Detection of shape outliers, with an application to complete unilateral cleft lip and palate in humans

Stanislav Katina

Department of Anthropology, University of Vienna, Austria, and
Department of Applied Mathematics and Statistics,
Comenius University, Bratislava, Slovakia

1 Introduction

Cleft lip/palate (CLP) is a relatively common birth defect (about 1 per 1000) so disfiguring that nowadays it is almost always corrected surgically as early as possible. The postnatal surgical correction does not, however, result in a normally growing maxilla (upper jaw), but instead, owing to scar tissue and altered mechanical function, one that grows abnormally. It is of interest to determine whether that abnormality is itself homogeneous. In this paper I comment on this question both as an example of geometric morphometrics and as an example of the kind of question that craniofacial surgeons might propose in their role as physical anthropologists, students of the variation of human form.

2 Data

The example involves data from digitally processed lateral X-ray films (under the standard conditions–focus-object distance =370cm, object-film distance=30cm, magnification=8.1%) of 48 boys who have complete unilateral CLP but no other malformation, born between the years $1985-1988$. All were surgically corrected in infancy by the same team of surgeons at the Clinic of Plastic Surgery in Prague, Czech Republic. Primary cheiloplasty performed according to Tennison's method at an average age of 9 month was associated with periosteoplasty (without the nasal septum repositioning) using a $5-7$mm wide and $15-20$mm long periostal flap obtained from the lateral maxillary segment (Velemínská *et al.* 2006). All patients underwent palatoplasty at an average age of 5 years, always by method of pushback with pharyngeal flap surgery. The corrective procedures were implemented in the patients with persisting soft tissue deformations like lip and nose, lip or nose (Velemínská *et al.* 2007). Data are from follow-up at two ages, 10 and 15 years, within the adolescent growth spurt (Fig.1).

3 Methods

22 cephalometric points (landmarks) were represented by their Procrustes shape coordinates, and the distribution of the 96 shapes was examined for outliers via Relative Warp Analysis (*RWA*, Dryden and Mardia 1999) in the affine subspace and the first few eigendimensions of the non-affine subspace of the full Procrustes shape space (Bookstein 1991). Using both subspaces, affine and non-affine, gives us a global view to the homogeneity of the sample. The craniofacial surgeons do not know by definition if possible inhomogeneity is affine or not-affine (local with small or large scale) but they might be interested in both. To separate outliers from inliers

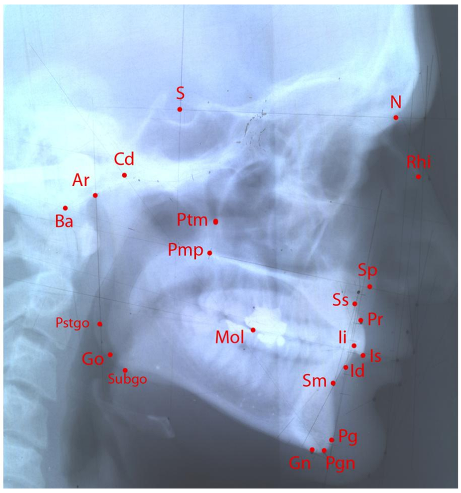

Figure 1: Example of X-ray with 22 landmarks (right lateral view; landmark definitions see in Velemínská *et al.* 2006)

bagplots (*bivariate boxplots*) as supplied by R were used and I will describe them in the next paragraphs.

For generalizing the median to higher-dimensional settings, a variety of different maximum depth estimators have been introduced. They extend, for example, the halfplane location depth of Tukey. Let $X_1, ..., X_N$ be independent and identically distributed bivariate random variables, $X_n \in \mathbb{R}^2$, $n = 1, 2, ..., N$. For given observations $x = (x_1, ..., x_N)$, we write $x_n = (x_{1n}, x_{2n})$, where x_{jn} are RW$_j$ scores (here $j = 1, 2$). The *halfplane location depth* of an arbitrary point $\theta \in \mathbb{R}^2$ relative to x is defined as $d_l(\theta, x) = \min_{\forall H} \# \{n : x_n \in H\}, n = 1, ..., N$, where H ranges over all closed halfplanes of which the boundary line passes through θ. The *deepest location* is defined as $\widehat{\theta}_{dl} = \arg \max\limits_{\forall \theta} d_l(\theta, x)$ and often called the *Tukey median*.

The *depth region of depth* l is defined as the set D_l of points θ_j with $d_l(\theta, x) \geq l$. Equivalently, D_l is the intersection of all closed halfplanes that contains at least $n - l + 1$ observations, hence D_l is a bounded convex polygon, and $D_{l+1} \subset D_l$. The boundary of D_l is a convex polygon, which is called the *contour of depth* l (lth *depth contour*, lth *hull*). Therefore, each vertex of a depth contour is the intersection point of two lines, each passing through two observations.

The univariate boxplot is based on the ranks since the box goes from the observation with the rank $\left\lfloor \frac{n}{4} \right\rfloor$ to that with the rank $\left\lceil \frac{3n}{4} \right\rceil$, and the central bar of the box is drawn at the median. A natural generalization of ranks to multivariate data is the notion of halfspace depth (Tukey 1975). Using this concept in shape analysis, we propose a *bivariate boxplot* (*bagplot*) with the *bag*, that contains 50% of the data points, a *fence* that separates inliers from outliers and a *loop* indicating the points outside the bag but inside the fence (Rousseeuw *et al.* 1999). The loop plays the same role as the two whiskers in one dimension, so we could call this plot also ”*bag-and-bolster plot*” to stress the analogy with ”box-and-whisker plot”. Like the univariate boxplot, bagplot visualizes several characteristics of the data - location, spread, correlation, skewness and tails. The construction of bag B is as follows. Let $\#D_l$ denote the number of points contained in D_l. One first determines the value l for which $\#D_l \leq \left\lfloor \frac{n}{2} \right\rfloor < \#D_{l-1}$ and then interpolates linearly between D_l and D_{l-1}, relative to $\widehat{\theta}_{dl}$, to obtain the set B (the bag B is a convex hull). The fence is obtained by inflating B, relative to $\widehat{\theta}_{dl}$, by a factor 3 found on the base of simulations (Rousseeuw *et al.* 1999). The loop contains all points between the bag and the fence and its outer boundary is the convex hull of the bag and the non-outliers. The points outside the fence are flagged as the outliers.

Above mentioned Tukey depth and depth contoures can be directly applied to *RW* scores x_n to detect possible outliers (Katina 2005).

4 Results and discussion

There are no outliers apparent in the affine (uniform) subspace. In the non-affine subspaces, we find no outliers for bending patterns at large scale but perhaps some outliers for local changes at small scale. These latter are associated with possible '*creases*' (extrema of directional derivative, Bookstein 2000) of the corresponding thin-plate splines (Fig.2). In those cases we can use the same spline formalism to relax the outlying form to an inlier by optimal relaxation along the '*curve décolletage*' that weighs bending energy against Procrustes distance (Fig.3).

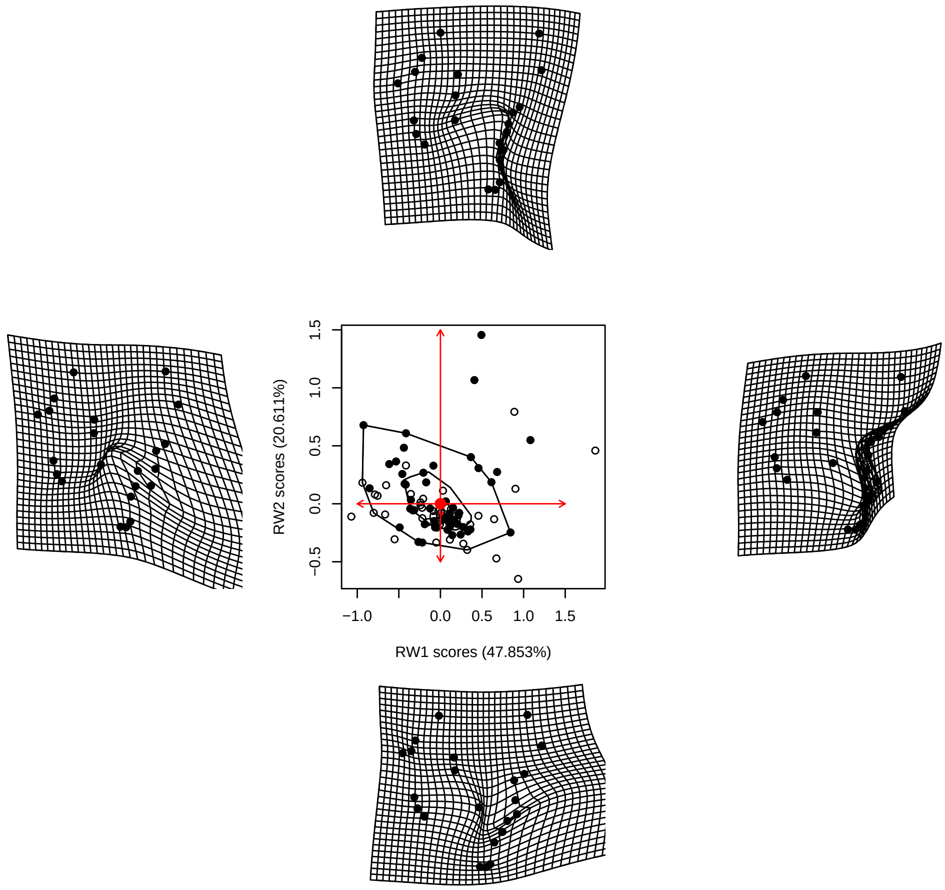

Figure 2: *RW* scores (RW_1 and RW_2) with the bag and the fence of the bagplot (middle) and TPS deformation grids of the mean shape to the estimated shapes indicated by the tips of the arrow heads in the middle figure – in shape space for local changes with small scale ($\alpha = -1$); 10 years old patients $\bullet$, 15 years old patients $\circ$

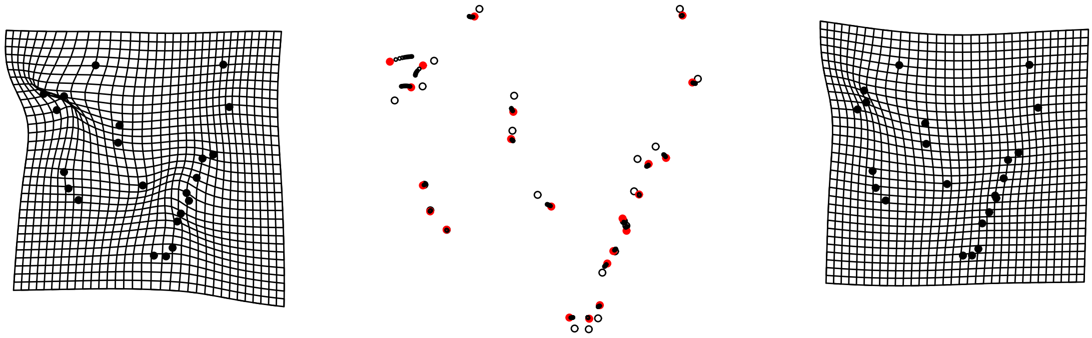

Figure 3: TPS deformation grids and relaxation in Procrustes shape coordinates; TPS grid of Procrustes mean shape to the shape '45' with crease in the upper right corner (left); 'curve décolletage' of shape '45' (mean shape ○, middle), TPS grid of Procrustes mean shape to the relaxed shape '45'

For LASR, these maneuvers suggest a possibly novel and interesting fusion of the Procrustes-spline toolkit with another theme, outlier detection, of the standard modern data-analytic toolkit. They also have practical implications for craniofacial management of CLP follow-up as well as suggestive implications for outlier detection in applied craniometrics and anthropometrics more generally.

References

Bookstein, F.L. (1991). *Morphometric tools for landmark data - geometry and biology*. Cambridge, Cambridge University Press.

Bookstein, F.L. (2000). Creases as local features of deformation grids. *Med. Image Anal.*, **4**, 93-110.

Dryden, I.L., Mardia, K.V. (1999). *Statistical Shape Analysis*. John Willey, New York.

Katina, S., 2005. Comparison of Procrustes and Bookstein 2D coordinates in shape analysis: simulations, bagplots for landmarks and applications. *Forum Statisticum Slovacum* **2**, 47-51.

Rousseeuw, P.J., Ruts, I., Tukey, J.W. (1999). The bagplot: a Bivariate Boxplot. *The American Statistician*, **53**, **4**, 382-387.

Tukey, J.W., 1975: Mathematics and the picturing of data. *Proc. Int. Congress Math.*, **2**, 523-531.

Velemínská, J., Katina, S., Šmahel, Z., Sedláčková, M. (2006). Analysis of facial skeleton shape in patients with complete unilateral cleft lip and palate: Geometric morphometrics. *Acta Chirurgiae Plasticae*, **48**, **1**, 26-32.

Velemínská, J., Šmahel, Z., Katina, S. (2007). Development prediction of sagittal intermaxillary relations in patients with complete unilateral cleft lip and palate during puberty. *Acta Chirurgiae Plasticae*, **49**, **2**, 41-46.

Acknowledgement

This research was supported by $EU\ FP6$ Marie Curie Actions grant MRTN-CT-2005-019564 (EVAN, Human Resource and Mobility Activity) and VEGA grant No. $1/3023/06$ from the Slovak Grant Agency. For comments I thank Fred L. Bookstein and Paul O'Higgins. For data acquisition and pre-processing I thank Jana Velemínská.