly, this could be of much use in forensics for the identification of missing persons. Other parameters such as sex or ethnicity at level 1 can be changed also once $\{a_l^2\}$ are fixed. Facial shapes for syndromes such as Treacher Collins syndrome or foetal alcohol syndrome have been investigated previously (e.g.) in the UK. Rather less information is available in some countries or ethnicities. This approach could therefore also be of much use in simulating facial types for syndromes such as Treacher Collins syndrome or foetal alcohol syndrome in new populations where we have less information. In turn, this could be very useful for clinical diagnosis in that it can assist in delineating characteristic facial features, indicative of theses underlying disorders, in patients from these countries or ethnicities where less information is available. Finally, the methods presented here could have many other potential applications in future, including: 3D facial reconstruction, facial identification, human-computer interfaces, and computer-generated imagery.

5 Conclusions

We have shown in this article that mPLSR can be applied to dense 3D point data representing facial shape. These calculations also demonstrate that this regression method can be used to describe such changes in facial shape due to age in adolescents. mPLSR models age as an explicit covariate, which is a strong advantage of mPLSR over mPCA because we can make a prediction of facial shape at any age. mPLSR might therefore be of much use in future for the prediction of facial shapes for missing persons at specific ages or for simulating shapes for syndromes that affect facial shape in new subject populations.

Acknowledgments

This investigation was supported by the KU Leuven, BOF (C14/15/081), NIH (1-RO1-DE027023) and the FWO Flanders (G078518N).

Appendix A: mPCA Mathematical Formalism

An excellent review of PCA and PLSR methods is given in Ref. [63] and we adopt a broadly similar mathematical formalism here. To begin with, however, 3D landmark points are represented by a vector z (length 3000 = number of points, 1000 × 3 dimensions) for each shape. Single-level PCA is carried out by finding the mean shape vector $\bar{z}$ over all shapes and a covariance matrix

$$K_{k_1,k_2} = \frac{1}{N-1}\sum_{i=1}^{n}(z_{ik_1} - \bar{z}_{ik_1})\,(z_{ik_2} - \bar{z}_{ik_2}) \quad . \tag{A1}$$

k_1 and k_2 indicate elements of this covariance matrix and i refers to a given subject. The eigenvalues λ_l and (orthonormal) eigenvectors u_l of this matrix are found readily. For PCA, one ranks all the eigenvalues into descending order, and one retains the first p_1 components in the model. The shape z is modeled by

$$z^{\mathrm{PCA}} = \bar{z} + \sum_{l=1}^{p_1} a_l u_l \quad . \tag{A2}$$

The coefficients $\{a_l\}$ (also referred to as "component scores" here) are found readily by using a scalar product with respect to the set of orthonormal eigenvectors, i.e., $a_l = u_l \cdot (z - \bar{z})$, for a fit of the model to a new shape vector z. Note that Eq. (A2) can be restated in a more general matrix form as $z^{\mathrm{PCA}} = \bar{z} + Au$, where U is formed of p_1 column of orthogonal vectors u_l of length d (= 3000 here) and A is formed of p_1 row vectors of the component scores a_l of length n (= 729 here).

mPCA allows us to isolate the effects of various influences on shape at different levels of the model and this is illustrated schematically in Fig. 1 given above. The covariance matrix at level 3 is formed with respect to all "repeated measurements" of 3D facial shape for each subject individually (at different ages) and then these covariance matrices are averaged over all subjects to give the level 3 covariance matrix, K^3. The average shapes over all ages for each subject l_2 in each sex / ethnicity group l_1 are found, such that each subject "forms their own group" at level 2. This covariance matrix K^2 at level 2 is formed with respect to these shapes and they are averaged over all four sex / ethnicity groups (here) to give the level 2 covariance matrix, K^2. Finally, the covariance matrix K^1 at level 1 is formed with respect to these average shapes for the four sex / ethnicity groups. The overall "grand mean" shape is again denoted by $\bar{z}$. The l-th eigenvalue at level 1 is denoted by λ_l^1, with associated eigenvector u_l^1, the l-th eigenvalue at level 2 is denoted by λ_l^2, with associated eigenvector u_l^2, and the l-th eigenvalue at level 3 is denoted by λ_l^3, with associated eigenvector u_l^3. We rank the eigenvalues into descending order at each level of the model separately, and then we retain the first p_1, p_2 and p_3 eigenvectors of largest magnitude at the three levels. The shape z is modeled by

$$z^{\text{mPCA}} = \bar{z} + \sum_{l=1}^{p_1} a_l^1 u_l^1 + \sum_{l=1}^{p_2} a_l^2 u_l^2 + \sum_{l=1}^{p_3} a_l^3 u_l^3 \ . \tag{A3}$$

The coefficients $\{a_l^1\}$, $\{a_l^2\}$, and $\{a_l^3\}$ (also referred to as "component scores" here) are determined for mPCA by using a global optimization procedure in MATLAB with respect to a "least-squares-type" cost function (i.e., $cost = \varepsilon^2 = \varepsilon \cdot \varepsilon$, where ε is the residual error vector). Equation (A3) may be expressed in matrix form given by

$$Z^{\text{mPCA}} = \bar{Z} + A^1 U^1 + A^2 U^2 + A^3 U^3 \ , \tag{A4}$$

where U^1 is formed of p_1 column of orthogonal vectors u_l^1 of length d (= 3000 here) and A^1 is formed of p_1 row vectors of the component scores a_l^1 of length n (= 729 here), etc.

Appendix B: mPLSR Mathematical Formalism

Here we use Partial Least-Squares Regression (PLSR) in order to carry out multivariate regression. PLSR is a method that uses a weight matrix W that aims to reflect the covariance structure between a response (dependent) Z (i.e., a matrix containing shape vectors for all subjects) and predictor (independent) variables X (such as sex, ethnicity, and age here). (Z and X are assumed to have means of zero.) Ultimately, PLSR aims to model

$$Z = XB \quad . \tag{B1}$$

However, we define firstly a "factor score" matrix T, which is given by

$$T = XW \quad , \tag{B2}$$

where W is an appropriate matrix of weights (mentioned above). Linear regression of Z on T produces regression coefficients Q, such that

$$Z = TQ \quad . \tag{B3}$$

Equations (B1) to (B3) imply that

$$B = WQ \quad . \tag{B4}$$

There are different ways of defining the weight matrix W and here we use the MATLAB "plsregress" command, which uses the SIMPLS algorithm [68]. The first single-level PSLR model of shape vector z with respect to age only is therefore given by

$$z^{PLSR\ 1} = B_{age} \times age \tag{B5}$$

where *age* is the age of subject for shape vector z. (Lower case z used here as this is a single shape.) The second single-level PSLR model including age, sex, and ethnicity is therefore given by

$$z^{PLSR\ 2} = B_{age} \times age + B_{sex} \times sex + B_{ethnicity} \times ethnicity. \quad \text{(B6)}$$

where *sex* (0 = male; 1 = female) and *ethnicity* (0 = Welsh; 1 = Finnish). Equations (B5) and (B6) are fixed-effects models. (A quadratic form with age can be formulated readily also, although only marginal gains were found here.)

A simple form of mPLSR [58] is carried out here by performing PLSR at level 3 for data Z' given by shape vectors $z_i{}' = z_i - \gamma_i$. The index i indicates a specific subject here. The vector γ_i indicates the mean shape over facial shapes at different ages for this subject i. Hence, we have "re-centred" the data at level 3 with respect to the "group mean shape" for each subject. We have therefore removed much of the between-subject variations, including variation between sexes and ethnicities. We now find the "gradients" B^3 via PLSR at level 3 for dependent variables Z' with respect to the independent variables X used at this level (here age only). A first (very simple) mPLSR model of shape vector z that is linear with age is therefore given by

$$z^{mPLSR\ 1} = \bar{z} + B^3 \times age \ . \quad \text{(B7)}$$

Note that *age* in Eq. (B7) again indicates the age of subject for a shape vector z. A slightly more complicated model includes also the effects of sex and ethnicity at level 1 and the simplest method of achieving this is to use the mean shapes μ_l for the given sex / ethnicity group l for this subject (level 1), such that

$$z^{mPLSR\ 2} = \mu_l + B^3 \times age \ . \quad \text{(B8)}$$

An extension of this equation is to include terms that account for "between-subject variations" at level 2. Here we use the principal components obtain at level 2 presented above to represent this source of variation. The final mPLSR models are now given by

$$z^{mPLSR\ 3} = \bar{z} + A^2 U^2 + B^3 \times age \ , \quad \text{(B9)}$$

and

$$z^{mPLSR\ 4} = \mu_l + A^2 U^2 + B^3 \times age \ . \tag{B10}$$

Arguably, this approach is analogous to a (multivariate multilevel) mixed-effects model. It is straightforward to demonstrate that the coefficients $\{a_l^2\}$, e.g., for a specific shape z with respect to Eq. (B10), are given by

$$a_l^2 = u_l^2 \cdot (z - \mu_l - B^3 \times age) \ . \tag{B11}$$

As in active shape models (ASMs) and active appearance models (AAMs), one can constrain the expansion coefficients (e.g., $|a_l^2| \leq 3\sqrt{\lambda_l^2}$) so that the model never "strays too far" from a sensible solution. (Note that this is used here when finding measures of model fit.) Indeed, this approach could be used in both ASMs and AAMs to account for continuous covariates (etc.) when segmenting features in images, although this lies beyond the scope of this paper. Quadratic forms with age in Eqs. (B7) to (B10) can also be formulated readily also, although again this process only yielded marginal gains and so this is not discussed here further here.

References

1. A.R. Al-Khatib, Facial three dimensional surface imaging: An overview. *Archives of Orofacial Sciences* **5** (2010) 1–8.
2. C.H. Cau, A. Cronin, P. Durning, A.I. Zhurov, A. Sandham, S. Richmond, A new method for the 3D measurement of postoperative swelling following orthognathic surgery. *Orthodontic Craniofacial Research* **9** (2006) 31–37.
3. B. Krneta, J. Primožič, A.I. Zhurov, S. Richmond, M. Ovsenik, Three-dimensional evaluation of facial morphology in children aged 5–6 years with a Class III malocclusion. *European Journal of Orthodontics* **36** (2012) 133–139.
4. J. Djordjevic, D.A. Lawlor, A.I. Zhurov, A.M. Toma, R. Playle, S. Richmond, A population-based cross-sectional study of the association between facial morphology and cardiometabolic risk factors in adolescence. *BMJ Open* **3** (2013) e002910.
5. H. Popat, A.I. Zhurov, A.M. Toma, S. Richmond, D. Marshall, P.L. Rosin, Statistical modeling of lip movement in the clinical context. *Orthodontic Craniofacial Research* **15** (2015) 92–102.
6. M. Alqattan, J. Djordjevic, A.I. Zhurov, S. Richmond, Comparison between landmark and surface-based three-dimensional analyses of facial asymmetry in adults. *European Journal of Orthodontics* **37** (2013) 1–2.
7. A. Al Ali, S. Richmond, H. Popat, R. Playle, T. Pickles, A.I. Zhurov, D. Marshall, P.L. Rosin, J. Henderson, K. Bonuck, The influence of snoring, mouth breathing and apnoea on facial morphology in late childhood: A three-dimensional study. *BMJ Open* **5** (2015) e009027.
8. K. Jones and D. Smith, Recognition of the fetal alcohol syndrome in early infancy. *The Lancet* **302** (1973) 999–1001.

9. A.P. Streissguth, J.M. Aase, S.K. Clarren, S.P. Randels, R.A. LaDue, and D.F. Smith, Fetal alcohol syndrome in adolescents and adults. *Journal of the American Medical Association* **265** (1991) 1961–1967.
10. J. McKenzie and J. Craig, Mandibulo-facial dysostosis (Treacher Collins syndrome). *Archives of Disease in Childhood* **30** (1955) 391.
11. D. Poswillo, The pathogenesis of the Treacher Collins syndrome (mandibulofacial dysostosis). *British Journal of Oral Surgery* **13** (1975) 1–26.
12. J.E. Allanson, P. O'Hara, L.G. Farkas, and R.C. Nair, R.C., Anthropometric craniofacial pattern profiles in Down syndrome. *American Journal of Medical Genetics* **47** (1993) 748–752.
13. J. Vandeventer, 4D (3D Dynamic) Statistical Models of Conversational Expressions and the Synthesis of Highly-Realistic 4D Facial Expression Sequences. Ph.D. Thesis, Cardiff University, Cardiff, UK, 2015.
14. J. Vandeventer, L. Graser, M. Rychlowska, P.L. Rosin, D. Marshall, Towards 4D coupled models of conversational facial expression interactions. In *Proceedings of the British Machine Vision Conference;* BMVA Press: Durham, United Kingdom (2015) 142–141.
15. K. Al-Meyah, D. Marshall, P. Rosin, 4D Analysis of Facial Ageing Using Dynamic Features. *Procedia Computer Science* **112** (2017) 790–799.
16. S.D.S. Rocha, D.L.D.P. Ramos, and M.D.G.P. Cavalcanti, Applicability of 3D-CT facial reconstruction for forensic individual identification. *Pesquisa Odontológica Brasileira* **17** (2003) 24–28.
17. W.J. Lee, C.M. Wilkinson, and H.S. Hwang, An accuracy assessment of forensic computerized facial reconstruction employing cone-beam computed tomography from live subjects. *Journal of Forensic Sciences* **57** (2012) 318–327.

18. G. Fatemifar, C.J. Hoggart, L. Paternoster, J.P. Kemp, I. Prokopenko, M. Horikoshi, V.J. Wright, J.H. Tobias, S. Richmond, A.I. Zhurov, Genome-wide association study of primary tooth eruption identifies pleiotropic loci associated with height and craniofacial distances. *Human Molecular Genetics* **22** (2013) 3807–3817.
19. P. Claes, H. Hill, M.D. Shriver, Toward DNA-based facial composites: Preliminary results and validation. *Forensic Science International: Genetics* **13** (2014) 208–216.
20. P. Claes, D.K. Liberton, K. Daniels, K.M. Rosana, E.E. Quillen, L.N. Pearson, B. McEvoy, M. Bauchet, A.A. Zaidi, W. Yao, and H. Tang, Modeling 3D facial shape from DNA. *PLoS Genetics* **10** (2014) e1004224.
21. P. Claes, J. Roosenboom, J.D. White, T. Swigut, D. Sero, J. Li, M.K. Lee, A. Zaidi, B.C. Mattern, C. Liebowitz, and L. Pearson, Genome-wide mapping of global-to-local genetic effects on human facial shape. Nature Genetics **50** (2018) 414–423.
22. S. Richmond, L.J. Howe, S. Lewis, E. Stergiakouli, A. Zhurov, Facial Genetics: A Brief Overview. *Frontiers in Genetics* **9** (2018) 462.
23. S.M. Hopman, J.H. Merks, M. Suttie, R.C. Hennekam, and P. Hammond, Face shape differs in phylogenetically related populations. *European Journal of Human Genetics* **22** (2014) 1268–1271.
24. S. Leslie, B. Winney, G. Hellenthal, D. Davison, A. Boumertit, T. Day, K. Hutnik, E.C. Royrvik, B. Cunliffe, D.J. Lawson, and D. Falush, D., 2015. The fine-scale genetic structure of the British population. *Nature* **519** (2015) 309.
25. M. Nelis, T. Esko, R. Mägi, F. Zimprich, A. Zimprich, D. Toncheva, S. Karachanak, T. Piskáčková, I. Balaščák, L. Peltonen, and E. Jakkula, Genetic structure of Europeans: a view from the North–East. *PloS One* **4** (2009) e5472.
26. L. Mamluk, H.B. Edwards, J. Savović, V. Leach, T. Jones, T.H. Moore, S. Ijaz, S.J. Lewis, J.L. Donovan, D. Lawlor, and G.D. Smith, G.D., Low alcohol consumption and pregnancy

and childhood outcomes: time to change guidelines indicating apparently ‘safe’ levels of alcohol during pregnancy? A systematic review and meta-analyses. *BMJ Open* **7** (2017) e015410.

27. E. Muggli, H. Matthews, A. Penington, P. Claes, C. O’Leary, D. Forster, S. Donath, P.J. Anderson, S. Lewis, C. Nagle, and J.M. Craig, Association between prenatal alcohol exposure and craniofacial shape of children at 12 months of age. *JAMA Pediatrics* **171** (2017) 771–780.
28. L.J. Howe, G.C. Sharp, G. Hemani, L. Zuccolo, S. Richmond, and S.J. Lewis, Prenatal alcohol exposure and facial morphology in a UK cohort. *Drug and Alcohol Dependence* **197** (2019) 42–47.
29. H. Mirghani, N. Osman, S. Dhanasekaran, H.M. Elbiss, and G. Bekdache, Transplacental transfer of 2-naphthol in human placenta. *Toxicology Reports* **2** (2015) 957–960.
30. N. Pound, D.W. Lawson, A.M. Toma, S. Richmond, A.I. Zhurov, I.S. Penton-Voak, Facial fluctuating asymmetry is not associated with childhood ill-health in a large British cohort study. *Proceedings of the Royal Society B: Biological Sciences* **281**, no. **1792** (2014) 20141639.
31. J. Djordjevic, A.I. Zhurov, S. Richmond, Visigen Consortium. Genetic and Environmental Contributions to Facial Morphological Variation: A 3D Population-Based Twin Study. *PLoS ONE* **11** (2016) e0162250.
32. J.T. Richtsmeier, J.M. Cheverud, and S. Lele, S., Advances in anthropological morphometrics. *Annual Review of Anthropology* **21** (1992) 283–305.
33. R.M. Ricketts, The influence of orthodontic treatment on facial growth and development. *The Angle Orthodontist* **30** (1960) 103–133.
34. E.O. Bergersen, The male adolescent facial growth spurt: its prediction and relation to skeletal maturation. *The Angle Orthodontist* **42** (1972) 319–338.

35. S.J. Chaconas, and J.D. Bartroff, Prediction of normal soft tissue facial changes. The Angle Orthodontist *The Angle Orthodontist* **45** (1975) 12–25.

36. A. Verdonck, M. Gaethofs, C. Carels, and F. de Zegher, Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *The European Journal of Orthodontics* **21** (1999) 137–143.

37. S.E. Bishara, Facial and dental changes in adolescents and their clinical implications. *The Angle Orthodontist* **70** (2000) 471–483.

38. S.N. Bhatia, G.W. Wright, and B.C. Leighton, A proposed multivariate model for prediction of facial growth. *American Journal of Orthodontics* **75** (1979) 264–281.

39. B.A. Chvatal, R.G. Behrents, R.F. Ceen, and P.H. Buschang, Development and testing of multilevel models for longitudinal craniofacial growth prediction. *American Journal of Orthodontics and Dentofacial Orthopedics* **128** (2005) 45–56.

40. D.J. Rudolph, S.E. White, and P.M. Sinclair, Multivariate prediction of skeletal Class II growth. *American Journal of Orthodontics and Dentofacial Orthopedics* **114** (1998) 283–291.

41. B.J. Turchetta, L.S. Fishman, and J.D. Subtelny, Facial growth prediction: a comparison of methodologies. *American Journal of Orthodontics and Dentofacial Orthopedics* **132** (2007) 439–449.

42. M.L. Moss, R. Skalak, M. Shinozuka, H. Patel, L. Moss-Salentijn, H. Vilmann, and P. Mehta, Statistical testing of an allometric centered model of craniofacial growth. *American Journal of Orthodontics* **83** (1983) 5–18.

43. H.Y. Suh, S.J. Lee, Y.S. Lee, R.E. Donatelli, T.T. Wheeler, S.H. Kim, S.H. Eo, and B.M. Seo, A more accurate method of predicting soft tissue changes after mandibular setback surgery. *Journal of Oral and Maxillofacial Surgery* **70** (2012) e553–e562.

44. H.S. Matthews, A.J. Penington, R. Hardiman, Y. Fan, J.G. Clement, N.M. Kilpatrick, and P. Claes, P.D., Modelling 3D craniofacial growth trajectories for population comparison and classification illustrated using sex-differences. *Scientific Reports* **8** (2018) 4771.

45. E.D. Schneiderman, S.M. Willis, C.J. Kowalski, and T.R. Ten Have, T.R., A PC program for growth prediction in the context of Rao's polynomial growth curve model. *Computers in Biology and Medicine* **22** (1992) 181-188.

46. J. Koudelová, J. Dupej, J. Brůžek, P. Sedlak, and J. Velemínská, Modelling of facial growth in Czech children based on longitudinal data: Age progression from 12 to 15 years using 3D surface models. *Forensic Science International* **248** (2015) 33–40.

47. R.J. Hennessy and J.P. Moss, Facial growth: separating shape from size. *The European Journal of Orthodontics* **23** (2001) 275–285.

48. E.S.J. Abu Alhaija and A. Richardson, Growth prediction in Class III patients using cluster and discriminant function analysis. *The European Journal of Orthodontics* **25** (2003) 599–608.

49. A.J. Lin, S. Lai. and F. Cheng, Growth simulation of facial/head model from childhood to adulthood. *Computer-Aided Design and Applications* **7** (2010) 777–786.

50. F. Lecron, J. Boisvert, M. Benjelloun, H. Labelle, and S. Mahmoudi, Multilevel statistical shape models: A new framework for modeling hierarchical structures, *9th IEEE International Symposium on Biomedical Imaging (ISBI)* (2012) 1284–1287.

51. D.J.J. Farnell, H. Popat, and S. Richmond, Multilevel principal component analysis (mPCA) in shape analysis: A feasibility study in medical and dental imaging, *Computer Methods and Programs in Biomedicine* **129** (2016) 149–159.

52. D.J.J. Farnell, J. Galloway, A.I. Zhurov, S. Richmond, P. Perttiniemi, and V. Katic, Initial Results of Multilevel Principal Components Analysis of Facial Shape, *Annual Conference on Medical Image Understanding and Analysis*. Springer, Cham. (2017) 674–685.

53. D.J.J. Farnell, J. Galloway, A.I. Zhurov, S. Richmond, P. Perttiniemi, and R. Lähdesmäki, What's in a Smile? Initial Results of Multilevel Principal Components Analysis of Facial Shape and Image Texture. In *Medical Image Understanding and Analysis*, Springer, Cham. **894** (2018) 177–188.

54. D.J.J. Farnell, J. Galloway, A.I. Zhurov, S. Richmond, D. Marshall, P.L. Rosin, K. Al-Meyah, P. Perttiniemi, and R. Lähdesmäki What's in a Smile? Initial Analyses of Dynamic Changes in Facial Shape and Appearance, *Journal of Imaging* **5** (2019) 2.

55. J. Galloway, D.J.J. Farnell, S. Richmond, and A.I. Zhurov. Multilevel Analysis of the Influence of Maternal Smoking and Alcohol Consumption on the Facial Shape of English Adolescents. Journal of Imaging 6(5), (2020) 34.

56. D.J.J. Farnell, J. Galloway, A. Zurov, S. Richmond. Multilevel Models of Age-Related Changes in Facial Shape in Adolescents. Communications in Computer and Information Science 1065, (2020) 101–113.

57. D.J.J. Farnell, S. Richmond, J. Galloway, A.I. Zhurov, P. Pirttiniemi, T. Heikkinen, V. Harila, H. Matthews, and P. Claes. Multilevel Principal Components Analysis (mPCA) of Age-Related Changes in Facial Shape in Adolescents. Computer Methods and Programs in Biomedicine 188, 105272 (2020).

58. O.E. de Noord and E.H. Theobald. Multilevel component analysis and multilevel PLS of chemical process data. Journal of Chemometrics: A Journal of the Chemometrics Society, 19, (2005) 301–307.

59. S. De Jong. SIMPLS: an alternative approach to partial least squares regression. Chemometrics and intelligent laboratory systems, 18(3), (1993) 251–263.

60. J.D. White, A. Ortega-Castrillon, H. Matthews, A.A. Zaidi, O. Ekrami, J. Snyders, Y. Fan, T. Penington, S. Van Dongen, M.D. Shriver, and P. Claes, MeshMonk: Open-source large-scale intensive 3D phenotyping. Scientific reports 9, (2019) 6085.

61. S. Richmond, A.I. Zhurov, A.S.M. Ali, P. Pirttiniemi, T. Heikkinen, V. Harila, S. Silinevica, G. Jakobsone, I. Urtane, Exploring the midline soft tissue surface changes from 12 to 15 years of age in three distinct country population cohorts. *European Journal of Orthodontics* 42(5), (2019) 517–524.
62. F.L. Bookstein. Pathologies of Between-Groups Principal Components Analysis in Geometric Morphometrics. *Evolutionary Biology* 46, (2019) 271–302.
63. P. Geladi and B.R. Kowalski, Partial least-squares regression: a tutorial. Analytica Chimica Acta, 185, (1986) 1–17.