

Applications of Procrustes analysis in orthodontic diagnosis and treatment planning

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 21 maart 2019
om 12.30 uur precies

door
Hans Wellens

geboren op 24 september 1973
te Wilrijk

Promotor:

Prof. dr. A.M. Kuijpers-Jagtman

Doctoral Thesis Committee:

Prof. dr. M.A.W. Merkx (voorzitter)

Prof. dr. Y. Ren (RUG)

Prof. dr. G. Willems (KU Leuven, België)

Paranymphs:

Dr. Alejandra Ortega Castrillon

mej. Julie Wellens

Institute for Health Sciences
Radboudumc

PhD thesis Radboud Institute for Health Sciences, Department of Dentistry, section of Orthodontics & Craniofacial Biology, Radboud university medical center, Nijmegen, the Netherlands

The research presented in this thesis was performed at the author's private practice, located at the Groene-Poortdreef 16, 8200 Sint-Michiels, Belgium.

ISBN: 978-94-6380-238-3

Applications of Procrustes analysis in orthodontic diagnosis and treatment planning

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of Deans
to be defended in public on Thursday, March 21st, 2019
at 12.30 hours

by
Hans Wellens

born on September 24th, 1973
in Wilrijk (Belgium)

Supervisor:

Prof. dr. A.M. Kuijpers-Jagtman

Doctoral Thesis Committee:

Prof. dr. M.A.W. Merkx (chair)

Prof. dr. Y. Ren (RUG, Netherlands)

Prof. dr. G. Willems (KU Leuven, Belgium)

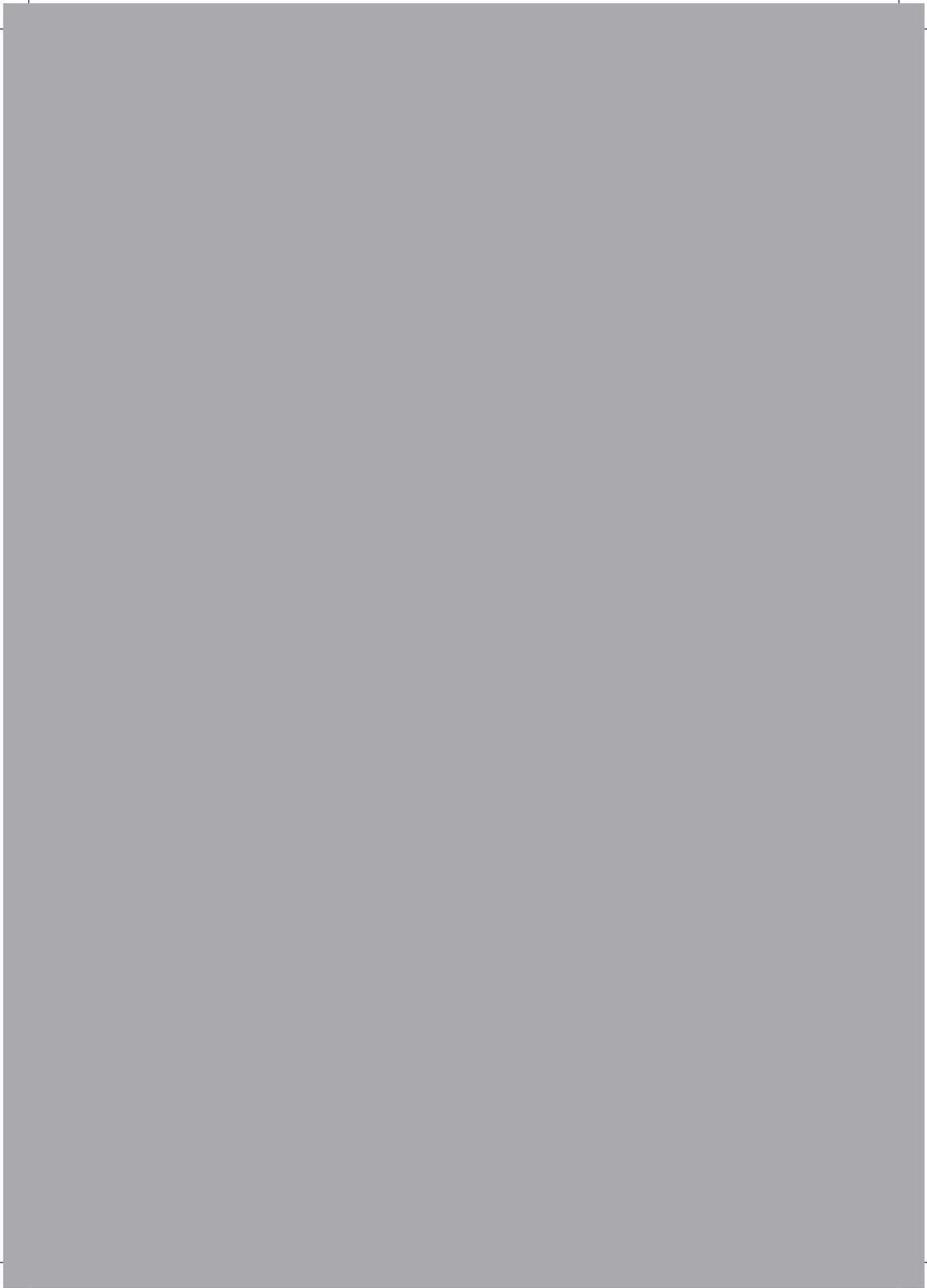
Paranymphs:

Dr. Alejandra Ortega Castrillon

mej. Julie Wellens

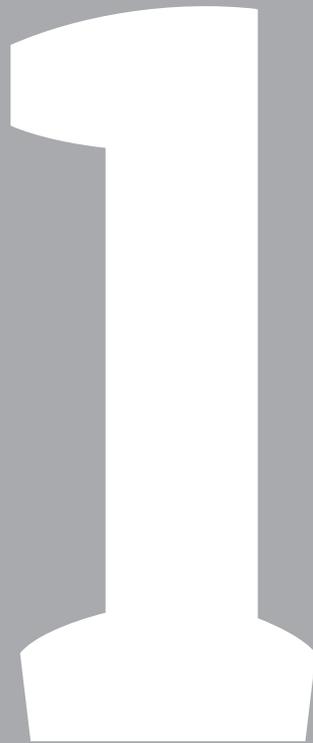
Contents:

Chapter 1	General introduction	9
Chapter 2	Improving the concordance between the ANB and Wits appraisal using Procrustes analysis. <i>Eur J Orthod 2009;31:503-515</i>	41
Chapter 3	Geometric morphometric analysis of craniofacial variation, ontogeny and modularity in a cross-sectional sample of modern humans. <i>J Anat 2013;222:397-409</i>	59
Chapter 4	Connecting the new with the old: modifying the combined application of Procrustes superimposition and principal component analysis, to allow for comparison with traditional lateral cephalometric variables. <i>Eur J Orthod 2016;38:569-576</i>	87
Chapter 5	ROC surface assessment of the ANB angle and Wits appraisal's diagnostic performance with a statistically derived 'gold standard': does normalizing measurements have any merit? <i>Eur J Orthod 2017; 39:358-364</i>	107
Chapter 6	General Discussion	123
Chapter 7	Summary	151
	Samenvatting	156
	Acknowledgements	165
	Curriculum Vitae	169
	List of publications	170
	Awards and honors	171



CHAPTER 1

General introduction



1. Background

The orthodontic diagnostic toolset traditionally comprises both a clinical and a radiological investigation, whereby the latter usually consists of a panoramic radiograph and/or peri-apical series, as well as a lateral (and possibly frontal) cephalogram (1,2). Most textbooks advocate lateral cephalometric analysis as an integral part of orthodontic diagnosis and treatment planning (1,2), while both the American and the European Boards of Orthodontics require these analyses when presenting cases (3,4). When lateral cephalometry was introduced in orthodontics in 1932 (5,6), it was indeed expected to revolutionize the diagnosis of maxillofacial discrepancies, which before represented a largely subjective exercise. It was assumed that the plethora of angular and linear measurements, surface areas and proportions subsequently introduced for analyzing the resulting radiographs would enable orthodontists and maxillofacial surgeons to more objectively, and in greater detail assess craniofacial form. This in turn was expected to translate into improved diagnostic power, with treatment decisions being based upon sound, numerical data.

An increasing volume of research seems to suggest these promises have never been fully met (7–14). Although it is not unexpected that treatment plans seem to be based mainly upon dental records (8,9,15), the addition of a lateral cephalogram to the diagnostic dataset appears to instigate relatively few treatment planning changes (8–12,15). Although the lateral cephalogram was rated somewhat higher than the panoramic radiograph and/or peri-apical series in terms of diagnostic productivity (8), it mainly seemed to influence extraction decisions (12), and overall did not improve diagnostic consistency (8,11). The latter finding seems to align poorly with the perceived importance for the treatment plan reported by orthodontists (8,10). The seemingly low impact of lateral cephalometry on the diagnostic process prompted Nijkamp et al. to advise against its routine use in clinical practice (11); a conclusion shared by Rischen et al. in their meta-analysis investigating the records needed for orthodontic diagnosis and treatment planning (13).

The intriguing incongruence between the perceived importance of lateral cephalometry and its observed impact was confirmed in a recent randomized controlled trial by Durão et al., which failed to find a significant influence on the diagnostic outcome, although the colleagues participating in the investigation did judge the lateral cephalogram highly important to producing a treatment plan. One of the authors' main considerations was that the diagnostic efficiency and therapeutic efficacy of lateral cephalometry needed to improve to justify its routine application in clinical practice. The efficacy of diagnostic imaging was defined by Fryback and Thornbury using a six level hierarchical model (16), the first three of which pertain to the images' technical quality (level one), diagnostic

accuracy (level two), and influence on the practitioner's diagnostic thinking (level three). Because of the aforementioned failure of lateral cephalometry to instigate changes to treatment plans formulated without it, it seems to score low on level three.

2. Diagnostic confusion

One of the earliest publications in the literature discussing how diagnostic confusion impacts lateral cephalometry was presented by Demisch et al. (17) when they measured the ANB angle (18), Wits appraisal (19) and AB-plane angle (20) in four groups of twenty patients, classified occlusally as normal, Class II division 1, Class II division 2 or Class III. The age distribution in these four groups was about the same. For both the ANB angle and Wits appraisal they reported regions of overlap where the same cephalometric value could be found in Class I and Class II, as well as Class I and Class III patients. For the ANB angle, overlap was even found between Class I and Class III patients (Figs. 1 and 2).

A similar finding was reported by Kim and Vietas (21) when they measured the Wits appraisal, ANB and AB-plane angle, and APDI index (antero-posterior dysplasia index) angle in 102 patients with a normal occlusion, and 874 malocclusion patients (of which 214 were Class I, 624 were Class II and 36 Class III). Apart from matching the groups' age distributions, they also accounted for/excluded patient with early loss of deciduous molars or with an obvious posterior lateral shift. All measurements were found to exhibit overlapping values for Class II and Class III patients, although the percentage of the measurement range where the overlap occurred varied, with the highest values reported for the Wits appraisal, and the lowest one for the APDI measurement. When Hurmerinta et al. (22) measured the ANB angle and Wits appraisal in 497 Finnish boys aged four to twenty years, the ANB angle and Wits appraisal derived diagnoses were found to agree in 59 per cent of cases, while they disagreed (i.e. between Class I or II or between Class I or III) in 31 per cent. More troublingly, both analyses contradicted one another in 10 per cent of cases (i.e. patient designated Class II by one and Class III by the other). The occasional disagreement or even contradiction among the different measures of sagittal discrepancy has prompted some authors to advocate applying even a third analysis (23) to accommodate situations where both analyses disagree, obviating the lack of sensitivity and specificity of the various tests.

2. Possible explanations

Many possible explanations have been put forward to explain the origin of 'diagnostic confusion' in lateral cephalometry and its surprisingly low diagnostic impact, which can be categorized loosely as technical and analytical aspects of lateral cephalometry, as well as geometric distortion (24). The **technical aspects**, which would belong to level 1 of the

Fryback and Thornbury hierarchical model of diagnostic efficacy (16), usually pertain to various factors confounding image acquisition and interpretation, such as the diverging nature of the x-ray beam, doubling of bilateral anatomical features, and structural shrouding (25). Whereas the first leads to differential cephalometric enlargement as a function of the distance to the x-ray source, generating two differently sized outlines of bilaterally symmetric structures, the latter two compound the problem by structural superimposition in the resulting two-dimensional representation of three-dimensional morphology.

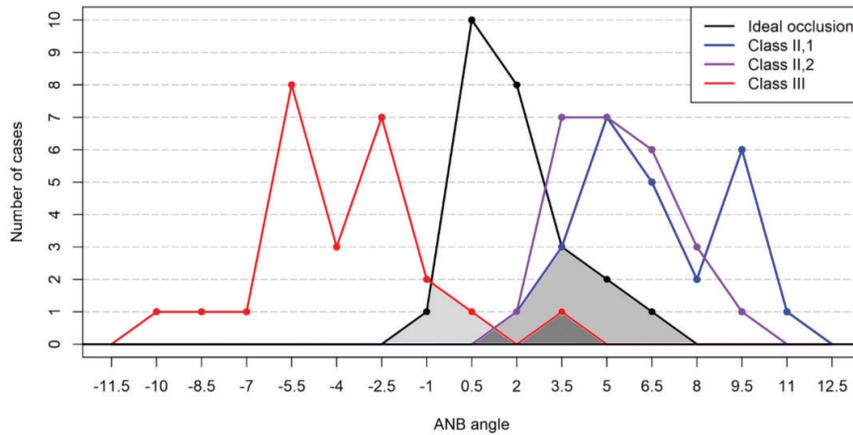


Figure 1. ANB angle values recorded in four groups of n=25 ideal occlusion (in black), Class II division 1 (in blue), Class II, division 2 (in purple) and Class III patients (in red). The regions of diagnostic disagreement (Class I/II or Class I/III overlap) or contradiction (Class II/III overlap) are colored light grey and dark grey, respectively. Recreated from Demisch et al. (17)

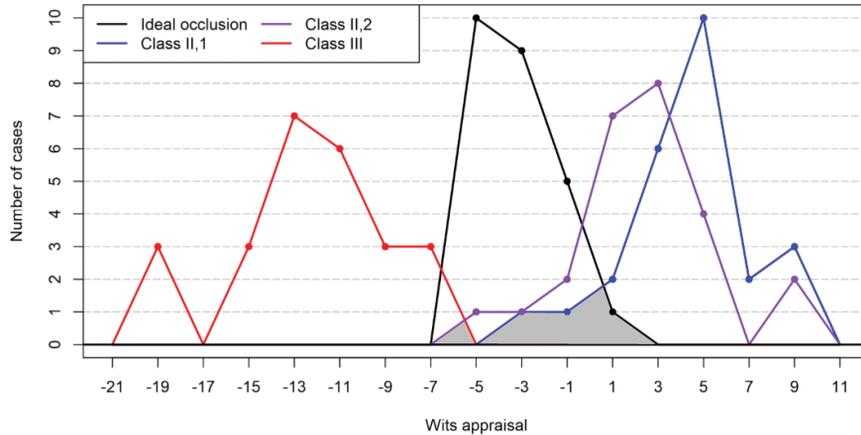


Figure 2. Wits appraisal values recorded in four groups of n=25 ideal occlusion (in black), Class II division 1 (in blue), Class II, division 2 (in purple) and Class III patients (in red). The regions of diagnostic disagreement (Class I/II or Class I/III overlap) are colored light grey. Recreated from Demisch et al. (17).

Difficulties associated with head rotation in the cephalostat have been mentioned as well (26,27). Combined, the aforementioned factors impede accurate and precise landmark identification (25,28), thus lowering diagnostic power. As an example, multiple authors have reported on difficulties associated with identifying the Wits appraisal's occlusal plane landmarks (17,29–31), which might lower the reproducibility of the obtained values (32).

Analytical aspects mainly pertain to, or are associated with the choice and definition of (reference) landmarks, as well as the fixed or floating, age, gender and/or ethnicity dependent value of the associated normative values (12,14). The Hurmerinta study cited above (22) provides one example, illustrating how in brachyfacial growth patterns, the distribution of the ANB angle was skewed in the Class III direction, while the opposite held true for the Wits appraisal; a phenomenon which was confirmed by Tanaka et al. (33). The final diagnosis may therefore be directly dependent on the cephalometric parameter of choice. Moreover, Hurmerinta showed an age-related decreasing mean value for the ANB angle and increasing mean value for the Wits appraisal (24), putting into question the use of the fixed normative cut-off values. Pancherz and Sack indeed described how growth may displace point N anteriorly and vertically (34), influencing the ANB angle; an effect which was confirmed by Lux et al. (35). Intriguingly, despite the aforementioned effect of the low-angle growth pattern on the ANB angle and Wits appraisal (22), Del Santo et al. mainly found inconsistencies between them for high-angle patients (36), largely due to their high occlusal plane angles. Indeed, similar to the aforementioned forward and upward growth of point N (34), the Wits appraisal's occlusal plane has been shown to exhibit a tendency to rotate forward with increasing age (30,37). A more in-depth description of the analytical factors, and the approaches introduced to account for them can be found in Wellens (24).

One confounding factor which often seems to be overlooked is geometric distortion, which might be described as 'the geometric phenomenon allowing patients with seemingly identical mandibulomaxillary relationships to exhibit markedly different cephalometric values', and has frequently been associated with measures of sagittal discrepancy such as the ANB angle and the Wits appraisal. For instance, Taylor pointed out that the ANB angle may be influenced by the anteroposterior position of point N (i.e. anterior skull base length) (38), which Freeman basically rephrased, when he pointed out the effects of relative bimaxillary protrusion or retrusion (39). Similarly, Binder reported on the influence of varying the vertical position of point N (40), while Jacobson explained the possible effects of rotations of the maxillomandibular complex relative to the skull base (41). The role of the vertical dento-alveolar dimension was described by Bishara (29).

Whereas the analytical aspects mentioned above mostly relate to how measurements in individual patients may be influenced by the choice of landmarks or their positional changes through growth and development, the geometric ones illustrate how relative changes in the position of the reference landmarks impact the measurements inter-individually, in patients exhibiting identical intermaxillary relationships. Although the above mentioned factors relate only to the ANB angle, the Wits appraisal certainly is not immune to this effect either. Roth (1982) clearly demonstrated its dependence on the vertical dentoalveolar dimension and on the cant of the occlusal plane. One way of summarizing all aforementioned examples, is to state that geometric distortion encompasses all influences on cephalometric measures originating from inter-individual variation in the position of the reference landmarks and planes used in said measure.

Together, the effects of the aforementioned technical, analytical and geometric factors combined serve to dilute the diagnostic consistency between the various lateral cephalometric analyses. This can be confirmed indirectly from the generally low correlations reported between them (between 0.62 to 0.73) (24,31,36,42–44), which would seem to strengthen the suggestion that the various measurements' results are a representation of factors other than sagittal intermaxillary relationships.

3. The land-surveyor analogy

The central axiom of this doctoral thesis is that the problems facing lateral cephalometry can be more readily understood by considering possible analogies between biological measurement and land surveying. Land surveyors equally make extensive use of angular and linear measurements in order to determine the exact position of (amongst others) property boundaries: a so-called "cadastral survey".

If an orthodontist were to attempt a survey of the property depicted in Fig. 3A, one possible approach might be to position the land surveyor's measuring instrument in the lower left corner of the said property, recording the distance relative to the upper left and right corners, as well as the angles formed between a line joining each vertex with the lower left corner, and another one joining the two lower corners (Fig 3B). The resulting ANB angle-like measurement could then be compared to cadastral data in order to draw conclusions regarding the upper corners' locations. Challenging the conclusions would be trivial, as they are based on the potentially false assumption that the reference landmarks (the lower left and right corners) are located correctly.

Forced to look for an alternative, the colleague might instead choose to construct a diagonal through the corners of the building, deriving the location of the two upper corners from the location of the intersection of perpendiculars drawn from these corners

onto the diagonal (recording all associated distances in the process) (Fig. 3C). In this Wits appraisal-type measurement, the uncertainty regarding the lower two corner landmarks is indeed avoided, although it is clearly substituted with uncertainty regarding the location of the building's corners. Therefore it would seem the very same reservations with respect to the ANB-like measurement apply here as well. Land-surveyors cleverly avoid this measurement conundrum by using an external frame of reference: an intricate network of "benchmarks" (points of which the location has been highly accurately determined in three spatial dimensions) from which to perform all measurements, for which there seems to be no readily available biological alternative.

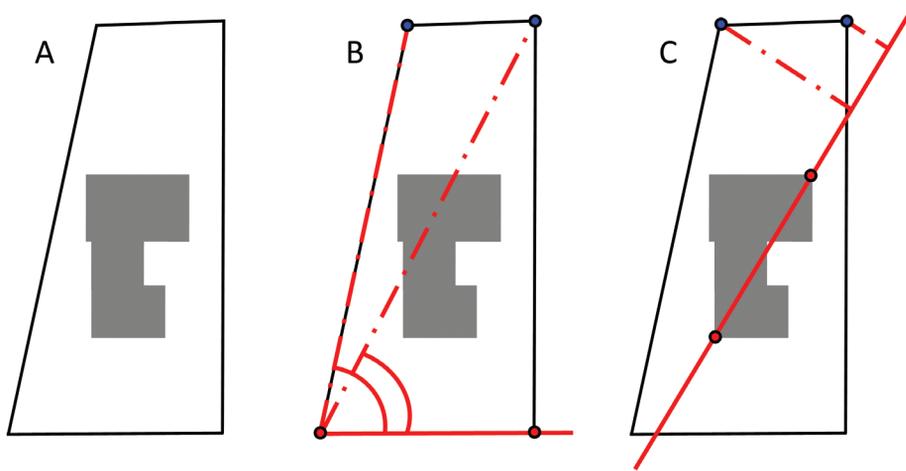


Figure 3. The land-surveyor analogy, illustrating the ANB analogous measurement in the middle panel (B), while the right panel (C) depicts the Wits analogue.

Although it might appear so at first glance, the analogy does not necessarily imply that all lateral cephalometric measurements are inherently flawed. Instead, it provides a logical possible explanation as to why they seem to work most of the time in daily clinical practice: in most individuals, the measurements' reference landmarks are simply located 'averagely enough' to ensure its applicability or validity. Unfortunately, there appears to be no straightforward methodology to establish in which subjects this is the case.

The analogy also suggests the traditional approach to problem solving in cephalometrics, simply moving around the 'cephalometric measuring-tripod' to a different set of reference landmarks to perform linear or angular measurements from, would seem to serve little purpose other than shifting positional uncertainty from one set of reference landmarks to the next.

One intriguing suggestion arising from this analogy is that lateral cephalometry might benefit if the intended angular/linear measurements could somehow be performed from a patient-independent reference frame, instead of utilizing the patients' own reference landmarks and planes for this purpose. This would of course require a suitable reference frame to be constructed, relative to which measurements are to be performed. Furthermore, a mathematically correct methodology would have to be established, for superimposing this reference frame on the patient's landmarks. Finally, it would have to be established whether the proposed methodology has any merit, by comparing its diagnostic performance to the traditional cephalometric measurements.

4. The geometric morphometric toolset

Orthodontics' and oral-maxillofacial surgery's interest in appraising (craniofacial) shape certainly is not unique to these fields. In fact, anthropologists and biologists alike have invested considerable effort in developing methods to describe, quantify and compare craniofacial morphology, giving rise to **geometric morphometrics**: an elaborate statistical toolset designed specifically for the study of **shape** (45–48). Shape is defined as “all geometric information remaining after removing the influences of position, size and orientation” (49). In comparison to traditional craniometry or cephalometry, the usual approach of geometric morphometrics involving generalized Procrustes superimposition (50,51) and principal component analysis (52), is largely operator independent and hence provides a more objective, methodologically superior approach to studying shape. As such, it might provide the required tools to fulfill the first two of the requirements mentioned above. Geometric morphometrics represents a very wide research field, of which only a very concise and therefore limited and incomplete overview will be presented here, with the aim of providing some background information for the following chapters.

David Kendall's above mentioned definition of shape (49) treats location, size and orientation as nuisance factors, which are removed using **generalized Procrustes superimposition (GPS)**: an iterative superimposition algorithm which centers, scales and rotates landmark configurations, minimizing the distances between corresponding landmarks (45,48,50,51,53). The centering operation simply shifts the centroid of all configurations to the origin, while the scaling step rescales them to a standardized size measure. This is usually **centroid size**: the square root of the sum of the squared distances of all the configuration's landmarks to its centroid (48,53), which is set to one. The third transformation involves rotating all configurations such that the sum of the squared distances between the corresponding landmarks is minimized (the statistically ubiquitous 'least squares criterion'). This three-step procedure is repeated multiple times

(i.e. iterated), until the inter-landmark distances can no longer be diminished (50,51). The 'goodness of fit' resulting from the Procrustes superimposition can thereafter be assessed using the **Procrustes distance** (45,48): the square root of the sum of the squared distances between the corresponding landmarks. This value will be smaller for configurations which are more similar in shape.

The number of ways landmark data is free to vary is usually referred to as the **number of degrees of freedom** (DOF). The latter equals the number of landmarks times the number of dimensions (e.g. for 16 landmarks in 2 dimensions, there are 32 DOFs). Procrustes superimposing landmark configurations has a constraining effect on variability, since by standardizing position one DOF is lost for the x, and another one for the y dimension, while scaling to centroid size and rotating to 'best-fit' decreases the total with two more DOFs. Therefore, 4 DOFs are lost as a consequence of the Procrustes superimposition for two-dimensional data. Similarly, seven DOFs are lost for three-dimensional

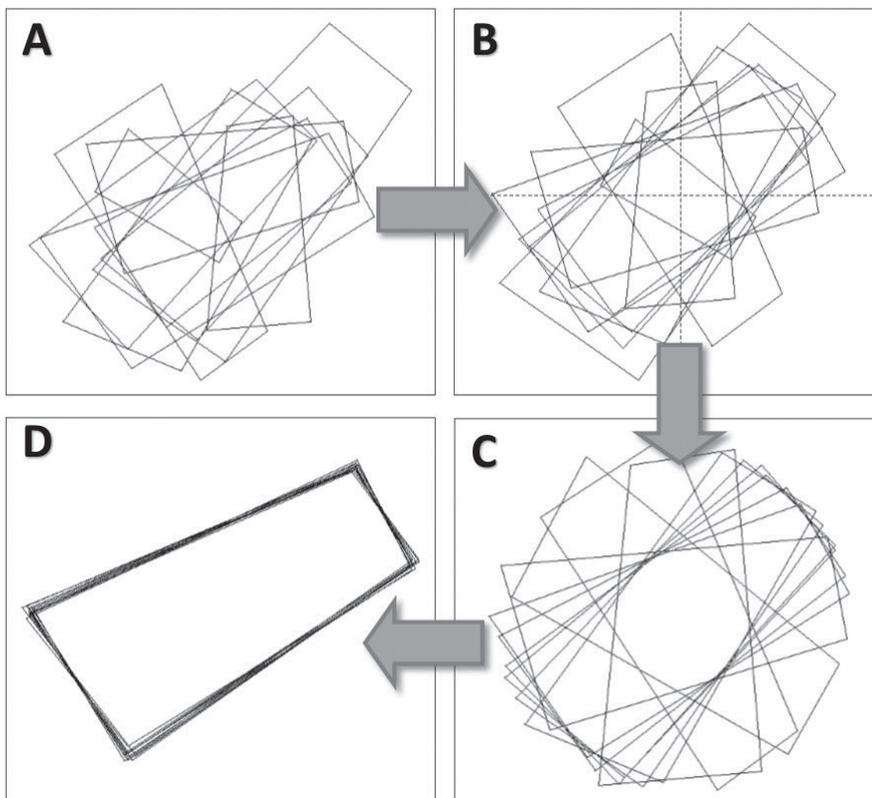


Figure 4. Procrustes algorithm, applied to the collection of rectangles depicted in panel A. The configurations are subsequently centered (panel B), scaled (panel C) and rotated (panel D), minimizing the distance between the corresponding landmarks using the least squares criterion.

configurations (i.e. three DOFs for centering, one for scaling, and three for rotation). The remaining number of degrees of freedom effectively represents the dimensionality of the resulting **curved, higher dimensional shape space on which each Procrustes superimposed landmark configuration is represented by a specific point** (for triangles in two dimensions, this shape space is a sphere) (54). Landmark configurations which are more similar in terms of shape (exhibit a smaller Procrustes distance) are closer together in the resulting shape space. The curved nature of the latter however does not lend itself to classic multi-variate statistics, which requires flat, Euclidian space. Instead, the configurations are projected (either orthogonally or stereographically) onto a planar space, tangent to the Procrustes shape space at the point of the reference configuration, in order to analyze the results statistically (45,48,54).

Since the mathematical foundations of Procrustes superimposition have been thoroughly documented (50,51,54), it represents an interesting candidate methodology for superimposing an external reference frame onto a patient's landmark data, as suggested above. Furthermore, by allowing us to determine the mean shape of different patient groups (e.g. male/female, younger versus older), as well as the significance of the shape difference between them, it might provide the sought after external reference frame as well.

The landmark coordinates' variational patterns can subsequently be scrutinized by subjecting the Procrustes superimposed configurations to **principal component analysis (PCA)** (52,55). PCA provides the directions (in higher dimensional space) along which the superimposed configurations vary most. These directions consist of linear combinations of the original values, which are perpendicular to one another and therefore mutually independent, and are derived by basically rotating the landmark data's coordinate system such that the greatest direction of variation (i.e. the first principal component or PC1) is horizontal (Fig. 5A to D)(48,56).

Figure 5 depicts a fairly typical scatter pattern for biological landmark data, which is usually somewhat elliptically distributed, with a more or less pronounced elongation of the scatter along one axis (Fig 5,A). A confidence ellipse has been added for clarity. The most important direction of variation (i.e. first principal component, or PC1), is represented by the red diagonal in Fig. 5B, which unsurprisingly coincides with the major axis of the ellipse in Fig. 5A. The minor axis of the same ellipse represents the second PC. A third PC can similarly be constructed, perpendicular to the two previous ones, which explains the majority of the variation not explained by PC1 and PC2. Counterintuitively, that process can be continued (abstractly of course) with each subsequent PC oriented

perpendicular to all previous ones explaining the majority of the ever decreasing remaining variation, until no more explainable variation is left.

Performing the PCA as described above would be fairly trivial, albeit not particularly informative. Although individual landmark variation is clearly of interest, their covariation provides much additional morphological information. The PCA is therefore calculated using the landmarks' variance-covariance matrix (45,48,53,56). Performing an Eckhard-Young singular value decomposition of the variance-covariance matrix, yields 3 matrices of which the first (left singular) matrix contains the **PC coefficients** or '**PC loadings**' (corresponding to the cosine of the angle of the new, rotated axis system relative to the original one), whereas the middle matrix contains the eigenvalues of the decomposition on its diagonal (which represent the percentage of variability explained by each PC), and zeros everywhere else. The diagonal matrix is usually standardized to one (i.e. the total variability explained is 100 per cent). The covariance matrix is preferred over the correlation matrix, since using correlations would standardize all the

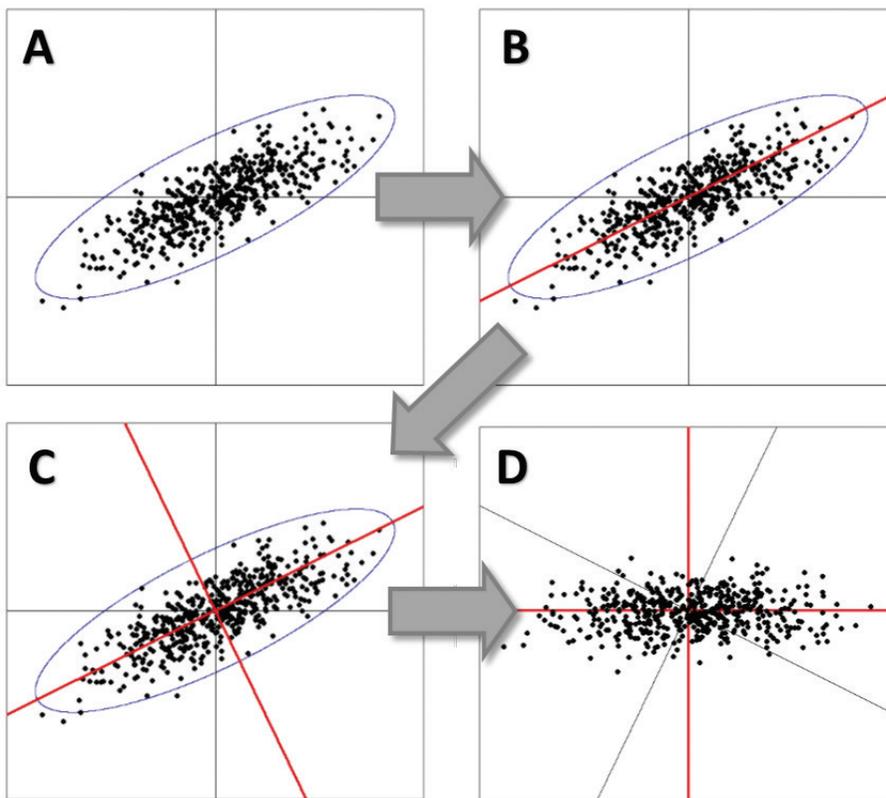


Figure 5. Conceptual representation of coordinate system rotation as applied during principal component analysis (48,56). Adapted from Zelditch et al. (48).

variables' variation to one, effectively eliminating the variation's scale. This would lead to different transformations being applied to each coordinate, since these usually differ in their variability. Finally, the effect of the PCA on the correlation matrix can be also orientation dependent.

Interestingly, the sharp decrease in the amount of variance explained by each subsequent PC implies that most of the shape variability can be described using a relatively small subset of the total number of principal components (48,57). This can be readily observed from a scree plot (Fig. 6), which visualizes the amount of the variability explained by each PC.

The question then arises how many PCs should be retained to provide a sufficiently complete picture of the shape variability. The latter is basically equivalent to the question which of these PCs describe biologically meaningful variability, as opposed to noise and measurement error. Some 'rules of thumb' have been described, such as only taking into consideration those PCs which explain more than five per cent of total variability (i.e. the first 5 PCs in Fig. 6), or to only retain those PCs to the left of the inflection point in the scree plot (48). The latter indicates the point beyond which subsequent PCs describe more or less the same amount of variability, since the decrease in explained variability between them is very small. More rigorous approaches for establishing the above mentioned 'biological interpretability' of a PC, such as the χ^2 -statistic based test proposed by Anderson (58), or permutation-based approaches such as those presented in Peres-Neto et al. (54), are also available.

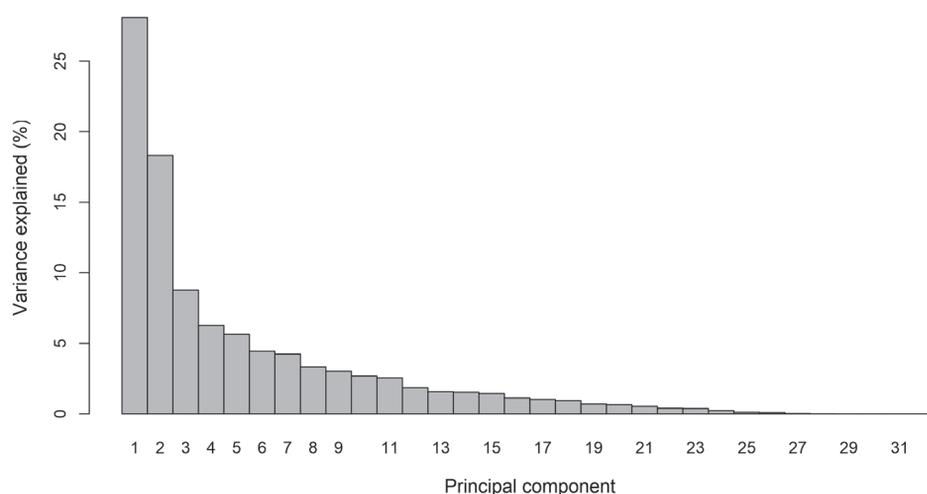


Figure 6. Scree plot depicting the amount of variability explained by each principal component. The sharp decrease in the amount of explained variability is clearly discernible. Consequently, most of the variability is explained by only a few of the first PCs.

Principal component analysis also generates **PC scores**, which indicate every patient's value (i.e. score) on each of the principal components. Principal components can be visualized by calculating the shape corresponding to a position of interest along a particular PC (e.g. plus and minus 3.5 standard deviations along PC1). Also, by plotting the PC scores in 2 or 3 dimensions, a patient's position within the plot may be ascertained relative to all other patients, and the sample mean shape (the origin of the axis system). The distance of a patient's location within the plot to other patients or the sample's mean shape, corresponds to the Procrustes distance between the associated Procrustes superimposed configurations. The PC scores are obtained by post-multiplying the matrix of Procrustes superimposed landmark coordinates with the aforementioned left singular matrix containing the PC coefficients (45,53). The shape corresponding to a position along a PC can be obtained by post-multiplying the matrix of (adjusted) PC scores with the transpose of the matrix containing the PC coefficients (45,53).

A very nice example of the application of geometric morphometrics in cephalometric diagnosis can be found in an article by Halazonetis (59). When applying GPS and PCA to the digitized lateral cephalograms of 150 Greek patients (15 landmarks), the first two principal components were found to explain 48.4 per cent of total variance. PC1 seemed to predominantly describe dolichofacial versus brachyfacial morphology, whereas PC2 mostly represented mandibulo-maxillary retrusion and protrusion. Combined, these two PCs seemed to "align" very favorably with orthodontists' preconceptions regarding the most important directions of craniofacial variability, as articulated by Sassouni and Nanda already in 1964 (60). In summary, geometric morphometrics allows us to superimpose our patients' landmark coordinates objectively, and to determine the mean shape of different patient groups (e.g. male/female, younger versus older) as well as the significance of the shape difference between them, both of which are useful for this thesis. It also allows us to determine and visualize the most important directions of shape variability, and to establish patients' scores on these directions.

5. Diagnostic performance

To ascertain whether the methodology proposed in this thesis has any merit, its diagnostic performance should be compared to that of the traditional lateral cephalometric tests. Diagnostic performance is usually expressed in terms of the sensitivity and specificity of the test under investigation (61,62): the **sensitivity**, or true positive ratio, represents the proportion of positively diseased patients correctly identified as such by the test, whereas **specificity**, or true negative ratio, refers to the proportion of healthy patients correctly identified as such. A high value for the true-positive ratio and a low one for the false-negative ratio (i.e. 1-specificity) are both highly desirable test characteristics. By

plotting one relative to the other for a full range of possible cut-off values, a **Receiver Operating Characteristic (ROC)** curve plot is obtained, whereby the area under the resulting curve serves as a measure its diagnostic power (61–63): the higher the test's diagnostic power, the more it approximates the upper left corner of the graph.

A test characterized by a diagonal ROC curve (from lower left to upper right) has no discriminatory power whatsoever. Values below 60 per cent area under the curve are usually designated 'very poor', whereas those between 60 and 70 per cent are designated as 'poor', between 70 and 80 as 'fair', between 80 and 90 as 'good', and anything above as 'excellent'. The orthodontic literature pertaining to the diagnostic performance of the lateral cephalometric tests is surprisingly scant (14). One of the most fundamental problems pertaining to ROC curve analysis applied to lateral cephalometric tests, is the absence of a true **gold standard**: a 'reference test' which provides the correct answer to the diagnostic question (61–63). In the absence of a true gold standard for cephalometric diagnosis, some authors have chosen to classify their study subjects

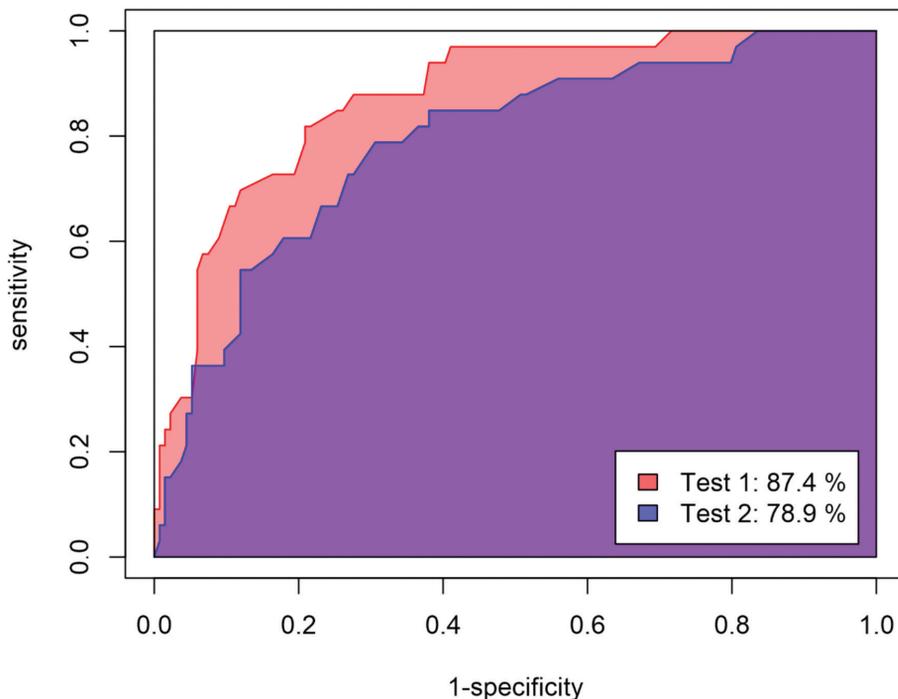


Figure 7. ROC curve plot of two competing diagnostic tests. Test one (in red) more closely approximates the upper left corner of the graph, compared to test two (in blue). Its area under the curve is also larger (87.4 per cent versus 78.9 per cent for test one), indicating it has greater diagnostic power, compared to the latter. ROC curve analysis also allows establishing the optimal cut-off value for the test, which is the value associated with the point on the curve closest to the upper left corner of the graph (62).

based upon occlusion (64,65) or existing cephalometric analyses, applied either singly (66) or combined (67). Other studies have included profile assessments (68,69), while one study applied a Delphi approach to establish their gold standard (69).

Another problem complicating the straightforward application of ROC curve analysis in lateral cephalometrics is the absence of a **true disease state**: since craniofacial variation does not represent a pathological condition, the cut-off values applied to classify individuals accordingly will inevitably be determined arbitrarily. This in turn implies that any gold standard cephalometric cut-off points used to determine the 'correct diagnosis', might in fact be sub-optimal, skewing the ROC curve analysis.

Since the geometric morphometric framework allows us to classify patients in terms of their skeletal makeup both sagittally (relative mandibulo-maxillary retrusion or protrusion) and vertically (dolichofacial versus brachyfacial growth pattern) (59), it might provide the gold standard for cephalometric diagnosis in terms of patterns of craniofacial variation, solving the above mentioned most fundamental problem. Also, there have been some recent developments in the field of ROC curve analysis to accommodate so called multi-class problems (as opposed to the classic two-class problem) (70–73). These thus apply in situations where multiple cut-off points are applicable not just to the diagnostic test, but also to the gold standard: instead of evaluating the diagnostic test's power at all possible values of its cut-off point relative to one gold standard value, the gold standard's cut-off point is varied as well, over different ROC curves. This produces multiple ROC curves, each representing the diagnostic performance of the test at a slightly different value of the gold standard cut-off. By stacking the resulting ROC curves side-by-side, a so called '**ROC surface**' may be constructed, the volume of which is a direct representation of the diagnostic power of the test, when considering multiple possible cut-off points for the gold standard (70,71).

6. Research questions

This thesis focused on the following research questions:

1. In light of the land-surveyor analogy, is there any merit in exchanging the patient-specific reference frame in lateral cephalometry (i.e. the patient's points N and occlusal plane for the ANB angle and Wits appraisal, respectively) with a Procrustes superimposed one (point N and the occlusal plane of the Bolton 12-year male-female averaged template)? (Chapter 2)
2. The choice of the Bolton 12-year male-female template as a reference frame in Chapter one is quite arbitrary. Could we develop a suitable reference frame applicable to Belgian patients? Can we describe/quantify/illustrate their craniofacial variation? Are the mean shapes of males and females comparable? Do we need a different reference frame for adults and children? (Chapter 3)
3. How does the newly created reference frame relate to our traditional lateral cephalometric variables? Can we visualize how the traditional variables behave within the geometric morphometric framework? (Chapter 4)
4. Given the geometric morphometrically derived reference frame, can we draw any definitive conclusions as to the diagnostic performance of the methodology presented in chapter 2? How do the traditional cephalometric tests compare to this proposed methodology? (Chapter 5)

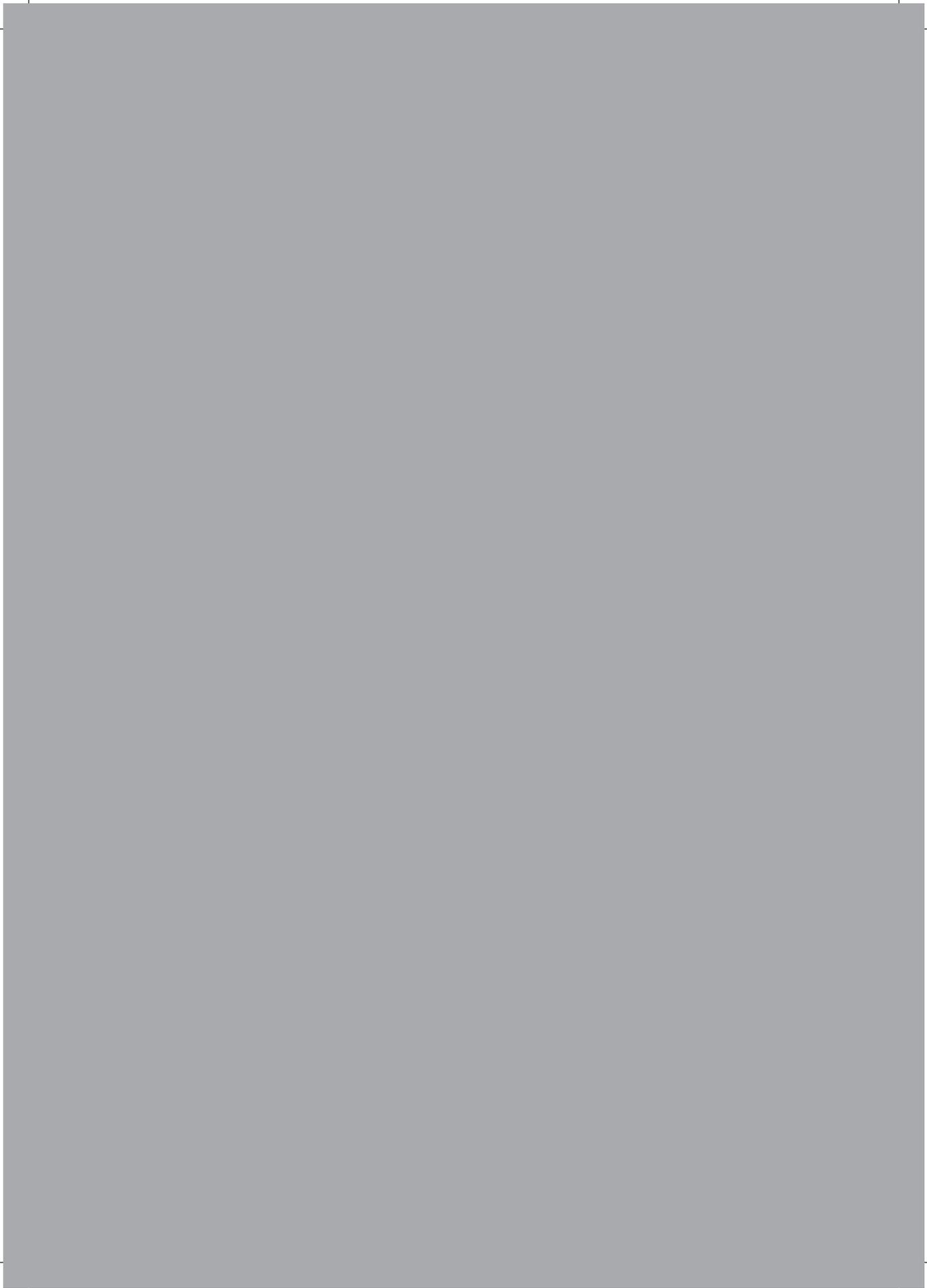
References

1. Graber T, Vanarsdall R, Vig K, Editors. *Orthodontics: Current Principles and Techniques*, 4th edition. St. Louis: Mosby; 2005.
2. Proffit WR, Jr HWF, Sarver DM. *Contemporary Orthodontics*. Elsevier Health Sciences; 2006.
3. Cephalometric Tracing Requirements [Internet]. [cited 20th januari 2018]. Available at: <https://www.americanboardortho.com/orthodontic-professionals/about-board-certification/clinical-examination/case-record-preparation/cephalometric-tracing-requirements/>
4. EBO Book 2013 4th Version [Internet]. [cited October 14th 2014]. Available at: <http://www.eoseurope.org/ebo/EBOBOOK20134thVERSION-web.pdf>
5. Broadbent BH. A NEW X-RAY TECHNIQUE and ITS APPLICATION TO ORTHODONTIA. *Angle Orthod*. 1931;1(2):45–66.
6. Hofrath H. Die Bedeutung der Röntgenfern- und Abstandsaufnahme für die Diagnostik der Kieferanomalien. *Fortschritte Orthod*. 1931;1:34–41.
7. Moyers RE, Bookstein FL. The inappropriateness of conventional cephalometrics. *Am J Orthod*. 1979;75(6):599–617.
8. Atchison KA, Luke LS, White SC. Contribution of pretreatment radiographs to orthodontists' decision making. *Oral Surg Oral Med Oral Pathol*. 1991;71(2):238–45.
9. Han UK, Vig KWL, Weintraub JA, Vig PS, Kowalski CJ. Consistency of orthodontic treatment decisions relative to diagnostic records. *Am J Orthod Dentofacial Orthop*. 1991;100(3):212–9.
10. Hansen K, Bondemark L. The influence of lateral head radiographs in orthodontic diagnosis and treatment planning. *Eur J Orthod*. 2001;23:452–453.
11. Nijkamp PG, Habets LLMH, Aartman IHA, Zentner A. The influence of cephalometrics on orthodontic treatment planning. *Eur J Orthod*. 2008;30(6):630–5.
12. Devereux L, Moles D, Cunningham SJ, McKnight M. How important are lateral cephalometric radiographs in orthodontic treatment planning? *Am J Orthod Dentofacial Orthop*. 2011;139(2):e175–81.
13. Rischen RJ, Breuning KH, Bronkhorst EM, Kuijpers-Jagtman AM. Records Needed for Orthodontic Diagnosis and Treatment Planning: A Systematic Review. *PLoS ONE*. 2013;8(11):e74186.
14. Durão AR, Alqerban A, Ferreira AP, Jacobs R. Influence of lateral cephalometric radiography in orthodontic diagnosis and treatment planning. *Angle Orthod*. 2015;85(2):206–10.
15. Bruks A, Enberg K, Nordqvist I, Hansson AS, Jansson L, Svenson B. Radiographic examinations as an aid to orthodontic diagnosis and treatment planning. *Swed Dent J*. 1999;23(2–3):77–85.
16. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Mak Int J Soc Med Decis Mak*. 1991;11(2):88–94.
17. Demisch A, Gebauer U, Zila W. Comparison of three cephalometric measurements of sagittal jaw relationship: angle ANB, Wits appraisal and AB-occlusal angle. *Trans Eur Orthod Soc*. 1977;269–81.

18. Steiner CC. Cephalometrics for you and me. *Am J Orthod.* 1953;39:729–55.
19. Jacobson A. The “Wits” appraisal of jaw disharmony. *Am J Orthod.* 1975;67(2):125–38.
20. Downs W. Variations in Facial Relationships - Their Significance in Treatment and Prognosis. *Am J Orthod Dentofacial Orthop.* 1948;34(10):812–40.
21. Kim Y. Anteroposterior Dysplasia Indicator - Adjunct to Cephalometric Differential-Diagnosis. *Am J Orthod Dentofacial Orthop.* 1978;73(6):619–33.
22. Hurmerinta K, Rahkamo A, Haavikko K. Comparison between cephalometric classification methods for sagittal jaw relationships. *Eur J Oral Sci.* 1997;105(3):221–7.
23. Ishikawa H, Nakamura S, Iwasaki H, Kitazawa S. Seven parameters describing anteroposterior jaw relationships: Postpubertal prediction accuracy and interchangeability. *Am J Orthod Dentofacial Orthop.* 2000;117(6):0714–20.
24. Wellens H. Improving the concordance between various anteroposterior cephalometric measurements using Procrustes analysis. *Eur J Orthod.* 2009;31(5):503–15.
25. Baumrind S, Frantz RC. The reliability of head film measurements. 1. Landmark identification. *Am J Orthod.* augustus 1971;60(2):111–27.
26. Yoon Y-J, Kim K-S, Hwang M-S, Kim H-J, Choi E-H, Kim K-W. Effect of head rotation on lateral cephalometric radiographs. *Angle Orthod.* 2001;71(5):396–403.
27. Malkoc S, Sari Z, Usumez S, Koyuturk AE. The effect of head rotation on cephalometric radiographs. *Eur J Orthod.* 2005;27(3):315–21.
28. Baumrind S, Frantz RC. The reliability of head film measurements. 2. Conventional angular and linear measures. *Am J Orthod.* 1971;60(5):505–17.
29. Bishara SE, Fahl JA, Peterson LC. Longitudinal changes in the ANB angle and Wits appraisal: clinical implications. *Am J Orthod.* 1983;84(2):133–9.
30. Sherman SL, Woods M, Nanda RS, Currier GF. The longitudinal effects of growth on the Wits appraisal. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod.* 1988;93(5):429–36.
31. Millett D, Gravely JF. The assessment of antero-posterior dental base relationships. *Br J Orthod.* 1991;18(4):285–97.
32. Rushton R, Cohen AM, Linney AD. The relationship and reproducibility of angle ANB and the Wits appraisal. *Br J Orthod.* 18(3):225–31.
33. Tanaka JLO, Ono E, Filho Medici E, Cesar de Moraes L, Cezar de Melo Castilho J, Leonelli de Moraes ME. Influence of the facial pattern on ANB, AF-BF, and Wits appraisal. *World J Orthod.* 2006;7(4):369–75.
34. Panchez H, Sack B. [A critical analysis of the SNA, SNB and ANB angles in assessing orthodontic treatments]. *Fortschr Kieferorthop.* 1990;51(5):309–17.
35. Lux CJ, Burden D, Conradt C, Komposch G. Age-related changes in sagittal relationship between the maxilla and mandible. *Eur J Orthod.* 2005;27(6):568–78.
36. Del Santo M. Influence of occlusal plane inclination on ANB and Wits assessments of anteroposterior jaw relationships. *Am J Orthod Dentofacial Orthop.* 2006;129(5):641–8.

37. Chang HP, Kinoshita Z, Kawamoto T. A study of the growth changes in facial configuration. *Eur J Orthod.* 1993;15(6):493–501.
38. Taylor CM. Changes in the relationship of nasion, point A, and point B and the effect upon ANB. *Am J Orthod.* 1969;56(2):143–63.
39. Freeman RS. Adjusting ANB angles to reflect the effect of maxillary position. *Angle Orthod.* 1981;51(2):162–171.
40. Binder RE. The geometry of cephalometrics. *J Clin Orthod JCO.* 1979;13(4):258–63.
41. Jacobson A. Update on the Wits Appraisal. *Angle Orthod.* 1988;58(3):205–19.
42. Rotberg S, Fried N, Kane J, Shapiro E. Predicting the “Wits” appraisal from the ANB angle. *Am J Orthod.* 1980;77(6):636–42.
43. Järvinen S. A comparison of two angular and two linear measurements used to establish sagittal apical base relationship. *Eur J Orthod.* 1981;3(2):131–4.
44. Richardson M. Measurement of dental base relationship. *Eur J Orthod.* 1982;4(4):251–6.
45. Dryden IL, Mardia KV. *Statistical Shape Analysis.* Wiley; 1998.
46. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: Ten years of progress following the ‘revolution’. *Ital J Zool.* 2004;71(1):5–16.
47. Mitteroecker P, Gunz P. Advances in Geometric Morphometrics. *Evol Biol.* 2009;36(2):235–47.
48. Zelditch ML, Swiderski DL, Sheets HD. *Geometric Morphometrics for Biologists: A Primer.* Academic Press; 2012.
49. Kendall DG. A Survey of the Statistical Theory of Shape. *Stat Sci.* 1989;4(2):87–99.
50. Gower JC. Generalized procrustes analysis. *Psychometrika.* 1975;40(1):33–51.
51. Rohlf FJ, Slice D. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Syst Zool.* 1990;39(1):40–59.
52. Pearson K. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philos Mag.* 1901;2:559–72.
53. Claude J. *Morphometrics with R.* 1 edition. Springer New York; 2008.
54. Rohlf FJ. Shape statistics: Procrustes superimpositions and tangent spaces. *J Classif.* 1999;16(2):197–223.
55. *Principal Component Analysis | I.T. Jolliffe | Springer [Internet].* [cited 28th of januari 2018]. Available at: <http://www.springer.com/us/book/9780387954424>
56. Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE. *Advances in Morphometrics.* Springer Science & Business Media; 2013.
57. Peres-Neto PR, Jackson DA, Somers KM. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Comput Stat Data Anal.* 2005;49(4):974–997.
58. Anderson TW. *An Introduction to Multivariate Statistical Analysis.* Wiley; 2003.
59. Halazonetis DJ. Morphometrics for cephalometric diagnosis. *Am J Orthod Dentofacial Orthop.* 2004;125(5):571–581.
60. Sassouni V, Nanda S. Analysis of dentofacial vertical proportions. *Am J Orthod.* 1964;50(11):801–823.

61. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med.* 1978;8(4):283–98.
62. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol.* 1979;14(2):109–21.
63. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29–36.
64. Han UK, Kim YH. Determination of Class II and Class III skeletal patterns: receiver operating characteristic (ROC) analysis on various cephalometric measurements. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod.* 1998;113(5):538–45.
65. Freudenthaler JW, Celar AG, Schneider B. Overbite depth and anteroposterior dysplasia indicators: the relationship between occlusal and skeletal patterns using the receiver operating characteristic (ROC) analysis. *Eur J Orthod.* 2000;22(1):75–83.
66. Kumar S, Valiathan A, Gautam P, Chakravarthy K, Jayaswal P. An evaluation of the Pi analysis in the assessment of anteroposterior jaw relationship. *J Orthod.* 2012;39(4):262–9.
67. Neela PK, Mascarenhas R, Husain A. A new sagittal dysplasia indicator: the YEN angle. *World J Orthod.* 2009;10(2):147–51.
68. Baik CY, Ververidou M. A new approach of assessing sagittal discrepancies: the Beta angle. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod.* 2004;126(1):100–5.
69. Anderson GW, Anderson GW, Fields HW, Beck FM, Chacon G, Vig KWL. Development of cephalometric norms using a unified facial and dental approach. *Am J Orthod Dentofacial Orthop.* 2006;129(5):710.
70. Mossman D. Three-way ROCs. *Med Decis Mak Int J Soc Med Decis Mak.* 1999;19(1):78–89.
71. Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. *Med Decis Mak Int J Soc Med Decis Mak.* 2000;20(3):323–31.
72. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurements. *Stat Med.* 2004;23(22):3437–49.
73. Li J, Fine JP. ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics.* 2008;9(3):566–76.
74. Smith BR, Park JH, Cederberg RA. An evaluation of cone-beam computed tomography use in postgraduate orthodontic programs in the United States and Canada. *J Dent Educ.* 2011;75(1):98–106.



CHAPTER 2

Improving the concordance between the ANB and Wits measures of sagittal discrepancy using Procrustes analysis

*This chapter is based on:
Wellens HL
Eur J Orthod, 2009;31(5):503-15*



Abstract

Objective: The aim of this study was to investigate a method which minimizes the effects of geometric distortion on various cephalometric measurements used to determine sagittal discrepancy, such as ANB angle, Wits appraisal, AB plane angle, projections on the palatal plane, Frankfort horizontal (FH) plane, the mandibulomaxillary bisector, and the SN line, in an attempt to optimize the correlation between them.

Methods: This was accomplished by superimposing the Bolton 12-year male-female averaged template on a patient's tracing using Procrustes analysis and performing measurements while exchanging the patient's reference landmarks/planes (point N, the mandibulomaxillary bisector, FH plane, occlusal plane, palatal plane, and SN line) with those of the template. The normalized measurements were then compared with their classic counterparts using correlation coefficients. The above cephalometric analyses, classic and normalized, were applied to 71 patients [26 males: mean age 13.1 years, standard deviation (SD) 1.1 years and 45 females: mean age of 14.6 years, SD 8.2 years].

Results: Spearman's rank correlation coefficient was calculated between the classic measurements and their normalized counterparts, resulting in a consistent increase in the correlation between the normalized measurements in comparison with the classic ones. This increase varied in absolute value from 0.052 to 0.405. All normalized measurements were highly correlated ($P > 0.742$, absolute value).

Conclusion: Although correlation calculations do not represent a true measure of diagnostic performance, it is hoped that improving their correspondence heightens the possibility of the different tests agreeing on the patient's sagittal discrepancy, decreasing the possibility of differing, or even totally opposing diagnostic outcomes resulting from their application to (clear-cut) Class I, II, and III patients.

Introduction

When Broadbent (1931) and Hofrath (1931) introduced standardized cephalometric radiography, it was hoped that the new diagnostic tool would provide a straightforward way of substantiating clinical dentofacial observations in orthodontic research as well as in everyday practice. In fact, today, orthodontic research without cephalometry seems almost unthinkable, providing a relatively precise and repeatable way of measuring and comparing growth, development, or treatment change (Baumrind and Frantz, 1971a, b). In daily clinical practice, however, the benefits of cephalometrics are somewhat less clear. Notwithstanding the fact that cephalometry is perceived to be an integral part of treatment planning (Atchison, 1986), several authors have demonstrated a disappointingly low 'weight' of lateral cephalometry in the therapeutic decision-making process: not only do treatment decisions seem to be made based mainly upon dental records (Atchison et al., 1991; Han et al., 1991) but also the addition of lateral cephalograms has been shown to cause few changes to treatment planning (Atchison et al., 1991; Han et al., 1991; Hansen and Bondemark, 2001). Furthermore, most of the orthodontists' certainty regarding his/her treatment plan seems to originate from dental records, the cephalometric data making only a small contribution (Atchison et al., 1991).

ANB and Wits analyses are popular for assessing dentoalveolar sagittal relationships. Riedel (1952) and Steiner (1953) recommended the ANB angle, 'triangulating' the relative sagittal positions of the mandible and maxilla using SNA and SNB. Jacobson (1975) suggested the Wits analysis, constructed by dropping perpendiculars from points A and B onto the occlusal plane, the distance between the resulting points A_o and B_o representing a more or less 'direct', linear measurement. Although both analyses are intended to measure anteroposterior (AP) dental base relationships, reported correlations between them are generally fairly low, values ranging from 0.62 to 0.76 have been mentioned (Rotberg et al., 1980; Järvinen, 1981; Richardson, 1982; Millett and Gravely, 1991; Del Santo, 2006). The apparent lack of 'absolute' correspondence between them suggests that both analyses do not only measure sagittal discrepancy, under the influence of various (geometric) confounding factors.

This became evident when Demisch et al. (1977) applied both cephalometric techniques to groups of fairly clear-cut Class I, II, and III patients: considerable areas of 'overlap' appeared where the same ANB or Wits value could belong to Class I as well as Class II patients or to Class I as well as Class III patients. For ANB analysis, overlapping was found between the values for Class II and III patients, leading Jacobson (1975) to recommend using the Wits analysis to confirm the possibly (more) compromised results of the ANB analysis. However, after analysing 872 malocclusion patients, Kim and Vietas (1978) also

found an overlapping of the Wits measurements for Class II and III patients, containing 36.4 per cent of the malocclusion group. Since the Wits appraisal evidently has its own limitations, a point could be made for applying even a third analysis (Ishikawa et al., 2000) to accommodate situations where both analyses disagree, obviating the lack of sensitivity and specificity of both tests. In view of the popularity of ANB and Wits analysis, it would therefore seem fair to say that even today, the orthodontic speciality is struggling to efficiently 'distil' the required diagnostic or scientific information from cephalometric images.

The aims for the current research were therefore as follows:

1. To first provide an overview of the factors responsible for the lack of correspondence between ANB and Wits measurements of sagittal discrepancy and of previously published attempts to increase this correspondence.
2. To investigate whether the correlation between ANB and Wits measurements of sagittal discrepancy could be improved by normalizing the patient's reference landmarks (S, N, and the occlusal plane). This would be performed by applying Procrustes analysis (Halazonetis, 2004) to the 12-year male – female averaged template of the Bolton-Brush growth study (Broadbent et al., 1975), calculating the normalized ANB and Wits measurements using the normalized reference landmarks and the patient's points A and B. This would allow determination of the correlation between the normalized ANB and Wits measurements and comparison of the resulting values to the correlation coefficients between the classical counterparts.
3. To evaluate whether or not the proposed technique performs similarly when applied to other methods for determining sagittal discrepancy (Figure 1): projections on the mandibulo-maxillary bisector (Hall-Scott, 1994), palatal plane (Ferrazzini, 1976), Frankfort horizontal (FH) plane (Chang, 1987), the SN line (Taylor, 1969), as well as the AB plane angle (AB to N – Pog, Downs, 1948). This would again be accomplished by comparing the correlations between the classic projections with those between the normalized counterparts.

Literature review

The apparent lack of correlation between ANB and Wits appraisal might, at least in part, originate from the radiographic technique itself (Broadbent et al., 1975). The combination of a diverging X-ray beam with slight errors in the positioning of the patient's head results in an enlarged, distorted radiographic image. Structures located on the principal axis of the X-ray beam tend to blur, while those located outside the patient's midsagittal plane are doubled, rendering two separate, more or less superimposed images. Furthermore, the two-dimensional (2D) superimposition of three-dimensional (3D) structures often shrouds the structures of interest, rendering reliable landmark

identification difficult. While the recent introduction of 3D computed tomography (CT) suitable for use in the orthodontic office will solve some of these issues associated with 2D lateral cephalometric radiography, there are some important problems associated with conventional cephalometry for which 3D-CT does not provide a solution. One is the difficulty associated with finding appropriate ways of relating the various structures to each other. For instance, Taylor (1969) pointed out that ANB is influenced by the relative AP position of point N (cranial base length), enabling patients with the same maxillomandibular relationship to have different ANB angles. The same point was made by Freeman (1981), this time by varying the AP position of the jaws in relation to point N (mandibular-maxillary prognathism). Likewise, vertical changes in the position of point N influence ANB, even in the absence of changes in the sagittal jaw relationships: an upward movement of point N will decrease ANB while a downward movement will lead to an increase in this angle (Binder, 1979).

The aforementioned AP and vertical changes in point N are not only important for interindividual comparison of ANB values in patients with comparable maxillomandibular relationships but also for the intraindividual comparison of radiographs taken at different points in time, as growth may displace point N anteriorly and vertically, influencing ANB (Panchez and Sack, 1990). In fact, Bishara et al. (1983) demonstrated significant changes in ANB between 5 and 25 years of age, without a similar change in Wits appraisal. The same conclusions were reached by Lux et al. (2005) between the ages of 7 (ANB) to 9 (Wits appraisal) and 15 years, at least for their Class I and 'ideal occlusion' groups.

Furthermore, Jacobson (1976 , 1988) made reference to the rotational effect of the jaws, whereby a clockwise or counterclockwise rotation of the jaws in relation to cranial reference structures, such as the SN line (and without changing their mutual relationship), will tend to increase/decrease ANB. Although these effects might appear theoretical, Tanaka et al. (2006), in a recent clinical study, clearly demonstrated the influence of facial type on the magnitude of ANB, reporting a lower mean ANB in brachyfacial patients versus higher mean values in dolichofacial patients. Another factor that may influence ANB is the vertical dentoalveolar dimension. Hussels and Nanda (1984) explained that an increase in the vertical dimension in the form of an increase in the length of NB, or of the distance 'point A-occlusal plane', will result in a decrease in ANB. This is, of course, in analogy to the findings of Binder (1979).

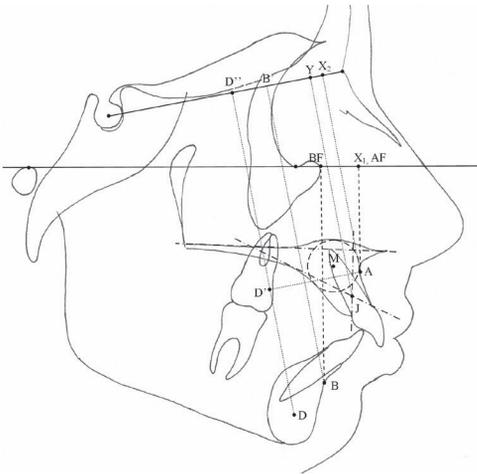


Figure 1. Illustration of the points, angles and planes mentioned in the text. D: midpoint of the cross-section of the mandibular symphysis; X: constructed by dropping a perpendicular from point A on the Frankfort horizontal (FH) plane (X1). Alternatively, a perpendicular may be dropped from point A on the SN line (X2); D': constructed by dropping a perpendicular from point A on the DD'' line, where DD'' is constructed by dropping a perpendicular from point B on the SN line; J: tangents are drawn to the nasal floor through ANS and to the hard palate through prosthion. Point J is located where the latter tangent crosses a perpendicular to the first tangent, which is dropped from the palatal analogue of prosthion (palatal margin of the alveolar bone); Y: constructed by dropping a perpendicular from point J on the SN line; M: midpoint of a circle that best fits the anterior, superior, and palatal outlines of the premaxilla; AF, BF: constructed by dropping a perpendicular from points A and B, respectively, on FH.

Jacobson (1976, 1988) proposed the Wits analysis, in an attempt to circumvent some of the problems associated with the use of cranial landmarks, as in ANB. By projecting points A and B onto the occlusal plane, points S and N were no longer needed. It was further hypothesized that this would solve problems associated with rotations of the mandibulomaxillary complex relative to the cranial base, as the occlusal plane would rotate together with the jaws. Unfortunately, the Wits appraisal also proved not to be immune to geometric influences. Roth (1982) clearly demonstrated its dependence on vertical dentoalveolar dimensions and on the cant of the occlusal plane. The combination of these two factors results in a more significant influence of occlusal plane changes on the Wits value in high- compared with low-angle patients (Del Santo, 2006).

Furthermore, a number of studies have reported that the occlusal plane tends to tip upward anteriorly with increasing age, influencing the Wits appraisal (Sherman et al., 1988; Hall-Scott, 1994). Several authors have also commented on difficulties associated with precisely locating the necessary landmarks (Demisch et al., 1977; Bishara et al., 1983; Sherman et al., 1988; Millett and Gravely, 1991). Originally defined as situated in the 'region of maximal cuspal overlap', occlusal plane landmarks are often somewhat difficult to discern, can be influenced by the stage of dental eruption, and are generally close together, promoting the occurrence of errors. As reproducibility is equally important for the clinical usefulness of a cephalometric technique, Rushton et al. (1991) advised not to ignore these imprecisions: the reproducibility of the Wits value was found to be rather poor, mainly due to difficulties in accurately locating the occlusal plane. Thoroughly defining the various landmarks (Baumrind and Frantz, 1971b) or using specific constructions to determine their position (Demisch et al., 1977) can, at least

in part, solve some of these problems. Many different solutions have been formulated in order to deal with the above geometric problems.

Using different reference points

In an attempt to overcome some of the anatomic and geometric problems associated with ANB and Wits measurements, some authors have suggested using different anatomical points (Figure 1), such as point D as defined by Steiner (1953), point X (in AXB angle; Freeman, 1981), points X and D' as proposed by Beatty (1975); in the AXD angle and AD' distance, points J and Y as introduced by Järvinen (1982); in the JYD angle, as well as point M (Nanda and Merrill, 1994) and points AF and BF (Chang, 1987; in the AF – BF distance). Many of these angular techniques aim at eliminating point N from the equation, as this is believed to diminish variability.

Selecting different reference planes

Other authors have tried limiting geometric or growth influences by selecting different reference planes to which points A and B are related. Hall-Scott (1994) introduced the mandibulomaxillary bisector, the maxillary plane (ANS – PNS) as proposed by Ferrazzini (1976), FH as suggested by Chang (1987; in the AF – BF distance and the A – NV and B – NV distance) and by Yang and Suhr (1995) in the F – H to AB plane angle, the nasion perpendicular recommended by McNamara (1984), the N – Po line as suggested by Holdaway (1983), and the anterior cranial base according to Taylor (1969). Since the palatal plane was found to be relatively stable in longitudinal cephalometric studies and because it is more easily located compared with the occlusal plane, Williams et al. (1985) proposed projecting points A and B onto a constructed occlusal plane, angled 8 degrees to the palatal plane. The reasons listed for selecting these reference planes include their superior anatomic stability over time, both in absolute terms (Williams et al., 1985), as relative to the jaws (Hall-Scott, 1994; i.e. the reference plane follows the rotation of the jaws), or that the anatomic points defining the reference planes are more easily discernible (Holdaway, 1983; McNamara, 1984; Williams et al., 1985; Chang, 1987; Hall-Scott, 1994; Ferrazzini, 1976; Yang and Suhr, 1995).

Floating norms and geometric calculations

The previously mentioned points and planes were all proposed in order to minimize geometric distortions. In doing so, it was hoped that the technique-specific cut-off points used for discriminating the various Classes of skeletal discrepancy would maintain their applicability throughout the highly variable population. Another approach would be to accept geometric distortion of the measurements, but to compensate for these distortions by modifying the cut-off points accordingly: cut-off points are individualized/calculated using various statistically determined cephalometric parameters. From a sample of 96 (dental) Class I patients, Panagiotidis and Witt (1976) calculated the

correlation between ANB, SNA, and SN – MP (mandibular plane to SN line) angles. They derived the following formula: $ANB = -35.16 + 0.4 SNA + 0.2 SN - MP$, reflecting the ANB angle that would be found in a Class I patient with angles SNA and SN – MP ($r = 0.808$). Calculating the theoretical ‘individualized’ ANB angle allows comparison with the actual measured value, the difference between the two representing a measure of the ‘true’ sagittal discrepancy. An analogous approach was adopted by Järvinen (1986) who found that 63 per cent of the variability in ANB could be explained by variations in SNA and SN – MP angles. Including the NSAr angles increased this figure to 65.9 per cent. This led to the formula $y = 0.472 x_1 + 0.204 x_2 - 43.386$, where $y = ANB$, $x_1 = SNA$, and $x_2 = SN - MP$ angle. As explained by Järvinen (1986), the calculated ANB value represents a floating (or individualized) norm, which can be compared with the measured value.

Yet another approach was proposed by Hussels and Nanda (1984). They calculated ANB geometrically using the formula:

$ANB = \tan^{-1} \left(\frac{a \sin \gamma}{b - a \cos \gamma} \right)$, where a is the distance from point A to B, b is the distance from point N to B, and $\gamma = SNB + NS - MP - 90$.

The resulting value is compared with the measured one to assess the true sagittal discrepancy. However, as pointed out by Järvinen (1986), their geometric approach supposes the AB plane is perpendicular to the occlusal plane in ‘normal’ patients. This may or may not be true, depending on the degree of eruption of the teeth and on the ever-present interindividual variations. Significant individual error may therefore result. In keeping with the studies of Panagiotodis and Witt (1976), Järvinen (1986), and Hussels and Nanda (1984), Kim and Vietas (1978) introduced the AP dysplasia indicator (APDI), calculated using the facial angle \pm the A-B plane angle \pm the palatal plane angle. The underlying philosophy was that it might be more advantageous to combine various single measurements in order to obtain a more robust interpretation of sagittal discrepancy. In fact, when correlating various cephalometric analyses to occlusal relationships, their index showed the highest coefficient among those investigated. Furthermore, although the difference with the Wits appraisal was small, the APDI showed a superior separation of the three skeletal Classes.

Optimizing cut-off points

It has been suggested that the disagreement between ANB and Wits measurements of skeletal discrepancy might be caused by inadequacies in the proposed cut-off points. For instance, Walker and Kowalski (1971) investigated the cephalograms of 474 males and 630 females and found an average ANB of 4.5 degrees [males: 4.65 degrees, standard deviation (SD) 2.23 degrees; females: 4.34 degrees, SD 2.66 degrees], which is considerably different from the originally proposed ideal value of 2 degrees (Steiner,

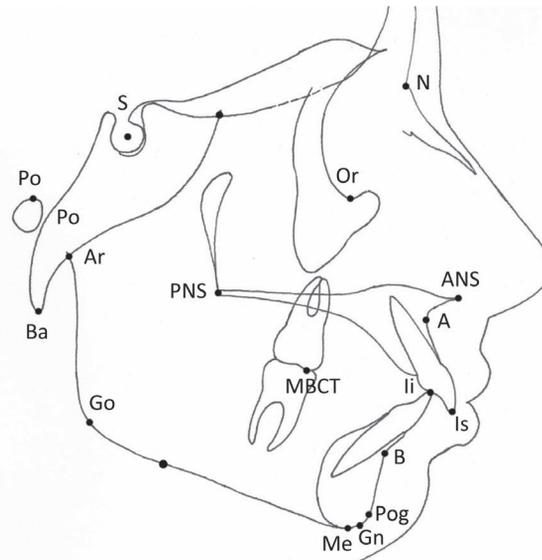
1953). Even when they included only patients classified as having a dental Class I occlusion, they found an average ANB of 4 degrees in approximately 1000 patients. Therefore, strict adherence to the ideal values of Steiner (1953), who emphasized that his proposed values should rather be used as 'rough estimates', could lead to Class I patients being misclassified as Class II. In a more recent report, Anderson et al. (2006) used receiver operator characteristic curves to determine and subsequently test optimized cut-off points for various cephalometric analyses, including the ANB and Wits analyses. Their conclusion was that using the optimized cut-off points improved accuracy in diagnosis for (among others) the Wits appraisal, in comparison with the conventional cephalometric norms. For the ANB analysis, the difference between the traditional and optimized cut-off points was not statistically significant.

Materials and methods

The study sample was obtained from the records of the author's private practice and consisted of 71 prospectively and consecutively collected patients, for whom good quality lateral cephalograms were available, using as the only additional inclusion criterion, the absence of craniofacial deformities. Of these 71 patients, 26 were male, mean age 13.1 years (SD 1.1 years, range 10.8 – 15.4) and 45 were female, mean age 14.6 years (SD 8.2 years, range 8.4 – 44.9). None had received previous orthodontic treatment. The lateral cephalograms were traced on a light box in a darkened room, using matte acetate tracing paper and a sharp pencil (Staedler 100-HB). The tracing paper was fixed to the cephalogram using tape, and background light due to size differences between the light box and the cephalogram was blocked out using cardboard. The landmarks used in the current research are shown in Figure 2 .

The finished tracing was placed approximately in the middle of the scanning surface of a desktop scanner (Scanjet8200, Hewlett-Packard, Palo Alto, California, USA). After scanning, the resulting image file was then imported into a digitizing software program (Digitizelt 1.5.7, I. Bormann, Bormisoft, Braunschweig, Germany). This program was used to determine the landmark co-ordinates using three calibration points, located on a transparent calibration sheet, which was included in the scan. The co-ordinates were subsequently imported into a graphing and curve-fitting program (FindGraph for Windows, version 1.482, UNIPHIZ Lab, 2001 – 2004, Tver, Russia), which was used to perform Procrustes analysis, a statistical technique which allows comparison of shape, independent of size (Halazonetis, 2004).

Figure 2. Digitized landmarks — point S: midpoint of the pituitary fossa of the sphenoid bone; point N: most anterior point of the frontonasal suture; porion: highest point of the meatus acousticus externus; orbitale: lowest point on the averaged left and right inferior margin of the orbit; articulare: intersection between the posterior border of the mandible, with the inferior outline of the cranial base; posterior nasal spine: the most posterior point in the median plane on the bony hard palate; anterior nasal spine, the tip of the median anterior process of the maxilla; basion: lowest point on the anterior margin of the foramen magnum, in the midsagittal plane; MBCT: the mesiobuccal cusp tip of the upper first molar; Is, tip of the crown of the most anterior maxillary central incisor; li, tip of the crown of the most anterior mandibular central incisor; interincisal point: the midpoint between Is and li; point A, deepest point on the anterior surface of the maxilla between ANS and prosthion; point B, deepest point on the anterior surface of the mandibular symphysis between infradentale and pogonion; pogonion: most anterior point of the mandibular symphysis; gnathion: most anterior and inferior point on the contour of the mandibular symphysis, constructed by bisecting the angle formed by the mandibular plane and N – Pog line; menton, most inferior point of the mandibular symphysis; gonion: most posterior and inferior point of the mandibular angle, determined by bisecting the angle formed by the tangent to the posterior border of the mandible and the mandibular plane.



Procrustes analysis attempts to find a ‘best fit’ of various clusters of analogous points. For this project, one cluster of points consisted of the template’s landmarks (12-year male-female average template of the Bolton-Brush growth study; Broadbent et al., 1975), while the other was represented by the tracing’s reference points, which were digitized earlier. The procedure consists of three discrete steps: firstly, the template is shifted to align its centroid (its centre of gravity or midpoint) with that of the tracing’s landmarks. Secondly, the template is rotated, minimizing the distances between the corresponding points of the template to those of the tracing. Finally, the template is scaled (inflated or shrunk ‘isomorphically’, without changing proportions) in order to remove size differences between the two clusters of points. The latter is performed by calculating the centroid size: the square root of the sum of the squared distances of each point to the centroid. The centroid size of the translated and rotated template is then matched to that of the tracing.

The Bolton template co-ordinates resulting from the Procrustes analysis (the template co-ordinates after it was translated, rotated, and scaled, in order to find the best fit relative to the patient's tracing) were exported back to Excel, where all further calculations were performed: the patient's digitized landmarks were used to determine ANB and Wits values, the individualized ANB angle according to Hussels and Nanda (1984), the floating norm according to Järvinen (1986), as well as the APDI as proposed by Kim and Vietas (1978). Also calculated were the AB-BB measurement according to Hall-Scott (1994), the AP-BP measurement introduced by Ferrazzini (1976), the AF-BF value proposed by Chang (1987), the ASN-BSN measurement according to Taylor (1969), and the AB plane angle as suggested by Downs (1948).

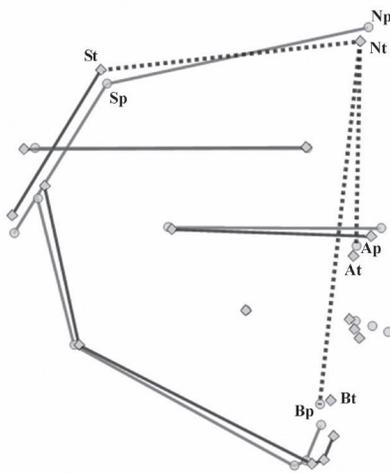
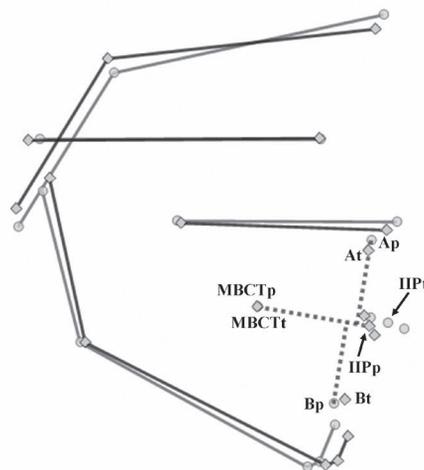


Figure 3. ANB analysis, as performed in the proposed technique, is calculated by connecting the Bolton template's point N, designated Nt (which is the point N resulting from the application of Procrustes analysis to the Bolton template), with the patient's points Ap and Bp, rendering the angle Ap-Nt-Bp. Also depicted are the Bolton template's points A, B, and S, designated At, Bt, and St, as well as the patient's points S and N, designated Sp and Np. The patient's landmarks are depicted as circles while the squares represent the Bolton template points.

Figure 4. The Wits appraisal, as performed in the proposed technique, is determined by projecting the patient's points A and B, designated Ap and Bp, onto the Bolton template's occlusal plane MBCTt – IIPt (where MBCTt and IIPt represent the mesiobuccal cusp tip and interincisal point resulting from the application of Procrustes analysis to the Bolton template). Also depicted are the patient's mesiobuccal cusp tip and interincisal point, designated MBCTp (coincides with MBCTt) and IIPp, as well as the Bolton template's points A and B, designated At and Bt. The patient's landmarks are depicted as circles while the squares represent the Bolton template points.



In addition, ANB and Wits values were normalized: the post-Procrustes coordinates for the Bolton template's points S and N (designated St and Nt, Figure 3) and the patient's points A and B were used to calculate the normalized ANB angle (ANBn), while the Bolton template's occlusal plane (constructed from MBCTt and IIpt, Figure 4) and the patient's points A and B were used to compute the normalized Wits analysis (WITSn). Similarly, normalized versions were calculated for the AB-BB measurement (AB-BBn) by determining the Bolton template's mandibulomaxillary bisector (again after performing Procrustes analysis), on which the patient's points A and B were projected. The Bolton template's palatal plane was used to determine the normalized APP-BPP value (APP-BPPn). Projecting the patient's points A and B onto the Bolton template's FH rendered the normalized AF-BF measurement (AF-BFn). Finally, the normalized AB plane angle was determined by measuring the inner angle formed by the AB plane (constructed from the patient's points A and B) and the Bolton template's post-Procrustes N-Pog line, rendering ABPAn, whereas projecting the patient's points A and B onto the Bolton template's SN line revealed the normalized ASN-BSN measurement (ASN-BSNn). The term normalized refers to the template (or norm's) reference landmarks/planes used in the analysis. Correlation calculations were performed between classic ANB and Wits, the AB-BB, AP-BP, AF-BF, and ABPA, values as well as ASN-BSN. The same procedure was repeated for their normalized counterparts (i.e. after performing Procrustes analysis).

Statistical procedure

All tests were performed using the Statistical Package for Social Sciences (SPSS Inc., Chicago, Illinois, USA). Significance was predetermined at the 0.05 per cent level of confidence. Intragroup comparison of males and females regarding SNA, SNB, and ANB (Riedel, 1952), Wits (Jacobson, 1975), Järvinen's floating norm (1986), individualized ANB (Hussels and Nanda, 1984), and APDI (Kim and Vietas, 1978) were performed using either t- or Mann-Whitney U-tests, depending on Levene's test to confirm homogeneity of variance and the Shapiro-Wilk test to assess normality of the distribution. The correlation between the various measurements for sagittal discrepancy was calculated using Spearman's rank correlation coefficient.

Error analysis

The entire procedure was repeated for 15 randomly selected cases, at least 2 weeks apart. Statistical significance was determined using paired t-tests. The overall method error was determined using the standard error of the method: $s = \sqrt{\frac{2d^2}{2n}}$, where d represents the difference between the corresponding repeated measurements, and n the number of measurements performed.

Results

Repeated measurements for ANB, Wits, AB-BB, AP-BP, AF-BF, ABPA, ASN-BSN, and their normalized counterparts (ANBn, WITSn, AB-BBn, AP-BPn, AF-BFn, ABPAn, and ASN-BSNn) did not reveal any statistically significant results (Table 1). Pearson's correlation coefficient varied from 0.957 (AB-BBn) to 0.984 (ASN-BSNn), all coefficients being highly significant ($P < 0.001$). Overall method error ranged from 0.34 mm (ASN-BSNn) to 0.66 mm (AF-BF), which was deemed to be acceptable when compared with the SDs. Comparing the overall method error for the classic tests with those of the normalized counterparts, it was found that the values were comparable, indicating that the normalized ANB, Wits, AB-BB, AP-BP, AF-BF, AB plane angle, and ASN-BSN measurements are as reproducible as their classic counterparts.

Table 2 summarizes the results for the intragroup comparison of males and females. Both groups were significantly different in age, the females being approximately 1.5 years older than the males. However, since no statistically significant differences were found in any of the classic methods to describe the mandibular and maxillary sagittal position and relationship (SNA, SNB, ANB, and Wits), in the floating norm methods of assessing sagittal discrepancy as proposed by Järvinen (1986) and Hussels and Nanda (1984), in the APDI by Kim and Vietas (1978), and in the vertical dimension (GoGn-SN values), it was considered both groups were skeletally sufficiently 'matched' to group males and females for further analysis.

Table 3 lists the descriptive statistics for the pooled sample. Since several grouped parameters were not normally distributed, correlations were determined using Spearman's rank correlation coefficient instead of Pearson's correlation coefficient. The classic and normalized cross-tabulated correlation coefficients for the measurements are listed in Tables 4 and 5, respectively. Applying the currently proposed technique to ANB, Wits, AB-BB, AP-BP, AF-BF, ABPA, and ASN-BSN heightened the correlation coefficients between all tests.

The smallest improvement in concordance was found for the correlation ABPA versus ANB, from -0.907 (classic tests) to -0.959 (normalized tests). The greatest improvement was seen for the correlation ABPA versus AP-BP, from -0.545 (classic tests) to -0.950 (normalized tests). All other correlation coefficients obtained after applying Procrustes analysis were above 0.742 (in absolute value): from 0.742 for the ASN-BSNn versus AB-BBn to -0.995 for the ABPAn versus WITSn. Normalizing ANB and Wits measurements increased the correlation to 0.964 in comparison with the rather modest value of 0.624 between their classic counterparts.

The results for the application of ANB and Wits to each patient, compared with the normalized counterparts, are shown in Figure 5. From this graph, it is evident that the cluster of points representing the classic tests is scattered rather loosely around the regression line $ANB = 3.26 + 0.38 \text{ Wits}$ (95 per cent confidence interval lower bound: $ANB = 2.81 + 0.28 \text{ Wits}$; upper bound: $ANB = 3.70 + 0.48 \text{ Wits}$), in comparison with the much smaller dispersion of the normalized points cluster (regression line $ANB_n = 2.87 + 0.76 \text{ WITS}_n$, 95 per cent confidence interval lower bound: $ANB_n = 2.73 + 0.72 \text{ WITS}_n$, upper bound: $ANB_n = 3.00 + 0.81 \text{ WITS}_n$).

Table 1: Method error analysis.

Measurement (original/repeated)	Paired Differences		t	Significance (2-tailed)	Correlation (Pearson)	Standard error of the method
	Mean	SD				
ANB	0.25	0.76	1.27	NS	0.982	0.43
Wits	0.09	0.86	0.40	NS	0.965	0.59
AB-BB	0.31	0.74	1.65	NS	0.976	0.55
AP-BP	0.27	0.80	1.31	NS	0.977	0.58
AF-BF	0.13	0.96	0.54	NS	0.969	0.66
ABPA	-0.20	0.94	-0.81	NS	0.970	0.37
ASN-BSN	-0.05	0.70	-0.27	NS	0.982	0.48
ANB _n	0.12	0.56	0.85	NS	0.961	0.39
WITS _n	0.17	0.68	0.99	NS	0.965	0.48
AB-BB _n	0.22	0.79	1.09	NS	0.957	0.56
AP-BP _n	0.09	0.56	0.65	NS	0.977	0.39
AF-BF _n	0.07	0.53	0.48	NS	0.980	0.37
ABPA _n	-0.22	0.92	-0.92	NS	0.973	0.65
ASN-BSN _n	-0.01	0.50	0.05	NS	0.984	0.34

NS, non-significant.

Table 2: Intergroup comparison of males and females using Mann-Whitney U-test.

Measurement (original/repeated)	Males		Females		P	Sign.
	Mean	SD	Mean	SD		
Age	13.1	1.1	14.6	8.2	0.04	*
SNA	81.4	3.0	81.1	3.6	0.91	ns
SNB	77.1	2.9	75.5	12.0	1.00	ns
ANB	4.3	2.1	3.9	2.3	0.15	ns
Wits	2.7	3.1	1.7	4.2	0.12	ns
Hussels & Nanda (1984)	4.3	2.1	3.9	2.4	0.19	ns
Järvinen (1986)	2.7	1.7	2.4	2.0	0.39	ns
APDI (Kim and Vietas, 1978)	81.1	4.3	81.8	6.8	0.17	ns
GoGn-SN	32.3	5.1	32.1	5.6	0.75	ns

NS, non-significant. * $P < 0.05$.

Table 3: Descriptive statistics for the pooled sample.

	Mean	SD	Minimum	Maximum
Age	14.1	6.6	8.5	44.9
SNA	81.2	3.4	74.5	89.0
SNB	76.1	9.7	0.0	62
ANB	4.0	2.2	-0.8	12.2
Wits	2.0	3.8	-4.5	17.8
Hussels & Nanda (1984)	4.0	2.3	-0.7	12.0
Järvinen (1986)	2.5	1.9	-1.0	8.4
APDI (Kim and Vietas, 1978)	81.6	6.0	63.4	93.2
GoGn-SN	32.2	5.4	20.4	47.3

Table 4: Correlation calculations for the classic measurements using Spearman's summed rank test.

	ANB	WITS	AB-BB	AP-BP	AF-BF	ABPA	ASN-BSN	H and N (1984)	Järvinen (1986)	APDI (1978)
ANB	1									
WITS	0.624	1								
AB-BB	0.732	0.807	1							
AP-BP	0.686	0.656	0.677	1						
AF-BF	0.725	0.598	0.679	0.842	1					
ABPA	-0.907	-0.685	-0.811	-0.545	-0.581	1				
ASN-BSN	0.690	0.596	0.571	0.884	0.844	-0.530	1			
H and N (1984)	0.967	0.580	0.664	0.643	0.701	-0.855	0.668	1		
Järvinen (1986)	0.705	0.769	0.932	0.665	0.711	-0.770	0.697	0.645	1	
APDI (1978)	-0.595	-0.437	-0.590	-0.477	-0.836	0.533	-0.534	-0.579	-0.631	1

H and N = Hussels and Nanda. All correlation coefficients are highly significant ($p < 0.001$)

Table 5

	ANBn	Witsn	AB-BBn	AP-BPn	AF-BFn	ABPA	ASN-ANBn	H and N (1984)	Järvinen (1986)	APDI (1978)
ANBn	1									
Witsn	0,964	1								
AB-BBn	0,946	0,98	1							
AP-BPn	0,944	0,966	0,908	1						
AF-BFn	0,919	0,938	0,865	0,993	1					
ABPAn	-0,959	-0,995	-0,985	-0,950	-0,9016	1				
ASN-BSNn	0,835	0,846	0,742	0,948	0,972	-0,816	1			
H and N (1984)	0,847	0,811	0,758	0,841	0,834	-0,793	0,798	1		
Järvinen (1986)	0,854	0,909	0,893	0,871	0,834	-0,91	0,744	0,645	1	
APDI (1978)	-0,653	-0,688	-0,675	-0,701	-0,688	0,683	-0,625	-0,579	-0,631	1

All correlations are highly significant ($p < 0,001$)

Discussion

The value for the correlation between ANB and Wits (0.624, Table 4) corresponds almost perfectly with those obtained by Roth (1982) in his positive Wits group and Järvinen (1981), and Richardson (1982), but is somewhat low in comparison with various other studies (Millett and Gravely, 1991; Del Santo, 2006). The latter is probably due to differences in sample size and selection criteria. The rather modest correlation of 0.624 signifies that about 39 per cent of the variability in ANB measurements can be explained by variations in the Wits value and vice versa. Consequently, both analyses tend to sometimes disagree, giving rise to a great deal of uncertainty as to which one of the two (if any) measurements is correct.

Figure 5 demonstrates the fairly large variation in ANB values that may be found associated with small variations in Wits appraisal. It seems most authors agree that this variation is caused by geometric distortion (Järvinen, 1981; Roth, 1982; Hussels and Nanda, 1984; Chang, 1987; Jacobson, 1988; Del Santo, 2006) which results from the application of ANB analysis and Wits appraisal to patients who, by definition, vary interindividually in the position of their reference landmarks: S, N, and the occlusal plane. Consequently, both analyses not only measure sagittal discrepancy but also equally represent variations in the AP and vertical position of N, in the vertical dentoalveolar dimension, in the rotation of both jaws in relation to the cranial base, and in molar and incisor eruptive status (which in turn determines the cant of the occlusal plane).

If the ANB analysis and Wits appraisal are to better correspond and hence more often agree on the admittedly ill-defined concept of sagittal discrepancy, it seems the key to improving their correlation lies in limiting geometric distortion of the measurements. If successful, this would not only improve the correlation between Wits appraisal and ANB analysis but could also lead to an improvement in their sensitivity and specificity. The presently proposed technique tries to limit geometric distortion by eliminating, as much as possible, individual variations in the position of the reference landmarks: S, N, and the occlusal plane. This is accomplished by 'fitting' a template on the patient's digitized landmarks using Procrustes analysis (Halazonetis, 2004). The measurements are then performed using the patient's points A and B and the template's points S and N as well as the template's occlusal plane.

From Table 5, it would appear that the technique performs adequately: the correlation between ANB_n and WITS_n increased to 0.964, in comparison with the value of 0.624 found for the classic tests (Table 4). Therefore, 93 per cent of the variability in the ANB_n measurement can be predicted by variations in the WITS_n measurement, in comparison with 39 per cent for their classic counterparts. This is graphically demonstrated by

the smaller dispersion of the points around the regression line in Figure 5. Although correlation itself does not represent a measure of diagnostic performance, the improved correlation between ANB and Wits considerably lowers the possibility of differing (Class I-Class II or Class I-Class III) or even totally opposing (Class II-Class III) diagnostic outcomes resulting from the application of both tests to (non-borderline) patients and could therefore possibly improve their diagnostic power.

Similar results were found when the present methodology was applied to other methods for determining sagittal discrepancy (Tables 4 and 5): projections on the mandibulomaxillary bisector (Hall-Scott, 1994), palatal plane (Ferrazzini, 1976), FH plane (Chang, 1987), the SN line (Taylor, 1969), as well as the AB plane angle (AB to N-Pog, Downs, 1948). Not surprisingly, the lowest improvements in correlation occurred for measurements that were already well correlated initially: the correlation between ABPAn and ANBn ($r = -0.959$, Table 5) improved only -0.052 , but was -0.907 initially. The second lowest improvement, 0.064 , was found for ASN-BSN versus AP-BPn ($r = 0.948$, Table 5), which started from an acceptable correlation of 0.884 (Table 4). Conversely, the highest improvements in correlation seem to have occurred for measurements that

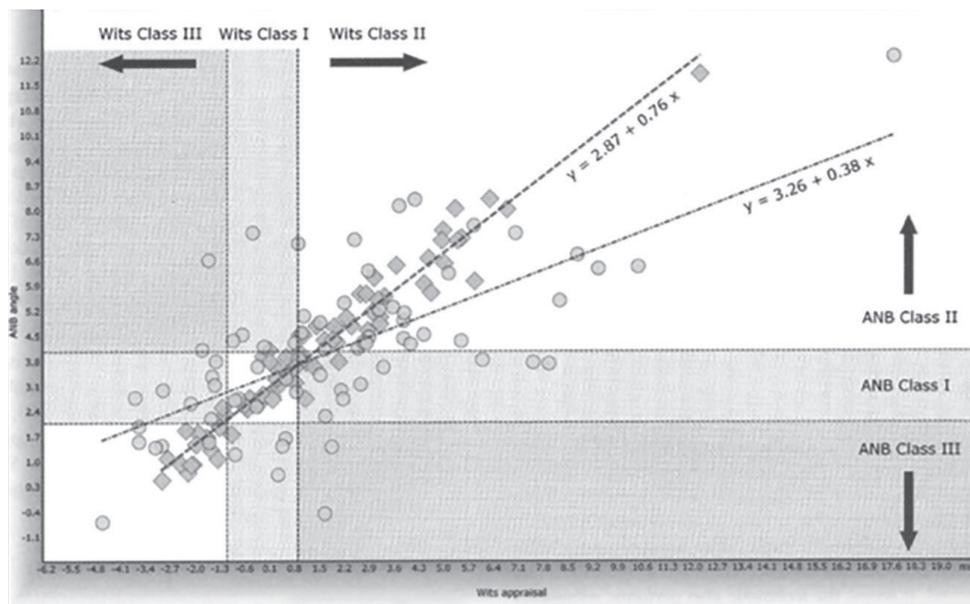


Figure 5. Relationship of the Wits appraisal (x -axis) to ANB angle (y -axis). The classic ANB and Wits analyses are depicted as circles, while the squares represent the normalized values. The cluster of points representing the classic tests is scattered loosely around the regression line $y = 3.26 + 0.38x$. In contrast, the normalized values correspond much closer their regression line $y = 2.87 + 0.76x$. The areas shaded light grey represent areas of ambiguity, while those shaded dark grey represent areas of contradiction.

were poorly correlated initially; the lowest correlation coefficient of -0.530 was found for ASN-BSN versus ABPA (Table 5). It improved from -0.286 after applying the currently proposed technique to -0.816 (Table 4). The correlation coefficient for AF-BF versus ABPA (Table 5) measured -0.581 before applying Procrustes analysis, while a value of -0.916 was found after normalization (Table 4). Since no deterioration was found for high initial correlations and low correlations were considerably improved, it seems the technique also performs adequately when applied to the aforementioned other methods for determining sagittal discrepancy.

The basic question of course is whether the proposed methodology is valid. The template fitted on the patient's tracing was the Bolton 12-year male-female averaged template, which by definition is indifferent to gender. However, as mentioned by Broadbent et al. (1975) and confirmed by Halazonetis (2007), males and females tend on average to differ mainly in size, more so than in shape. In fact, after removing gender-related size differences using Procrustes analysis, Halazonetis (2007) stated that the differences between the male and female average tracings were so small that they were hardly visible to the naked eye. Since Procrustes analysis was equally used in the current project to adjust the size of the template to the patient's tracing, the use of an averaged template in the current project would seem justified.

The choice of the 12-year template was rather arbitrary. Since the patient sample was gathered prospectively, it was not known at the start what the average patient age would be. Secondly, the classic templates as developed by Broadbent et al. (1975) and Popovich and Thompson (1977) are generally used for direct comparison, after being superimposed on the patient's tracing. The relevance of the measurements therefore depends on selecting the right age and consequently the right size of template. Since in the current project the template size (and position) was adjusted using Procrustes analysis, the choice of template was far less critical. The landmarks digitized in the present study were selected mainly because it was felt that they optimally characterized the anatomical structure of interest (mandible, maxilla, skull base, and FH). Points A and B were digitized, but they were omitted during the Procrustes analysis.

The correlation coefficient found in the current investigation was quite high. Tu et al. (2006) recently published a report regarding the problem of mathematical coupling, which concerns correlation and regression analyses. This is said to occur in situations where both aforementioned analyses are applied when 'the relationship between two variables is due to a common component, when one variable is part of the other, or when a third variable is common to both', mathematical coupling could cause misleading results.

It is, however, very unlikely that mathematical coupling applies to the current investigation for two reasons:

1. Direct mathematical coupling requires a relatively simple mathematical relationship between the variables under investigation (such as $ANB = SNA - SNB$; Tu et al., 2006), and therefore does not seem to apply. Although Järvinen (1985) demonstrated that the Wits analysis can be calculated from ANB, the required formula is quite complex:

$$Wits = \cos(\beta) \times \sqrt{NA^2 + NB^2 - 2 \times NA \times NB \cos(\alpha)} \text{ (Figure 6),}$$

where:

- β = angle formed between the occlusal plane and a line joining points A and B
- NA = distance between points N and A, NB = distance between points N and B
- α = ANB angle

It is interesting to note that Järvinen's formula uses a third factor, AB plane angle, to calculate the Wits from the ANB, which by virtue of its presence allows variability. Therefore, both analyses are no longer directly coupled. Another approach to calculate Wits from ANB uses the intersection between the SN line and the occlusal plane (Figure 7). As before, the below formula needs additional factors to differentiate one analysis from the other: adding the SN line to the equation introduces angle α and length IN, which again introduce variability. Since it seems impossible to calculate the Wits from ANB without additional factors, the presence of direct mathematical coupling in the present study seems highly unlikely.

$$A_o B_o = \frac{\sin(ANB) \sin(\alpha) IN}{\sin(180 - \alpha - SNA) \sin(180 - \alpha - SNB)} - \left(\frac{\sin(\alpha) IN}{\sin(180 - \alpha - SNA)} - NA \right) \\ \cos(180 - \alpha - SNA) - \cos(180 - \alpha - SNB) \left(NB - \frac{\sin(\alpha) IN}{\sin(180 - \alpha - SNB)} \right)$$

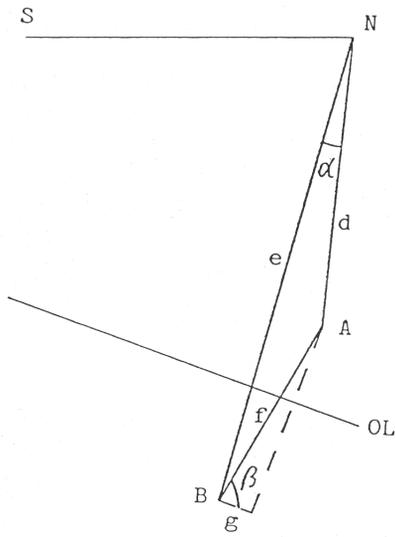
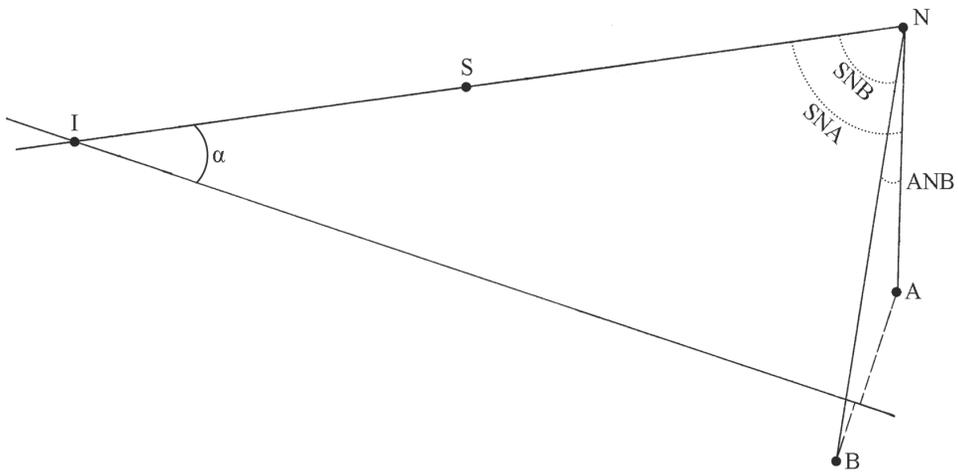


Figure 6. (left) Relationship between Wits appraisal and ANB angle, according to Järvinen (1985).

Figure 7. (below) Calculating the Wits appraisal from ANB angle, using the SN line. 'I' represents the intercept between the SN line and the occlusal plane, while alpha depicts the angle between these two lines. The definitions of the other landmarks can be found in Figure 2. The labels for the intersection between the perpendiculars dropped from A and B onto the occlusal plane, Ao and Bo, have been omitted for clarity.



2. Indirect mathematical coupling, which assumes changes in one variable, via a third variable or via an underlying physiological association to inevitably lead to related changes in the other variable (Tu et al., 2006) also does not seem to apply. Although ANB and Wits analysis are linked via the dentoalveolar relationship, simple geometric distortions can cause ANB to change without a concomitant change in the Wits appraisal or vice versa. This applies even in the current methodology, where individual variation is limited using reference landmarks from the superimposed template (hence fixing the mutual relationship of N to the occlusal plane). As an example, changing the vertical position of A and B perpendicular to the template's occlusal plane (without changing their mutual AP relationship) will change the value of ANBn, without changing WITSn.

Therefore, changes in one variable (ANBn), via a third variable or underlying physiological association (position of A and B), do not inevitably lead to changes in the other variable (WITSn).

Similar examples, in slight variations, could be presented for every pair of correlated measurements in Tables 4 and 5. As an example, when studying the correlation between AP–BP and AF–BF, changing the vertical position of points A and B perpendicular to the palatal plane will change the reading of the projection on FH, without a concomitant change in AP – BP. Since the prerequisites for the presence of indirect mathematical coupling are also not fulfilled, it seems the use of correlation analysis to describe the interrelationship between ANB and Wits measurements in the present investigation is warranted.

The correlation coefficients could probably have been improved further by limiting variability in the vertical dentoalveolar dimension. Roth (1982) proposed drawing a line through points A and B and constructing two alternative points for A' and B' on this line, at a fixed distance from one another (for instance, distance A' to B' is always 50 mm, the midpoint between them being located on the occlusal plane). The Wits appraisal would then be obtained using the two alternative points A' and B'. In doing so, Roth (1982) observed that the separation between the various Classes of sagittal discrepancy improved (less overlapping of these Classes was observed, where a Wits value could belong to Class I, Class II, as well as Class III patients).

Why standardizing the vertical dentoalveolar dimension might have improved the correlation between for instance ASN-BSN and other measurements, such as ANBn, WITSn, and ABPAN, can be explained by the oblique orientation of the SN line relative to the occlusal plane. Therefore, the ASN-BSNn measurement would be very sensitive to remaining variations in the vertical dentoalveolar dimension, in comparison with the less oblique planes such as the palatal plane or the mandibulomaxillary bisector. Standardizing the vertical dentoalveolar dimension would remove this variation and hence improve correlations. However, a conscious decision was taken not to apply this approach in the current project. Firstly, there is no real consensus on how the vertical dentoalveolar dimension should be standardized: as an alternative to the technique proposed by Roth (1982), one could construct perpendiculars on the occlusal plane through points A and B, to then locate points A' and B' on these lines, at a fixed distance from the occlusal plane. Furthermore, it might be contended that standardizing the vertical dentoalveolar dimension in addition to the proposed methodology makes it very hard to assess what is finally being measured: the measurement of sagittal discrepancy may be diluted by all these manipulations to the point of having little left to do with this measurement. Finally,

it would have been considerably more difficult to disprove the presence of mathematical coupling, should the vertical dentoalveolar dimension have been standardized.

The currently proposed methodology can also be used morphometrically, directly comparing the patient's tracing to the size- and position-corrected template. Figure 8a-c illustrates three patients, the first two of whom will be treated orthodontically, while the third represents the application of the current technique to a prospective surgery patient (the tracings were not corrected for cephalometric enlargement). As evident from these examples, information may be obtained in addition to the patient's sagittal discrepancy, such as growth pattern, mandibular length and position, palatal plane inclination, dental protrusion or retrusion, by direct comparison. Although the analysis itself might seem complicated to perform technically, Procrustes analysis can be included in any of the modern computer programs for cephalometric analysis with little programming effort. For everyday use in a clinical setting, this would clearly be preferable. However, as illustrated by this research, the procedure may also be performed using affordable, publicly available software programs, albeit requiring more work exchanging information back and forth between the various software packages. As the currently proposed technique seems to considerably improve the correspondence between ANB and Wits appraisal, it could be used morphometrically and can easily be expanded to include the third dimension; it seems to at least warrant further investigation.

Conclusions

Applying Procrustes analysis to fit the 12-year male-female averaged Bolton template on the patient's digitized landmarks, and combining the template's reference landmarks/planes with the patient's points A and B to determine the normalized measurements for determining sagittal discrepancy, increases the correlation between the various analyses in comparison with their classic counterparts. The significantly improved correspondence between the normalized analyses heightens the possibility of these tests agreeing on the patient's sagittal discrepancy and decreases the possibility of differing or even totally opposing diagnostic outcomes resulting from their application to (clear-cut) Class I, II, and III patients .

Acknowledgement

The author would like to express his gratitude towards the reviewers, for their much appreciated comments and suggestions, and to Professor E. A. BeGole, for her continuing support.

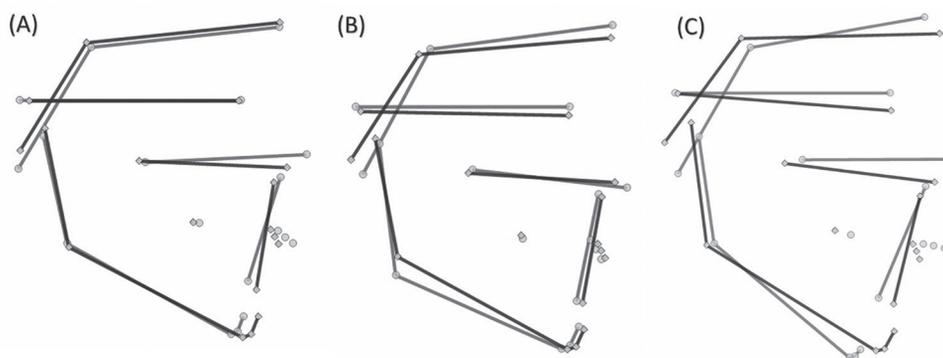


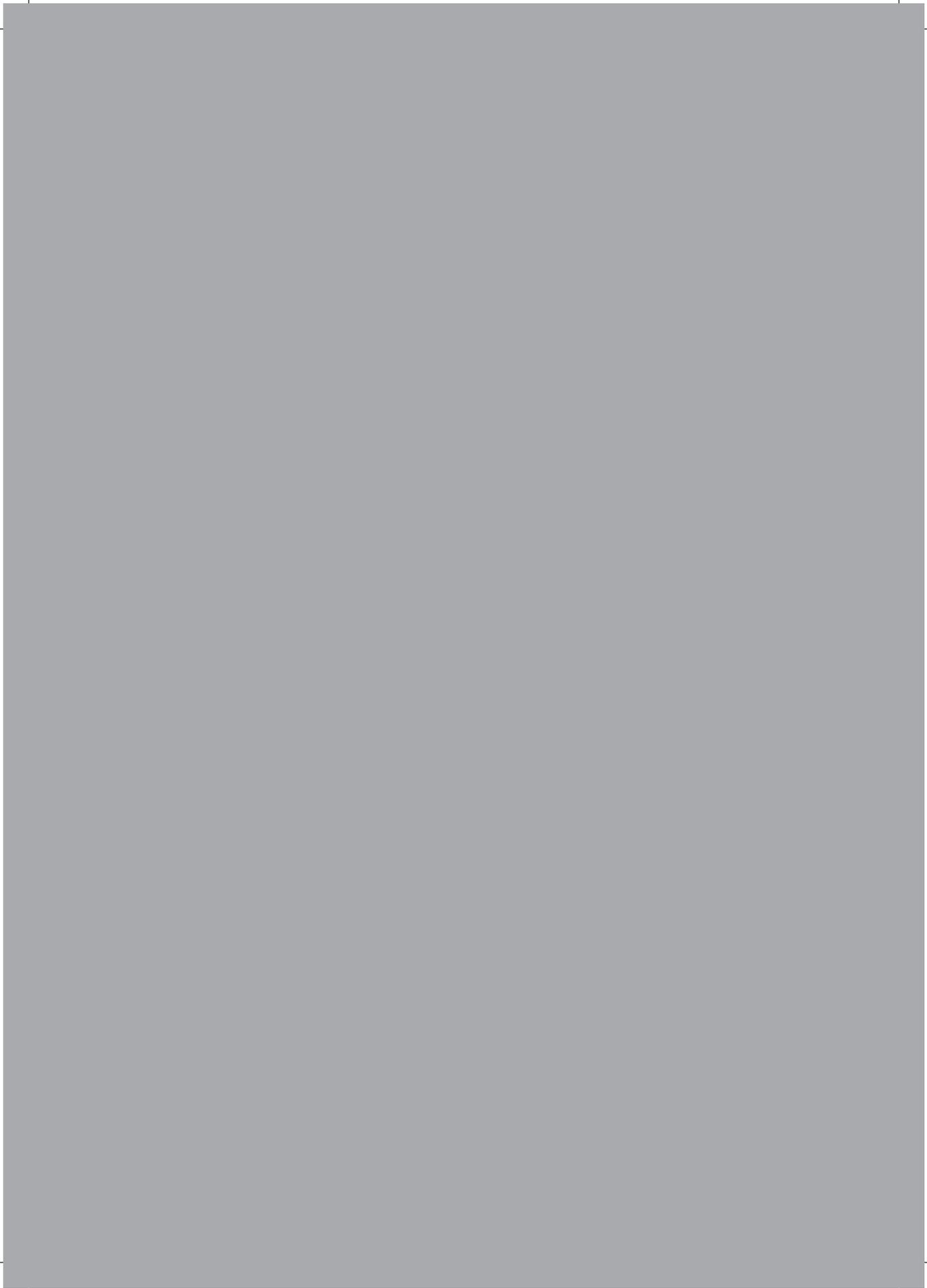
Figure 8. Examples of the morphometric application of the currently proposed technique to various patients. The template is depicted in black (squares) and the patient in grey (circles). In (a), the superimposition seems to suggest a short, retrognathic mandible, combined with slight maxillary prognathism, and a favourable mandibular plane angle. For the patient in (b), the maxillomandibular relationship seems to be normal, albeit with slight bimaxillary retrusion, a minor retrusion of the chin due to a somewhat shorter mandibular corpus, and an increased lower posterior face height. In (c), the proposed technique is applied to a surgery patient. A marked mandibular retrusion is combined with a minor maxillary protrusion and a considerably increased lower anterior face height. The dentition is clearly protrusive. Interestingly, the superimposition indicates posterior maxillary intrusion and mandibular autorotation/slight protrusion using a bimaxillary surgical approach as the treatment regimen for this patient.

References

1. Anderson G, Fields H W, Beck M, Chacon G, Vig K W 2006 Development of cephalometric norms using a unified facial and dental approach. *Angle Orthodontist* 76: 612–618
2. Atchison K A 1986 Radiographic examinations of orthodontic educators and practitioners. *Journal of Dental Education* 50: 651–655
3. Atchison K A, Luke L S, White S C 1991 Contribution of pre-treatment radiographs to orthodontists' decision making. *Oral Surgery, Oral Medicine, Oral Pathology* 71: 238–245
4. Baumrind S, Frantz R C 1971a The reliability of head film measurements. 1. Landmark identification. *American Journal of Orthodontics* 60: 111–127
5. Baumrind S, Frantz R C 1971b The reliability of head film measurements. 2. Conventional angular and linear measures. *American Journal of Orthodontics* 60: 505–517
6. Beatty E J 1975 A modified technique for evaluating apical base relationships. *American Journal of Orthodontics* 68: 303–315
7. Binder R E 1979 The geometry of cephalometrics. *Journal of Clinical Orthodontics* 13 : 258–263
8. Bishara S E, Fahl J A, Peterson L C 1983 Longitudinal changes in the ANB angle and Wits appraisal: clinical implications. *American Journal of Orthodontics* 84: 133–139
9. Broadbent B H 1931 A new X-ray technique and its application to orthodontia. *Angle Orthodontist* 1: 45–66
10. Broadbent B H Sr, Broadbent B H Jr, Golden W H 1975 Bolton standards of dentofacial developmental growth. Mosby, St Louis
11. Chang H P 1987 Assessment of anteroposterior jaw relationship. *American Journal of Orthodontics and Dentofacial Orthopedics* 92: 117–122
12. Del Santo M 2006 Influence of occlusal plane inclination on ANB and Wits assessments of anteroposterior jaw relationships. *American Journal of Orthodontics and Dentofacial Orthopedics* 129: 641–648
13. Demisch A, Gebauer U, Zila W 1977 Comparison of three cephalometric measurements of sagittal jaw relationship: angle ANB, Wits appraisal and AB-occlusal angle. *Transactions of the European Orthodontic Society*, pp. 269–281
14. Downs W B 1948 Variations in facial relationships: their significance in treatment and prognosis. *American Journal of Orthodontics* 34: 812–840
15. Ferrazzini G 1976 Critical evaluation of the ANB angle. *American Journal of Orthodontics* 69: 620–626
16. Freeman R S 1981 Adjusting A-N-B angles to reflect the effect of maxillary position. *Angle Orthodontist* 51: 162–171
17. Halazonetis D J 2004 Morphometrics for cephalometric diagnosis. *American Journal of Orthodontics and Dentofacial Orthopedics* 125: 571–581
18. Halazonetis D J 2007 Morphometric evaluation of soft-tissue profile shape. *American Journal of Orthodontics and Dentofacial Orthopedics* 131: 481–489

19. Hall-Scott J 1994 The maxillary-mandibular planes angle (MM degrees) bisector: a new reference plane for anteroposterior measurement of the dental bases. *American Journal of Orthodontics and Dentofacial Orthopedics* 105: 583–591
20. Han U K, Vig K W, Weintraub J A, Vig P S, Kowalski C J 1991 Consistency of orthodontic treatment decisions relative to diagnostic records. *American Journal of Orthodontics and Dentofacial Orthopedics* 100: 212–219
21. Hansen K, Bondemark L 2001 The influence of lateral head radiographs in orthodontic diagnosis and treatment planning. *European Journal of Orthodontics* 23: 452–453
22. Hofrath H 1931 Die Bedeutung der Röntgenfern- und Abstandsaufnahme für die Diagnostik der Kieferanomalien. *Fortschritte der Orthodontik* 1: 34–41
23. Holdaway R A 1983 A soft-tissue cephalometric analysis and its use in orthodontic treatment planning. Part I. *American Journal of Orthodontics* 84: 1–28
24. Hussels W, Nanda R S 1984 Analysis of factors affecting angle ANB. *American Journal of Orthodontics* 85: 411–423
25. Ishikawa H, Nakamura S, Iwasaki H, Kitazawa S 2000 Seven parameters describing anteroposterior jaw relationships: postpubertal prediction accuracy and interchangeability. *American Journal of Orthodontics and Dentofacial Orthopedics* 117: 714–720
26. Jacobson A 1975 The “Wits” appraisal of jaw disharmony. *American Journal of Orthodontics* 67: 125–138
27. Jacobson A 1976 Application of the “Wits” appraisal. *American Journal of Orthodontics* 70: 179–189
28. Jacobson A 1988 Update on the Wits appraisal. *Angle Orthodontist* 58: 205–219
29. Järvinen S 1981 A comparison of two angular and two linear measurements used to establish sagittal apical base relationship. *European Journal of Orthodontics* 3: 131–134
30. Järvinen S 1982 The JYD angle: a modified method of establishing sagittal apical base relationship. *European Journal of Orthodontics* 4: 243–249
31. Järvinen S 1985 An analysis of the variation of the ANB angle: a statistical appraisal. *American Journal of Orthodontics* 87: 144–146
32. Järvinen S 1986 Floating norms for the ANB angle as guidance for clinical considerations. *American Journal of Orthodontics and Dentofacial Orthopedics* 90: 383–387
33. Kim Y H, Vietas J J 1978 Anteroposterior dysplasia indicator: an adjunct to cephalometric differential diagnosis. *American Journal of Orthodontics* 73: 619–633
34. Lux C J, Burden D, Conrath C, Komposch G 2005 Age-related changes in sagittal relationship between the maxilla and mandible. *European Journal of Orthodontics* 27: 568–578
35. McNamara J A 1984 A method of cephalometric evaluation. *American Journal of Orthodontics* 86: 449–469
36. Millett D, Gravely J F 1991 Assessment of anteroposterior dental base relationships. *British Journal of Orthodontics* 18: 285–297
37. Nanda R S, Merrill R M 1994 Cephalometric assessment of sagittal relationship between maxilla and mandible. *American Journal of Orthodontics and Dentofacial Orthopedics* 105: 328–344

38. Panagiotidis G, Witt E 1976 The individualized ANB angle. *Transactions of the European Orthodontic Society*, pp. 255–260
39. Pancherz H, Sack B 1990 A critical analysis of the SNA, SNB and ANB angles in assessing orthodontic treatments. *Fortschritte der Kieferorthopädie* 51: 309–317
40. Popovich F, Thompson G W 1977 Craniofacial templates for orthodontic case analysis. *American Journal of Orthodontics* 71: 406–420
41. Richardson M 1982 Measurement of dental base relationship. *European Journal of Orthodontics* 4: 251–256
42. Riedel R A 1952 The relation of maxillary structures to the cranium in malocclusion and in normal occlusion. *Angle Orthodontist* 22: 142–145
43. Rotberg S, Fried N, Kane J, Shapiro E 1980 Predicting the Wits appraisal from the ANB angle. *American Journal of Orthodontics* 77: 636 – 642
44. Roth R 1982 The Wits appraisal—its skeletal and dento-alveolar background. *European Journal of Orthodontics* 4: 21–28
45. Rushton R, Cohen A M , Linney A D 1991 The relationship and reproducibility of angle ANB and the Wits appraisal. *British Journal of Orthodontics* 18: 225–231
46. Sherman S L, Woods M, Nanda R S, Currier G F 1988 The longitudinal effects of growth on the Wits appraisal. *American Journal of Orthodontics and Dentofacial Orthopedics* 93: 429–436
47. Steiner C C 1953 Cephalometrics for you and me. *American Journal of Orthodontics* 39: 729–755
48. Tanaka J L, Ono E, Filho Medici E, Cesar de Moraes L, Cezar de Melo Castilho J, Leonelli de Moraes M E 2006 Influence of the facial pattern on ANB, AF-BF, and Wits appraisal. *World Journal of Orthodontics* 7: 369–375
49. Taylor C M 1969 Changes in the relationship of nasion, point A, and point B and the effect upon ANB. *American Journal of Orthodontics* 56: 143–163
50. Tu Y K, Nelson-Moon Z L, Gilthorpe M S 2006 Misuses of correlation and regression analyses in orthodontic research: the problem of mathematical coupling. *American Journal of Orthodontics and Dentofacial Orthopedics* 130: 62–68
51. Walker G F, Kowalski C J 1971 The distribution of the ANB angle in “normal” individuals. *Angle Orthodontist* 41: 332–335
52. Williams S, Leighton B C, Nielsen J H 1985 Linear evaluation of the development of sagittal jaw relationship. *American Journal of Orthodontics* 88: 235–241
53. Yang S D, Suhr C H 1995 F-H to AB plane angle (FABA) for assessment of anteroposterior jaw relationships. *Angle Orthodontist* 65: 223–231



CHAPTER 3

Geometric morphometric analysis of craniofacial variation, ontogeny and modularity in a cross-sectional sample of modern humans

*This chapter is based on:
Wellens HLL, Kuijpers-Jagtman AM, Halazonetis DJ.
J Anat. 2013; 222(4):397-409.*

3

Abstract

Objective: This investigation aimed to quantify craniofacial variation in a sample of modern humans.

Methods: In all, 187 consecutive orthodontic patients were collected, of which 79 were male (mean age 13.3, SD 3.7, range 7.5-40.8) and 99 were female (mean age 12.3, SD 1.9, range 8.7-19.1). The male and female subgroups were tested for differences in mean shapes and ontogenetic trajectories, and shape variability was characterized using principal component analysis. The hypothesis of modularity was tested for six different modularity scenarios.

Results and conclusion: The results showed that there were subtle but significant differences in the male and female Procrustes mean shapes. Males were significantly larger. Mild sexual ontogenetic allometric divergence was noted. Principal component analysis indicated that, of the four retained biologically interpretable components, the two most important sources of variability were (i) vertical shape variation (i.e. dolichofacial vs. brachyfacial growth patterns) and (ii) sagittal relationships (maxillary prognathism vs. mandibular retrognathism, and vice versa). The mandible and maxilla were found to constitute one module, independent of the skull base. Additionally, we were able to confirm the presence of an anterior and posterior craniofacial columnar module, separated by the pterygomaxillary plane, as proposed by Enlow. These modules can be further subdivided into four sub-modules, involving the posterior skull base, the ethmomaxillary complex, a pharyngeal module, and the anterior part of the jaws.

Introduction

Human craniofacial growth and the morphological variance that comes with it have been the subject of long-standing interest. The functional matrix hypothesis (FMH) formulated by Moss (1962, 1997) suggests that craniofacial skeletal growth is directed mainly by the operational and spatial demands of developing neighboring 'functional volumes': skeletal muscles and multiple other tissues and organs, such as the brain, eyes, nasopharynx, masticatory system and even sinuses (Moss, 1962). According to this hypothesis, the craniofacial skeleton is therefore literally molded into shape, through time, by developing adjacent organs (Moss, 1962, 1997). The skeletal muscles represent the so-called periosteal matrix, while all other tissues and organs combined constitute the capsular matrix (Moss, 1962).

Although the basic principles of the FMH (Moss, 1962) are fairly widely accepted, there is some discussion regarding the amount to which basicranial growth might be under epigenetic (i.e. non-genetic) control, in addition to being molded into shape by developing surrounding tissues. Since the basicranium, by way of its central position in the skull, divides and connects the neuro- and viscerocranium, it might to some extent influence facial growth (Bastir & Rosas, 2006; Gkantidis & Halazonetis, 2011; Lieberman et al. 2000b, 2008). Lieberman & McCarthy (1999) and Lieberman et al. (2008) pointed out that basicranial growth occurs mainly through endochondral ossification at the synchondroses, thus allowing for cranial base flexion or extension. This occurs either through differing depositional or resorptive growth fields on either side (anteroposteriorly) of the synchondroses (i.e. drift) (Giles et al. 1981; Enlow & Hans, 1996; Lieberman & McCarthy, 1999) or through differential chondrogenic activity at the upper vs. lower margins (Giles et al. 1981; Lieberman & McCarthy, 1999). Contrary to the intramembranous ossification around many of the various organs constituting the capsular matrix, basicranial growth might therefore be under more intrinsic control (Jeffery & Spoor, 2002; Lieberman et al. 2008). Additionally, the midsphenoidal synchondrosis ossifies prior to birth, whereas the spheno-ethmoidal one usually does not fuse before 6 years of age (Lieberman & McCarthy, 1999). The spheno-occipital synchondrosis remains active up to approximately 12 years of age (Lieberman & McCarthy, 1999). The midline basicranium therefore reaches adult shape at about 7–8 years, contrary to the lateral cranial base (at about 11–12 years). Both structures attain their adult shape before the neurocranium and face (at 15–16 years) (Bastir et al. 2006), possibly constraining the growth and/or position of the latter structures (Lieberman et al. 2008).

From a general point of view, growing and developing organs should not impinge on one another. As a result, all organs must grow/develop in a more or less coordinated way.

The two closely related biological concepts of morphological integration and modularity have the potential to explain the aforementioned notion of coordination/balance of craniofacial growth (Moss, 1962, 1997; Lieberman et al. 2000a,b, 2008; Klingenberg et al. 2003; Bastir & Rosas, 2005; Mitteroecker, 2007; Bastir, 2008; Klingenberg, 2008; Mitteroecker & Bookstein, 2008; Klingenberg, 2009). Growing organs and their surrounding skeletal structures share abundant and strong interactions which can be anatomic, developmental, functional or genetic in nature (Bastir & Rosas, 2005; Mitteroecker, 2007; Bastir, 2008; Klingenberg, 2008, 2009; Mitteroecker & Bookstein, 2008). As such, they form morphologically tightly integrated organismal units, which are referred to as modules. The latter are usually defined as serving a common functional goal (Mitteroecker, 2007; Mitteroecker & Bookstein, 2008), being tightly integrated internally, while at the same time being relatively independent from other such units, with which they interact and from which they can be delineated clearly (Mitteroecker, 2007; Klingenberg, 2008, 2009; Mitteroecker & Bookstein, 2008). Morphological integration therefore refers to a high degree of structural interactivity, leading to tightly coordinated morphological development of the structures involved (e.g. strong covariation). Modularity, on the other hand, implies a relative independence thereof, due to a much lower degree of interactivity in terms of frequency and strength.

As pointed out by Bastir & Rosas (2005), the functional volumes and their associated skeletal structures in the FMH (Moss, 1962) are an example of morphological integration. On the other hand, Enlow's counterpart analysis (Enlow et al. 1969; Enlow, 1990; Enlow & Hans, 1996) can be regarded as an example of modularity: the growth counterparts are hypothesized to subdivide the skull into various relatively independent modules, both sagittally and vertically. It is important to note that modularity is not limited to a single developmental organizational level: what appears as a single module at a given level of complexity can represent multiple modules seen from the next, lower level of complexity (Bastir, 2008) (e.g. when inspecting specific substructures at a higher resolution). To shed some light on specific craniofacial mechanisms of morphological integration, it is essential to first identify and delimit modules (Klingenberg et al. 2003; Klingenberg, 2008, 2009). As stated above, the counterpart analysis (Enlow et al. 1969; Enlow, 1990; Enlow & Hans, 1996) divides the face into an anterior and posterior module, separated by the posterior maxillary plane. The anterior module, referred to as the nasomaxillary complex, has been identified as a tightly integrated facial block, together with the orbits (Enlow et al. 1969, 1990, 1996; Lieberman et al. 2000b; McCarthy & Lieberman, 2001). More specifically, the posterior maxillary plane has been found to maintain a 90° relationship to the neutral horizontal axis (NHA) in primates and humans (amongst others) (Enlow et al. 1969, 1990, 1996; Lieberman et al. 2000a,b; McCarthy & Lieberman, 2001). The NHA is defined anteriorly by the midpoint between the upper

and lower orbital rims, and posteriorly by the midpoint between the superior orbital fissures and the lower border of the optic foramen (McCarthy & Lieberman, 2001).

Biegert (1957) found that in non-human primates, basicranial flexion decreased as facial size increased. Combined with the fact that in humans, the much more rapid increase in brain size relative to facial size was accompanied by an increase in basicranial flexion, this led him to formulate the 'bi-directional hypothesis', which states that an increase in facial size relative to brain size is associated with a reduction in basicranial flexion (Biegert, 1957; Lieberman et al. 2008; Bastir et al. 2010). A strong morphological integration has also been reported between the bilateral middle cranial fossa and the width of the mandibular ramus (Bastir & Rosas, 2004; Lieberman et al. 2008), which in turn was found to be significantly less correlated with the midline cranial base (Bastir & Rosas, 2004). Bastir & Rosas (2005) concluded that the ethmomaxillary complex is tightly integrated with the mandible in modern humans.

Intriguingly, many studies report poor morphological correlations between the midline cranial base and various facial variables (facial breadth and height, vertical facial pattern, and mandibulo–maxillary relationships) (Lieberman et al. 2000b; Bastir & Rosas, 2006; Polat & Kaya, 2007; Proff et al. 2008). Based upon the 90° relationship of the posterior maxillary plane to NHA reported by Enlow et al. (1969, 1990, 1996), some studies have suggested that midline cranial base flexure could be developmentally limited for functional reasons, resulting in pharyngeal airway patency (Ross & Henneberg, 1995; McCarthy & Lieberman, 2001), although a definitive confirmation of this hypothesis has yet to be provided (Ross et al. 2004). Additionally, morphological correlations might change with growth and development (Arthur, 2002; Gkantidis & Halazonetis, 2011). Indeed, the midline cranial base was found to be slightly better correlated with the face compared with the lateral cranial base in children (Gkantidis & Halazonetis, 2011). The latter, however, retained and even strengthened its facial correlation during growth, contrary to the midline cranial base (Gkantidis & Halazonetis, 2011).

The aims of this study were threefold:

1. To evaluate craniofacial shape variance in a sample of modern humans (orthodontic patients) by applying principal component analysis.
2. To test the male and female subgroups for differences in mean shapes and ontogenetic trajectories.
3. To test six different hypotheses of modularity by applying the methodology proposed by Klingenberg (2009). Three scenarios involving two modules were considered, one involving three, and one, four separate modules.

Materials and methods

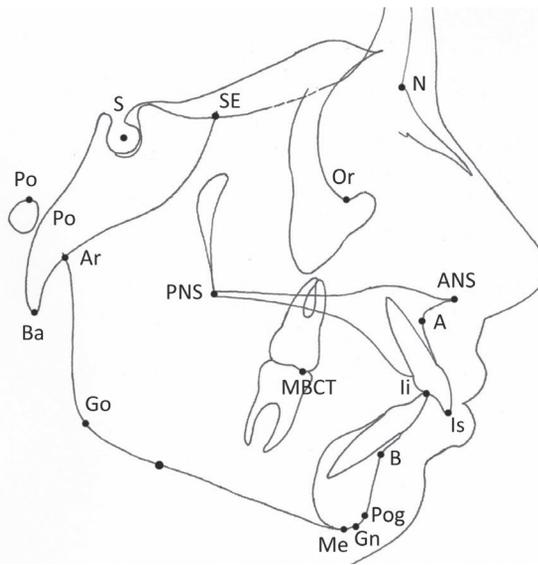
Lateral cephalometric radiographs of patients aged between 8 and 20 years, treated between April 2006 and May 2009 were collected from the records of the first author's private orthodontic practice. Additional inclusion criteria included the availability of good quality lateral cephalograms, the absence of craniofacial deformities, and that patients could only appear in the sample once. The resulting experimental group consisted of 178 patients, 79 of whom were male (mean age 13.3 years, SD 3.7 years, range 7.5–40.8 years) and 99 female (mean age 12.3 years, SD 1.9 years, range 8.7–19.1 years). All radiographs were taken with the same machine by a trained operator (H.L.L.W.), using a standardized technique. The lateral cephalograms were traced, by the same author, on a light box in a darkened room, using matte acetate tracing paper and a sharpened pencil. The landmarks used in the current project are shown in Fig. 1.

The finished tracing was placed approximately in the middle of the scanning surface of a desktop scanner (Scanjet 8200; Hewlett Packard, Palo Alto, CA, USA). The resulting image file was imported into a software program (DIGITIZEIT 1.5.7, I. Bormann; Bormisoft, Braunschweig, Germany) to record the landmarks' coordinates using three calibration points, located on a transparent calibration sheet, which was included in the scan. The recorded coordinates were then grouped in EXCEL (2010; Microsoft Corporation, Redmond, WA, USA) for subsequent analysis in R (<http://www.r-project.org>), MORPHOJ (Klingenberg, 2011) or VIEWBOX (dHal Software, Kifissia, Greece). Since radiographic magnification was the same for all lateral cephalograms, it was not accounted for.

Statistical analysis

All statistical analyses except the modularity test were programmed in R by the first author, and confirmed using VIEWBOX by the third author. To determine whether there were significant shape differences between male and female subjects, a permutation test was designed: the translated, scaled and rotated coordinates resulting from a pooled generalized Procrustes fit were used to calculate the Procrustes distance between the two groups' average configurations. Next, 1000 group pairs of the same size as the original male and female groups were created by randomly allocating the Procrustes coordinates to either group of each pair, without replacement.

Figure 1. Landmark definitions. Point S, midpoint of the pituitary fossa of the sphenoid bone; Point N, most anterior point of the frontonasal suture; Porion, highest point of the meatus acousticus externus; Orbitale, lowest point on the averaged left and right inferior margin of the orbit; Articulare, intersection between the posterior border of the mandible, with the inferior outline of the cranial base; Posterior nasal spine, the most posterior point in the median plane on the bony hard palate; Anterior nasal spine, the tip of the median anterior process of the maxilla; Basion, lowest point on the anterior margin of the foramen magnum, in the midsagittal plane; Point A, deepest point on the anterior surface of the maxilla between ANS and Prosthion; Point B, deepest point on the anterior surface of the mandibular symphysis between Infradentale and Pogonion; Pogonion, most anterior point of the mandibular symphysis; Gnathion, most anterior and inferior point on the contour of the mandibular symphysis, constructed by bisecting the angle formed by the mandibular plane and N-Pogonion line; Menton, most inferior point of the mandibular symphysis; Gonion, most posterior and inferior point of the angulus mandibulae, determined by bisecting the angle formed by the tangent to the posterior border of the mandible, and the mandibular plane; Spheno-ethmoidale, intersection between the anterior border of corpus of the Os sphenoidale with the inferior border of the Os ethmoidale.



The number of group pairs exhibiting a larger Procrustes distance as the one between the original two groups, divided by 1000, served as the P-value for the significance of the findings. A similar permutation test was used to evaluate potential size differences by randomly permuting the $\log(\text{centroid size})$ values (the natural logarithm of centroid size), without replacement. Finally, the male–female mean shape differences were revisited by rerunning the first permutation test while controlling for the effects of allometry.

For this purpose, the residuals resulting from the pooled within-group regression of shape over centroid size were used. In view of the relatively large age range in the experimental sample, it was deemed necessary to ascertain whether there were differences in the male and female ontogenetic shape trajectories: does craniofacial shape vary as a function of growth and development? If so, is this variation similar for male and female patients?

Size, in this context, is used as a (admittedly poor) proxy for development. The approach proposed by Mitteroecker et al. (2004) was applied to the $n \times m$ matrix of stereometrically projected Procrustes coordinates \mathbf{X} , whereby n is the number of rows, and m the number of columns. The vector of the 'common allometric component' (CAC), the component of shape change which is most closely aligned with size, was calculated as: $\mathbf{a} = \frac{\mathbf{X}'\mathbf{s}}{\mathbf{s}'\mathbf{s}'}$, and normalized as: $\mathbf{a} = \frac{\mathbf{a}}{\sqrt{\mathbf{a}'\mathbf{a}}}$. The CAC could then be visualized relative to $\log(\text{centroid size})$. The first residual component was calculated by first projecting out the CAC: $\mathbf{W} = \mathbf{X}(\mathbf{I} - \mathbf{a}'\mathbf{a})$, and then performing a singular value decomposition of $\mathbf{W}'\mathbf{W}$ into $\mathbf{V}\mathbf{D}_w\mathbf{V}'$. The columns of \mathbf{V} are the residual components, with scores $\mathbf{X}\mathbf{V}$. Additionally, the $n \times m$ matrix of stereometrically projected Procrustes coordinates \mathbf{X} was augmented with a column matrix \mathbf{s} , containing the logarithm of centroid size. Principal component analysis was applied to the resulting matrix, which allowed the plotting of the first three PC's of the resulting form space in a second three-dimensional plot.

The craniofacial variance of the experimental sample was further scrutinized by applying principal component analysis to the covariance matrix of the pooled generalized partial Procrustes coordinates (Zelditch et al. 2004). The number of biologically interpretable (i.e. non-trivial) PCs was determined using the 'Random average under permutation' rule, as outlined by Peres-Neto et al. (2005): the variables in the data matrix were randomized within variables 1000 times, and a PCA was performed on each reshuffled data matrix. The average eigenvalues were then calculated and compared with the ones obtained. If the observed exceeded the average random value, that axis was perceived as non-trivial. The percentage of variation explained by each of the non-trivial PCs was also calculated. Next, the hypothesis of modularity was tested with the MORPHOJ software package (Klingenberg, 2011), using the methodology proposed by Klingenberg (2009). Four scenarios involving two modules were considered, one with three modules, and one involving four modules. The same, pruned adjacency graph (Fig. 2b) was used for all scenarios, constructed beforehand using Delaunay triangulation (Delaunay, 1934), whereby connections that did not pass over continuous (skeletal) tissue were omitted (red lines in Fig. 2a) and, where needed, an additional diagonal was added to quadrilaterals (Klingenberg, 2009) (red lines in Fig. 2b).

Figure 3a depicts the structural subdivision, as used in the modularity scenarios involving two or three modules. In case of two modules, two of the three substructures were combined into one. Figure 3b represents the subdivision used when testing the counterpart principle, while Fig. 3c visualizes the location of the subdivisions in the four-module scenario. To correct for the effects of allometry, the modularity test was performed using the residuals resulting from a pooled within-group regression of shape over centroid size. The (multi-)RV coefficient (Klingenberg, 2009) could then be calculated for each scenario, to be compared with the corresponding value of randomly

generated alternative subdivisions into two to four spatially continuous modules. These modules would contain the same number of points as the corresponding modules in the proposed subdivision, the adjacency graph serving as an algorithmic tool to assure spatial continuity in these alternative modules.

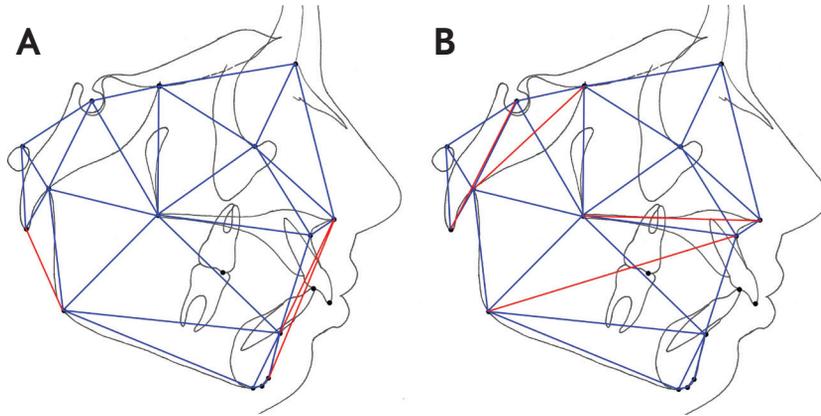


Figure 2. Adjacency graphs, constructed using Delaunay triangulation (Delaunay, 1934). The original adjacency graph is depicted in (A). The red lines in (A) indicate connections that were removed in (B) because they did not pass over continuous (skeletal) tissue (Klingenberg, 2009). The red lines in (B) represent diagonals that were added to selected quadrilaterals from (A) (Basion to Sella, Articulare to Spheno-ethmoidale, Gonion to Point A, and PNS to ANS) (Klingenberg, 2009).

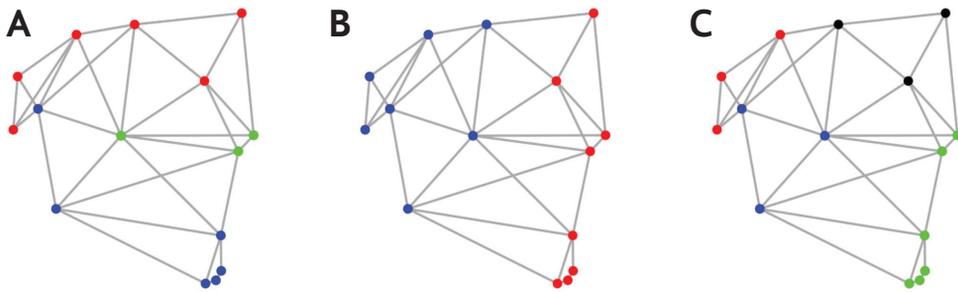


Figure 3. Subdivisions used during modularity hypothesis testing. The subdivisions in (a) were used either separately or combined in modularity scenarios 1, 2, 3 and 5. (b) The subdivisions employed when testing for the counterpart principle (modularity Scenario 4). (c) The four sub-modules which were considered in modularity Scenario 6. (Scenario 1) The presence of two modules (a): the skull base (in red) and maxilla (in green), vs. the mandible (in blue). (Scenario 2) The presence of two modules (a): the skull base and mandible (in red and blue, respectively) vs. the maxilla (in green). (Scenario 3) The presence of two modules (a): the skull base (in red) vs. the mandible and maxilla (in blue and green, respectively). (Scenario 4) The counterpart principle (b): the anterior vs. posterior module (in red and blue, respectively). (Scenario 5) The presence of three different modules (a): the skull base (in red), mandible (in blue) and maxilla (in green). (Scenario 6) Combining Scenarios 3 and 4: the four-module scenario (c). Note: The colors were used only to discriminate the various modules and are therefore not necessarily structurally consistent throughout the various scenarios depicted.

Each round of GPA superimpositions was performed using the simultaneous-fit approach (i.e. the superimpositions were performed while maintaining relative size among the modules). The multi-RV coefficient was used when testing for the presence of three or four modules, whereas the original RV coefficient was employed when only two modules were involved. All possible continuous alternative subdivisions were generated for comparison with the respective proposed modularity scenarios. The number of alternative subdivisions exhibiting a multi-RV coefficient lower than the proposed one was recorded as the P-value for the significance of the finding.

Error analysis

The digitizing procedure was repeated by the same author (H.L.L.W.) for 15 randomly selected cases, at least 2 weeks apart. Statistical significance was determined using a Procrustes analysis of variance.

Results

The error analysis did not reveal any statistically significant differences between the first and second digitizing round ($P = 0.175$) (Table 1). The male and female average configurations, calculated from a pooled generalized Procrustes fit, are depicted in Fig. 4. Very subtle differences can be observed, mainly at the Articulare and sphenio-ethmoidale landmarks. These shape differences were not significant, as indicated by the permutation test P-value of 0.33. Evidence of sexual dimorphism in size was found: the male patients' centroid sizes were found to be significantly larger ($P < 0.001$). None of the second permutation test's resampled datasets were found to have a larger difference in $\log(\text{centroid size})$ values. Rerunning the permutation test while controlling for the effects of allometry revealed modest, but highly significant, male–female mean shape differences ($P < 0.001$, Fig. 5a). These were exaggerated three times for visualization purposes in Fig. 5b. Apart from obvious differences at the Articulare and sphenio-ethmoidale landmarks, it appears females were slightly more orthognathic and dolichofacial.

Table 1: Summary of the ANOVA results for the error experiment.

Effect	SS	MS	df	<i>F</i>	<i>P</i>	
Centroid size						
Individual	0.120454	0.120454	1	0	0.9706	
Residual	2.439	87.120836	28			
Effect	SS	MS	df	<i>F</i>	<i>Pillai trace</i>	<i>P</i>
Shape						
Individual	0.0004516	1.73693E-05	26	0.13	0.97	0.1753
Residual	0.100258	0.000137717	728	1		

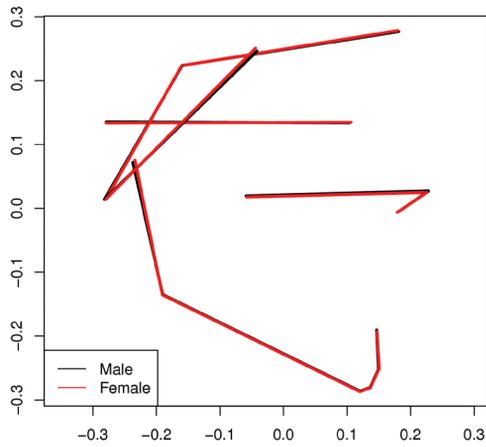


Figure 4. Mean configurations resulting from the pooled generalized Procrustes analysis. Subtle differences can be observed at the Articulare and Spheno-ethmoidale landmarks.

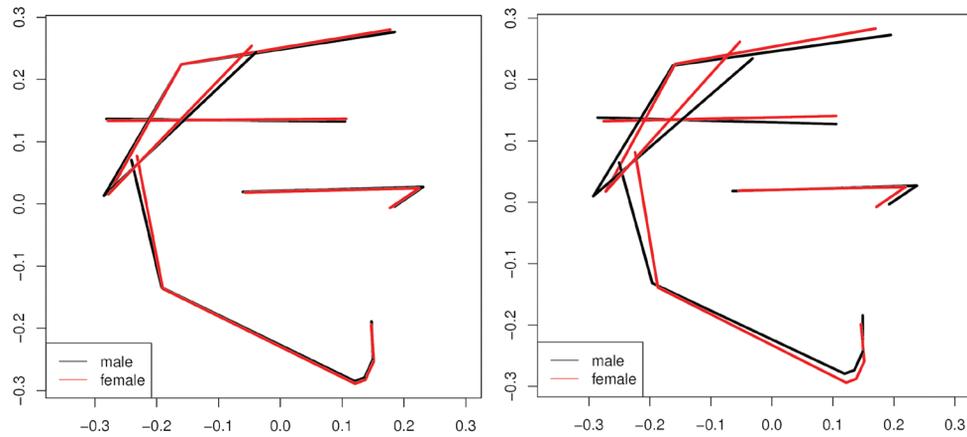


Figure 5. (a) (below) The male–female mean shape differences when controlling for the effects for allometry. Apart from obvious differences at the Articulare and Spheno-ethmoidale landmarks, it appears females were slightly more orthognathic and dolichofacial. (b) The differences exaggerated three times.

Two approaches were employed to assess possible divergences in the male and female ontogenetic allometric signals. The first method involved calculating the common allometric component (CAC: that component of shape change which most closely aligns with growth and development), as well as the residual shape components (RSC) (Mitteroecker et al. 2004). The first RSC, plotted relative to the CAC in Fig. 6, suggests that both sexes go through very similar ontogenetic shape changes: the shape trajectories are very similar, if not identical.

When plotting the CAC vs. $\log(\text{centroid size})$ in Fig. 7, a divergence in the male and female ontogenetic allometric signals could be observed. The statistical summary for

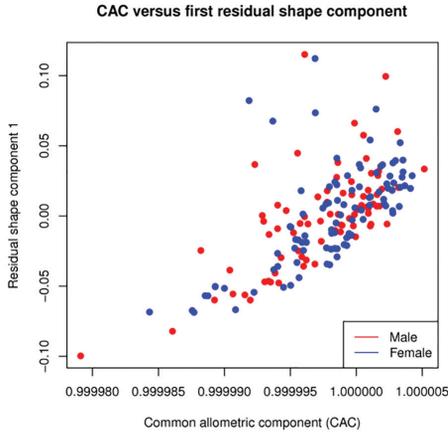


Figure 7. Scatterplot of the common allometric component (CAC, y-axis) vs. the natural logarithm of centroid size (x-axis). The slopes of both regression lines were now significant.

Figure 6. Two-dimensional projection of the rotated 3D scatterplot, representing the common allometric component (CAC, y-axis) vs. the first residual shape component (RSC1, x-axis). The sexual ontogenetic allometric trajectories seem to largely coincide.

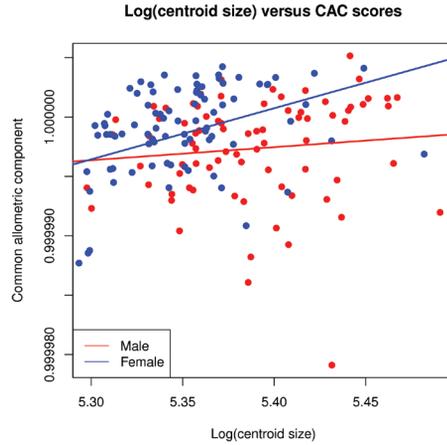


Figure 8. Scatterplot of the common allometric component (CAC, y-axis) vs. the natural logarithm of centroid size (x-axis), with the four most extreme male outliers removed. The male regression line was non-significant (Table 3) due to the presence of several outliers.

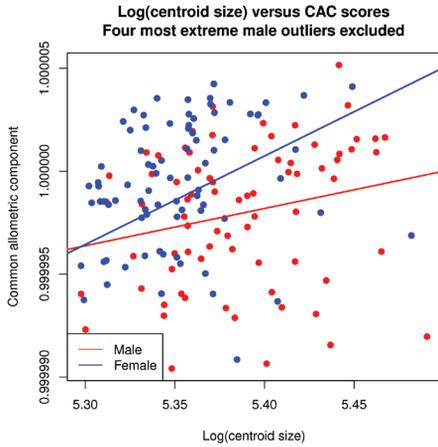
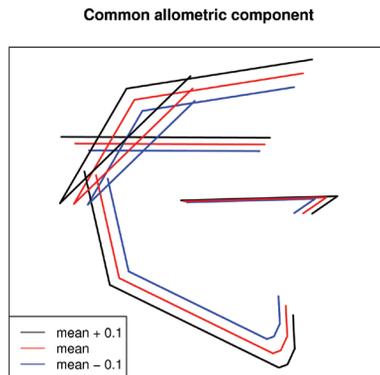


Figure 9. The common allometric component (CAC), visualized using the male Procrustes mean shape, with representations plus and minus 0.1 along the component axis. The CAC seems to represent a vector of almost pure size change, with little to no observable concomitant shape change.



the linear regression (Table 2) indicated that the slope for the male subgroup was not significant, probably due to the presence of several outliers in the male points cluster. Upon removing the four most extreme male outliers, the recalculated slope was found to be significant (Table 3 and Fig. 8), but the associated regression line still diverged from the female one. Intriguingly, the CAC was found to represent a vector of pure size change, with no clearly identifiable shape change (Fig. 9).

The second method involved augmenting the matrix of Procrustes coordinates with a column matrix holding the $\log(\text{centroid size})$ values, and subsequently performing a principal component analysis of the resulting matrix, premultiplied by its transpose. This allowed the first three PCs of the resulting Procrustes form space to be plotted in a three-dimensional plot (Fig. 10a,b). The resulting point scatter was quite spherical, with broad regions of overlap between the male and female point clusters.

In the PC1–PC2 view of the resulting 3D plot (Fig. 10a), there was a clearly discernible divergence in the male and female ontogenetic allometric signals. In smaller individuals, females tended to have higher PC2 scores in comparison with males, and vice versa for larger individuals. In contrast, the PC1–PC3 view revealed almost parallel trajectories (Fig. 10b). A bootstrap test was designed to confirm these visual impressions (10 000 iterations, with replacement). In the PC1–PC2 view, the angle between the male and female trajectories (0.257 radians) was found to be significant (P-value: 0.027; 95% confidence interval: -0.249 to 0.259 radians), contrary to the PC1–PC3 view (P-value: 0.406; 95% confidence interval: -0.164 to 0.169 radians).

Table 2: Summary of the linear regression results for the common allometric component (CAC) scores vs. $\log(\text{centroid size})$. The male regression line slope was not significant due to the presence of several outliers.

	Estimate	Std. Error	<i>t-value</i>	<i>P-value</i>
Male				
Intercept	0.9999	6.24E-05	16016.57	<0.001
Slope	1.81E-05	1.16E-05	0.95	0.346
Female				
Intercept	0.9998	4.10E-05	24417.46	<0.001
Slope	4.30E-05	7.67E-06	5.61	<0.001

Table 3: Summary of the linear regression results for the common allometric component (CAC) scores vs. log(centroid size), after removing the four most extreme male outliers. The slopes of both regression lines were now significant.

	Estimate	Std. Error	<i>t-value</i>	<i>P-value</i>
Male				
Intercept	0.9999	4.79E-05	20859.23	<0.001
Slope	1.81E-05	8.90E-06	2.04	0.0454*
Female				
Intercept	0.9998	4.10E-05	24417.46	<0.001
Slope	4.30E-05	7.67E-06	5.61	<0.001

Significance: *P < 0.05.

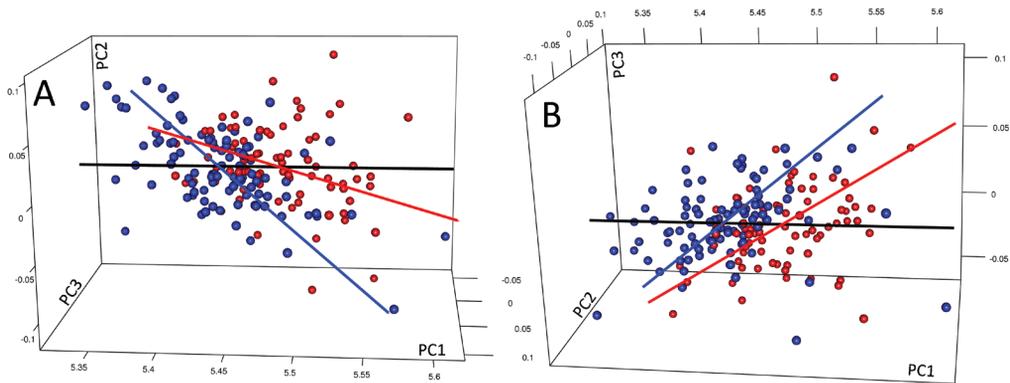


Figure 10. Three-dimensional plot of the first three principal components in form space. The red spheres represent males, the blue spheres, females. The black line indicates the direction of pure size change (Mitteroecker et al. 2004), which should be more or less parallel to the first PC. The red and blue lines represent respectively the three-dimensional male and female ontogenetic trajectories.

In view of the subtlety of the male–female average shape differences and the rather spherical nature of the depicted point clouds, we opted nevertheless to pool males and females for the principal component analysis, performed in shape space. Based upon Perez-Neto’s ‘Random average under permutation’ stopping rule (Perez-Neto et al. 2005), the first four principal components were found to be biologically interpretable ($P < 0.001$). These are depicted in Fig. 11(a-d). Together, the four PCs account for 59.45% of the total variance in the sample, ranging from 29.56 (first PC) to 6.47% (fourth PC). Their biological interpretation is provided in the Discussion section.

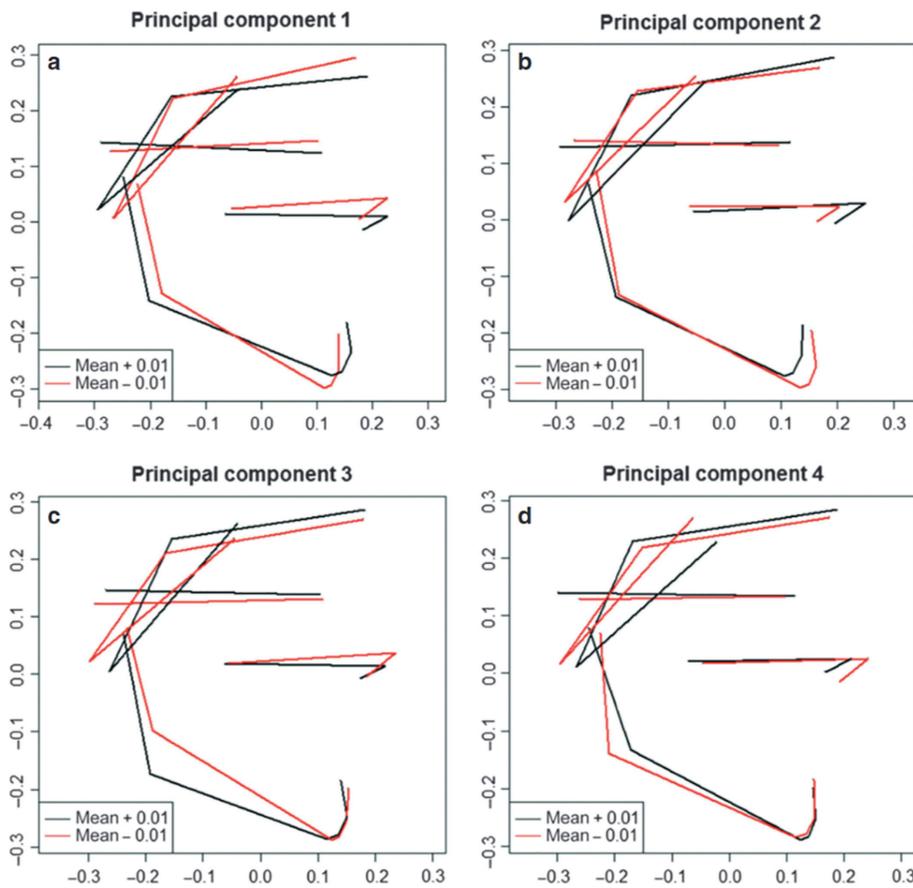


Figure 11. Visualization of the four retained biologically interpretable principal components, in shape space. The black wireframes represent the positive deformation of the male Procrustes mean shape along the PC axis. The red wireframes are the negative deformations. The percentage of variability explained by PC 1 through 4 is 29.7, 15.9, 7.68, and 6.5%, respectively.

Although there were no significant differences in mean shape between males and females, static allometry could still influence the modularity hypothesis test (Klingenberg, 2009). Since the male and female patients clearly differed in size, the modularity test was performed using the residuals of a pooled within-group regression of shape over size. The original adjacency graph, constructed using Delaunay triangulation, is visualized in Fig. 2a and the corrected adjacency graph for the three different modularity scenarios in Fig. 2b. The connections between Basion and Gonion, ANS and Pogonion, as well as between ANS and Point B were removed, and four diagonals were added: Articulare to Spheno-ethmoidale, Gonion to Subspinale, Basion to Point S and ANS to PNS. The subdivisions associated with each the three modularity scenarios are depicted in Fig. 3(a-c).

The results of testing the hypothesis of modularity are listed in Table 4. The RV coefficient for the subdivision into two modules, the skull base vs. the mandibulomaxillary complex, proved significant (Fig. 3a; the mandible and maxilla were combined into one structure for testing) ($P < 0.05$, Table 4). The same holds true for the subdivision representing the counterpart principle (Fig. 3b) ($P = 0.49$, Table 4). The modularity scenario involving four modules proved significant as well (Fig. 3c) ($P < 0.05$). The modularity hypothesis was rejected when considering the skull base, mandible and maxilla separately, as well as when combining the skull base with the maxilla, or with the mandible (Table 4).

Table 4: Results for the modularity hypothesis test using the (multi-)RV coefficient. The counterpart modularity scenario is visualized in Fig. 3b and the scenario involving four modules in Fig. 3c.

No. of modules	Skeletal parts in modules	(Multi-) RV coef.	Number of alt. subdivisions	No. of alt. subdivisions with lower (multi-)RV	P-value
2	Skb + Mx, Mnd	0.666846	265	224	0,845
2	Skb + Mnd, Mx	0.458984	61	27	0,443
2	Skb, Mx + Mnd	0.495572	265	8	0.030*
2	Counterparts	0.537199	305	15	0.049*
3	SkB, Mx, Mnd	0.445382	545	88	0,162
4	Four modules	0.365604	608	23	0.038*

SkB = Skull base, Mx = Maxilla, Mnd = Mandible
Significance: * < 0.05

Discussion

One of the aims of this study was to characterize craniofacial variation in a large, preferably unselected, sample of orthodontic patients. To sample realistically the highly variable contemporary (orthodontic patient) population, inclusion criteria need to be limited in scope and number. This might in turn lead to differences in the age and sex distribution of the experimental sample. It is important to consider the relevance of any such differences to the planned principal component analysis or the modularity hypothesis test. While the first permutation test found no statistically significant differences in the male and female Procrustes mean shapes (Fig. 4), rerunning the permutation test while controlling for the effects of allometry revealed highly significant, albeit surprisingly subtle, differences (Fig. 5a).

In view of the rather liberal inclusion criteria used and the heterogeneous nature of the resulting experimental group, the similarity between the mean configurations is striking.

The subtle nature of the male–female mean shape differences in the current study was an unexpected finding, since marked sexual dimorphism has frequently been reported, both in size and shape (Ursi et al. 1993; Humphrey, 1998; Rosas & Bastir, 2002; Bulygina et al. 2006). Bulygina et al. (2006) reported male–female size differences in the anterior part of the neurocranium already 1 year after birth (or earlier) which remained constant during growth, confirming earlier results by Ursi et al. (1993), who found the anterior cranial base to be larger in males from 6 years of age. In the early stages of growth, males tended to exhibit a more profound cranial base flexion, relatively smaller faces, and larger frontal bones (Bulygina et al. 2006). In the subsequent years this reversed, until at 6–12 years of age, the midline shapes of both sexes were very similar (Bulygina et al. 2006). The maxillary and mandibular position seemed to be dimorphic at any age, while their effective lengths did not exhibit male–female differences until about 9 years of age (Ursi et al. 1993).

With regard to adults, most authors seem to agree on the presence of facial dimorphism, mainly as a consequence of male hypermorphosis (Enlow, 1990; Ursi et al. 1993; Humphrey, 1998; Rosas & Bastir, 2002; Bulygina et al. 2006): the female growth spurt slows down at about 13 years of age (Enlow, 1990; Bulygina et al. 2006) while male pubertal growth peaks at 15 years of age (Dean et al. 2000), an age at which female growth is usually complete (Bulygina et al. 2006). Since the cranial base matures completely at about 11–12 years of age (Bastir et al. 2006), remaining craniofacial growth is spatially and functionally limited to the masticatory and facial structures (Enlow, 1990; Humphrey, 1998; Lieberman et al. 2000b, 2008; Bastir & Rosas, 2006; Gkantidis & Halazonetis, 2011).

In view of the large age range of the experimental sample, it was deemed interesting to further scrutinize craniofacial variance with regard to ontogenetic allometric differences. Apart from purely allometric shape changes occurring in the pooled sample (in the case of more or less parallel ontogenetic trajectories), the ontogenetic trajectories of males and females might also diverge (Mitteroecker et al. 2004). Both scenarios might necessitate a separate analysis of younger vs. older patients and/or males vs. females, with regard to the PCA and the modularity hypothesis test.

As suggested by Mitteroecker et al. (2004), the first three components of the data decomposition were visualized simultaneously. These were further analyzed by providing two-dimensional projections of the most relevant rotations of the 3D plot. Figure 6 shows a scatter plot of the common allometric component (CAC) vs. the first residual shape component (RSC1), illustrating ontogenetic shape changes. These components are the first and second PC, resulting from the principal component analysis, performed in shape space. The male and female trajectories are remarkably similar: if not identical, they are

virtually parallel. This would seem to indicate that within the growth period studied, developing males and females go through almost identical shape changes. These findings align with those of Viðarsdóttir et al. (2002), who found no discernible divergence in the male and female ontogenetic shape trajectories in any of the 10 populations they studied (at least for those that contained enough sexed males and females to draw this conclusion). It should be pointed out that the latter study, as well as the current one, was cross-sectional in nature, which would seem to limit the potential to pick up subtle ontogenetic shape trajectory variations. This might explain why Bulygina et al. (2006), using the longitudinal Denver Growth Study data, were able to demonstrate that the sexual ontogenetic shape trajectories are somewhat parallel until the beginning of puberty, but differed in direction thereafter.

The plot of the CAC vs. $\log(\text{centroid size})$ in Fig. 7 to visualize the allometric growth trajectories seems to confirm the common notion that males and females go through craniofacial shape changes at different rates. The female regression line is considerably steeper than the male one. Since the male regression line was found to be nonsignificant (Table 2), the four most extreme male outliers were removed and the regression line recalculated (Fig. 8). Although the male slope was now found to be significant (Table 3), the still steeper female slope confirmed that females on average tend to reach their adult size/shape earlier, whereas the male growth spurt, apart from exhibiting a delayed onset relative to females (Enlow, 1990), takes much longer to complete, especially in the mandibular and maxillary region (Mitani & Sato, 1992; Bastir et al. 2006).

Performing a linear regression of the CAC on $\log(\text{centroid size})$ was considered appropriate here, since the methodology proposed by Mitteroecker et al. (2004) specifically separates shape changes associated with allometry (CAC) from those that are not (RSCs). Even if one does not accept the use of regression lines in this scenario due to the multivariate nature of shape or the somewhat spherical nature of the point scatter, Figs 7 and 8 clearly contain evidence of allometric differences, the female scatter being located more to the top left of the graph.

Intriguingly, the CAC seems to represent a vector of almost pure size change, with little or no identifiable accompanying shape change (Fig. 9). This might again be explained by the fact that the brunt of the experimental sample fell within the age range for which Bulygina et al. (2006) demonstrated a surprising similarity in the midline cranial shape (8–12 years). Rosas & Bastir (2002) studied allometry and sexual dimorphism in two groups of 55 adult males and females. In terms of male–female shape differences, they found males to exhibit a relative forward angulation of the nasal bones, with a more pronounced glabella. The latter finding has also been reported by Bulygina et al. (2006). Furthermore, a downward rotation of the anterior nasal floor was noted, as well as a

more retro-positioned symphysis, leading to a more pronounced chin in comparison with females, who were more protrusive at the dento-alveolar level (Rosas & Bastir, 2002). Increased antegonial notching, as well as an antero-inferior displacement of the gonial angle and a more forward condylar position were also reported (Rosas & Bastir, 2002). With respect to allometric shape differences, smaller individuals exhibited a vertical decrease in the maxillary alveolar process and mandibular ramus, leading to a more retrognathic profile (Rosas & Bastir, 2002). The glabellar region moved slightly back and the occipital clivus was displaced downward, relative to large individuals (Rosas & Bastir, 2002).

As explained by Mitteroecker et al. (2004), the Procrustes form space can also be produced by performing a PCA on the matrix of shape coordinates, augmented with a column vector holding the log(centroid size) values, thus allowing visualization of the first three principal components in a 3D plot (Fig. 10a,b). In this scenario, log(centroid size) is part of the eigenanalysis, and usually largely dominates the first PC (Mitteroecker et al. 2004). This can be readily observed in Fig. 10(a,b), which in the PC1–PC2 view (top graph) shows a clearly discernible divergence in the male and female ontogenetic allometric signals, contrary to almost parallel trajectories in the PC1–PC3 view (bottom graph).

Although a bootstrap permutation test confirmed these findings, the spherical nature of the male and female point clouds and their poor separation seem to limit the conclusions that can be drawn from them. As evident from both views in Fig. 10, the most important male–female separating characteristic remains dimorphism in size. The lack of a clear correlation between the CAC and log (centroid size) in Figs 7 and 8 may be explained, in part, by the cross-sectional nature of the experimental sample. Since the various stages of craniofacial development are represented by different individuals, more variance is introduced along ages (Bulygina et al. 2006; Polanski, 2011). Additionally, log(centroid size) might be considered to be a poor proxy for development. It is therefore perfectly conceivable that small, precocious individuals as well as larger ones with a somewhat delayed craniofacial development were misclassified in Figs. 7 and 8.

Size and dental eruption staging are quite often the only measures available for estimating age in anthropological specimens; limitations that do not apply to the current experimental sample. However, several studies in modern Homo considering calendar age, hand-wrist radiographs, standing height or dental maturation have alluded to the poor correlation between the latter and craniofacial development. Additionally, it has been suggested that the growth spurt in standing height does not always coincide perfectly with that of the cranium, or that the latter growth spurt might be modular in nature, affecting some structures earlier or later than others (Mitani & Sato, 1992).

Although hand-wrist radiographs and standing height measurements were not available for the current sample, the cervical vertebrae maturation index (CVM) might have been used. This index attempts to quantify the remaining adolescent growth using the morphology of the first four cervical vertebrae, and has been suggested as an alternative to the use of hand-wrist radiographs. However, some recent publications have questioned the reproducibility of the CVM technique (Chatzigianni & Halazonetis, 2009; Fudalej & Bollen, 2010; Nestman et al. 2011).

The results for the (pooled) principal component analysis are depicted in Fig. 11(a-d). The number of biologically interpretable PCs was determined using the 'Random average under permutation' stopping rule, proposed by Peres-Neto et al. (2005). This more robust approach was preferred over the use of the '5% of explained variation' rule of thumb or the screeplot approach (Zelditch et al. 2004). The 5% rule is quite arbitrary, whereas screeplots do not necessarily exhibit a readily distinguishable abrupt change in curvature (Zelditch et al. 2004). The latter is required to make a clear-cut decision on which PCs precede this point and can therefore be regarded as being biologically pertinent.

According to the 'Random average under permutation' rule, the first four PCs are biologically interpretable ($P < 0.001$). It should be noted that although the first principal component (i.e. the vector of maximal variance) could be argued to have a clear biological justification, the subsequent ones are constrained to be orthogonal. As such, their individual biological interpretation requires some caution. The first PC (Fig. 11a) seems to deal mainly with vertical effects: hyperdivergency (in red) vs. hypodivergency (in black), with the accompanying decrease or increase in relative facial depth, respectively. Some rotation of the skull base can also be observed. The second PC (Fig. 11b) represents Class II vs. Class III skeletal patterns, with mandibular retrognathism and maxillary prognathism in black, and the exactly opposite arrangement in blue.

Equally intriguing is the skull base rotation which can be found in both PC1 and PC2, and which seems to confirm the conclusions by Kuroe et al. (2004) that skull base rotation might be more important than skull base flexure in reference to different growth patterns. Additionally, the vertical facial patterns in the first PC seem to be associated with pure skull base rotation, whereas the second PC suggests sagittal discrepancy to be correlated with the length of the anterior and posterior skull base as well, which confirms the results from Kerr & Adams (1988), and Kuroe et al. (2004). These first two PCs also compare very favorably to the results of Halazonetis (2004), who reported strikingly similar morphological associations.

For the second PC, however, he reported a relative superior/inferior positioning of the skull base, as opposed to the currently found combined skull base rotation and lengthening/shortening. The third PC (Fig. 11c) seems to pertain to the length of the mandibular ramus, represented by the vertical position of the Gonion landmark, as well as a rotation of the posterior skull base. The third PC skull base variation is somewhat more pronounced than that found by Halazonetis (2004), whereas a very slight vertical maxillary displacement in his study was found to be a slight maxillary rotation in the current one. Total relative skull base rotation around Sella as well as maxillary length seem to be the next most important areas of variation, as evidenced by the fourth PC (Fig. 11d).

The original and corrected adjacency graphs are depicted in Fig. 2a and 2b, respectively. The latter graph is missing the connections between Basion and Gonion, as well as those between ANS and Pogonion, and between ANS and Point B. These were removed since they were located 'outside of the skeletal structures of interest', as recommended by Klingenberg (2009), who also proposed removing connections between structures that do not typically interact, and adding a second diagonal to quadrilaterals, where needed (Basion to Sella, Articulare to Spheno-ethmoidale, Gonion to Point A, and PNS to ANS in Fig. 2b). Although personal considerations and/or preferences might come into play here, pilot studies indicated these modifications did not influence the result of the modularity hypothesis test; however, static allometry might (Klingenberg, 2008, 2009). As pointed out by Goshwami & Polly (2010), Procrustes analysis standardizes size but does not remove the component of shape which is correlated with size, potentially creating an appearance of complete integration, and masking modularity.

Since the second permutation test indicated that the male and female subgroups were significantly different in size, we opted to test the hypothesis of modularity using the shape coordinates resulting from a pooled within-group regression of shape over centroid size. Of the modularity scenarios involving the skull base, mandible and maxilla, either separately or combined (modularity scenarios 1, 2, 3 and 5), only the hypothesis involving the skull base vs. the mandibulomaxillary complex turned out to be significant (Table 4). In all, 265 possible alternative (continuous) subdivisions were tested, of which only eight (0.030%) exhibited a lower RV coefficient than the corresponding value of the original subdivision. Hence the latter's RV coefficient is located far enough in the left tail of distribution to be considered significant. This would seem to confirm the frequently reported poor correlation between the midline cranial base and the face (Lieberman et al. 2000b; Bastir & Rosas, 2006; Polat & Kaya, 2007; Proff et al. 2008). Indeed, Gkantidis & Halazonetis (2011) found the correlation between the midline cranial base and the face to decrease into adulthood, contrary to the lateral cranial base.

The subdivision into an anterior and posterior module (Fig. 3b) mimicking the counterpart principle (Enlow et al. 1969; Enlow, 1990; Enlow & Hans, 1996), was significant as well ($P = 0.049$, Table 4). This provides additional evidence, albeit not exceptionally strong, for the presence of an anterior and posterior craniofacial column, which could be regarded as two vertically oriented modules, each consisting of highly integrated sub-modules. The anterior column consisting of the anterior skull base, the ethmomaxillary complex and the mandibular corpus, and the posterior column of the posterior skull base and the mandibular ramus (Enlow et al. 1969; Enlow, 1990; Enlow & Hans, 1996). The presence of sub-modules within the previously outlined framework was confirmed using the sixth modularity scenario (Fig. 3c), which considered four modules ($P = 0.038$, Table 4). Interestingly, one of the submodules involves two mandibular ramal landmarks and a maxillary landmark, which could be regarded as delimiting pharyngeal space. These were previously found to predict vocal tract dimensions independently of cranial base flexion in a longitudinal sample of *Homo sapiens* (Lieberman & McCarthy, 1999). Although some evidence has been found that skull base angulation may be constrained by pharyngeal restructuring prenatally (Jefferey, 2005), the latter structure's growth peaks only after the ossification of most sphenoidal synchondroses, and is therefore correlated more strongly with mandibular and maxillary landmarks (Lieberman & McCarthy, 1998). Dento-alveolar modularity was not considered in this study.

Since this was a two-dimensional landmark-based investigation, although the anterior limits of the dento-alveolar regions could be pinpointed with relative accuracy, the posterior limits were often very hard to locate reliably due to cephalometric superimposition and/or differential enlargement of the bilateral landmarks. Also, due to the heterogeneous nature of the experimental sample, it was difficult to select landmarks which would consistently define these posterior limits of the dentition: some patients did not have erupted permanent second molars, whereas in others, the third molars were in place.

Conclusion

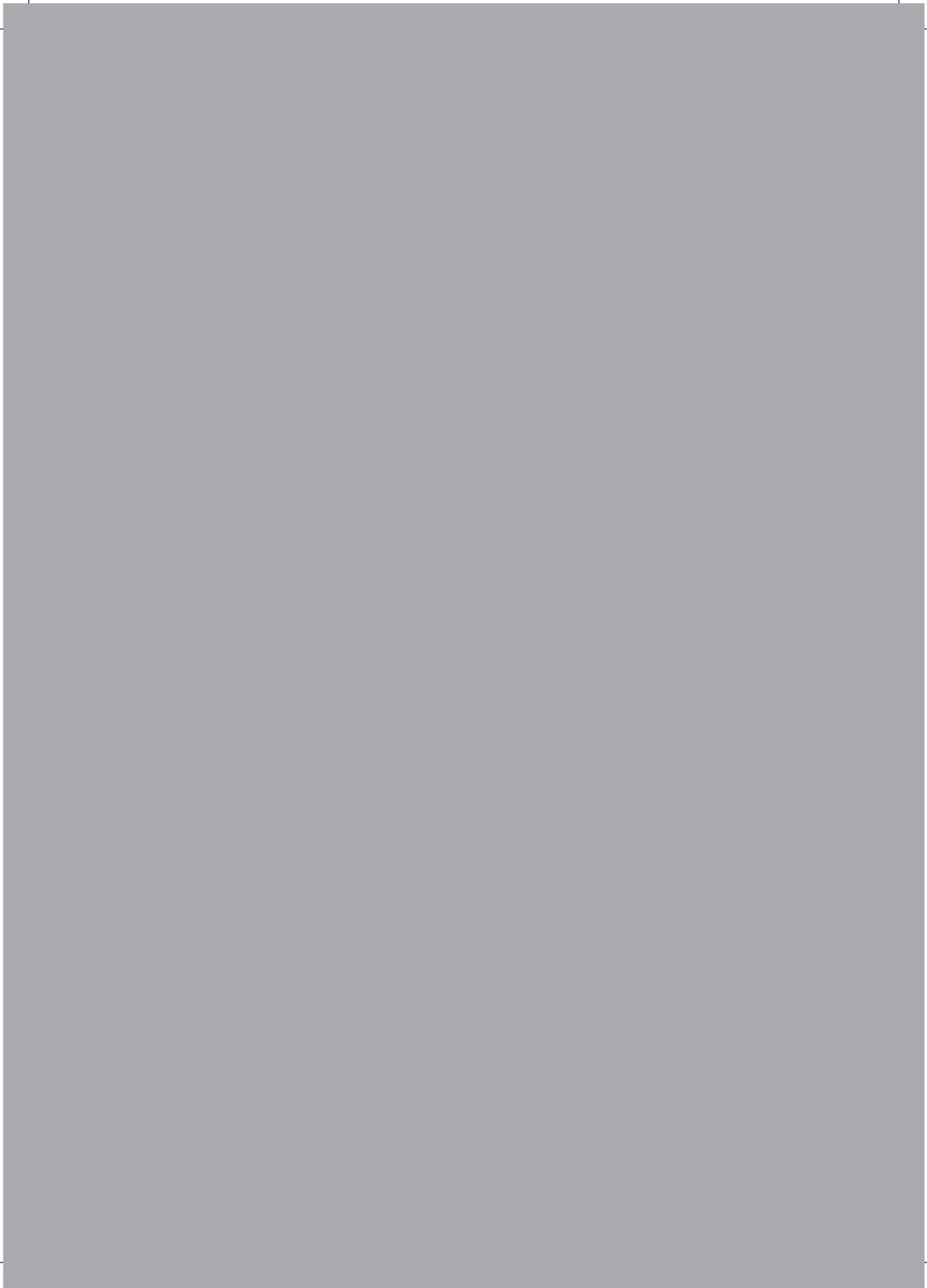
Within the age period studied and the limitations of this cross-sectional study, we found subtle but significant differences in the male and female Procrustes mean shapes. Males tended to be larger. Additionally, mild sexual ontogenetic allometric divergence was found. The principal component analysis retained four biologically interpretable components, the first two of which relate to vertical growth patterns (dolichofacial vs. brachyfacial) and sagittal skeletal relationships (maxillary prognatism vs. mandibular retrognathism, and vice versa), respectively. The mandible and maxilla were found to constitute one module, independent of the skull base. We were also able to provide evidence for the counterpart principle, in the form of an anterior and posterior module, separated by the pterygomaxillary plane, which could be further subdivided into four separate modules involving the posterior skull base, the ethmomaxillary complex, a pharyngeal module, and the anterior part of the jaws.

References

1. Arthur W (2002) The emerging conceptual framework of evolutionary developmental biology. *Nature* 415, 757–764.
2. Bastir M (2008) A systems-model for the morphological analysis of integration and modularity in human craniofacial evolution. *J Anthropol Sci* 86, 37–58.
3. Bastir M, Rosas A (2004) Facial heights: evolutionary relevance of postnatal ontogeny for facial orientation and skull morphology in humans and chimpanzees. *Hum Evol* 47, 359–381.
4. Bastir M, Rosas A (2005) Hierarchical nature of morphological integration and modularity in the human posterior face. *Am J Phys Anthropol* 128, 26–34.
5. Bastir M, Rosas A (2006) Correlated variation between the lateral basicranium and the face: a geometric morphometric study in different human groups. *Arch Oral Biol* 51, 814–824.
6. Bastir M, Rosas A, O'Higgins P (2006) Craniofacial levels and the morphological maturation of the human skull. *J Anat* 209, 637–654.
7. Bastir M, Rosas A, Stringer C, et al. (2010) Effects of brain and facial size on basicranial form in human and primate evolution. *J Hum Evol* 58, 424–431.
8. Biegert J (1957) Der Formwandel der Primatenschädels und seine Beziehungen zur ontogenetischen Entwicklung und den phylogenetischen Spezialisierungen der Kopforgane. *Gegenbaurs Morphol Jahrb* 98, 77–199.
9. Bulygina E, Mitteroecker P, Aiello L (2006) Ontogeny of facial dimorphism and patterns of individual development within one human population. *Am J Phys Anthropol* 131, 432–443.
10. Chatzigianni A, Halazonetis DJ (2009) Geometric morphometric evaluation of cervical vertebrae shape and its relationship to skeletal maturation. *Am J Orthod Dentofacial Orthop* 136, 481.e1–481.e9.
11. Dean D, Hans MG, Bookstein FL, et al. (2000) Three-dimensional Bolton-Brush Growth Study landmark data: ontogeny and sexual dimorphism of the Bolton standards cohort. *Cleft Palate Craniofac J* 37, 145–156.
12. Delaunay B (1934) Sur la sphère vide. *Izv Akad Nauk Sssr Otdelenie Khim Nauk* 7, 793–800.
13. Enlow DH (1990) *Facial Growth*, 3rd edn. Philadelphia: Saunders.
14. Enlow DH, Hans MG (1996) *Essentials of Facial Growth*. Philadelphia: W.B. Saunders Co.
15. Enlow DH, Moyers RE, Hunter WS, et al. (1969) A procedure for the analysis of intrinsic facial form and growth. An equivalent-balance concept. *Am J Orthod* 56, 6–23.
16. Fudalej P, Bollen AM (2010) Effectiveness of the cervical vertebral maturation method to predict postpeak circumpubertal growth of craniofacial structures. *Am J Orthod Dentofacial Orthop* 137, 59–65.
17. Giles WB, Phillips CL, Joondeph DR (1981) Growth in the basicranial synchondroses of adolescent *Macaca mulatta*. *Anat Rec* 199, 259–266.
18. Gkantidis N, Halazonetis DJ (2011) Morphological integration between the cranial base and the face in children and adults. *J Anat* 218, 426–438.

19. Goshwami A, Polly PD (2010) Methods for studying integration and modularity. In: Quantitative Methods in Paleobiology. (eds Alroy J, Hunt G), pp. 213–243, The Paleontological Society: The Paleontological Society Papers, 16.
20. Halazonetis DJ (2004) Morphometrics for cephalometric diagnosis. *Am J Orthod Dentofacial Orthop* 125, 571–581.
21. Humphrey LT (1998) Growth patterns in the modern human skeleton. *Am J Phys Anthropol* 105, 57–72.
22. Jefferey N (2005) Cranial base angulation and growth of the human fetal pharynx. *Anat Rec* 284A, 491–499.
23. Jeffery N, Spoor F (2002) Brain size and the human cranial base: a prenatal perspective. *Am J Phys Anthropol* 118, 324–340.
24. Kerr WJ, Adams CP (1988) Cranial base and jaw relationship. *Am J Phys Anthropol* 77, 213–220.
25. Klingenberg CP (2008) Morphological integration and developmental modularity. *Annu Rev Ecol Evol Syst* 39, 115–132.
26. Klingenberg CP (2009) Morphometric integration and modularity in configurations of landmarks: tools for evaluating a-priori hypotheses. *Evol Dev* 11, 405–421.
27. Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Mol Ecol Resour* 11, 353–357.
28. Klingenberg CP, Mebus K, Auffray JC (2003) Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible? *Evol Dev* 5, 522–531.
29. Kuroe K, Rosas A, Molleson T (2004) Variation in the cranial base orientation and facial skeleton in dry skulls sampled from three major populations. *Eur J Orthod* 26, 201–207.
30. Lieberman DE, McCarthy RC (1999) The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *J Hum Evol* 36, 487–517.
31. Lieberman DE, Ross CF, Ravosa MJ (2000a) The primate cranial base: ontogeny, function, and integration. *Am J Phys Anthropol* 31, 117–169.
32. Lieberman DE, Pearson OM, Mowbray KM (2000b) Basicranial influence on overall cranial shape. *J Hum Evol* 38, 291–315.
33. Lieberman DE, Hallgrímsson B, Liu W, et al. (2008) Spatial packing, cranial base angulation, and craniofacial shape variation in the mammalian skull: testing a new model using mice. *J Anat* 212, 720–735.
34. McCarthy RC, Lieberman DE (2001) Posterior maxillary (PM) plane and anterior cranial architecture in primates. *Anat Rec* 264, 247–260.
35. Mitani H, Sato K (1992) Comparison of mandibular growth with other variables during puberty. *Angle Orthod* 62, 217–222.
36. Mitteroecker P (2007) Evolutionary and developmental morphometrics of the hominoid cranium. Ph.D. Thesis.

37. Mitteroecker P, Bookstein F (2008) The evolutionary role of modularity and integration in the hominoid cranium. *Evolution* 62, 943–958.
38. Mitteroecker P, Gunz P, Bernhard M, et al. (2004) Comparison of cranial ontogenetic trajectories among great apes and humans. *J Hum Evol* 46, 679–697.
39. Moss ML (1962) The functional matrix. In: *Vistas in Orthodontics*. (eds Kraus B, Reidel R), pp. 85–98. Philadelphia: Lea and Febiger.
40. Moss ML (1997) The functional matrix hypothesis revisited. I. The role of mechanotransduction. *Am J Orthod Dentofacial Orthop* 112, 410–417.
41. Nestman TS, Marshall SD, Qian F, et al. (2011) Cervical vertebrae maturation method morphologic criteria: poor reproducibility. *Am J Orthod Dentofacial Orthop* 140, 182–188.
42. Peres-Neto PR, Jackson DA, Somers KM (2005) How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Comput Stat Data Anal* 49, 974–997.
43. Polanski JM (2011) Morphological integration of the modern human mandible during ontogeny. *Int J Evol Biol* 2011, 545879.
44. Polat OO, Kaya B (2007) Changes in cranial base morphology in different malocclusions. *Orthod Craniofac Res* 10, 216–221.
45. Proff P, Will F, Bokan I, et al. (2008) Cranial base features in skeletal Class III patients. *Angle Orthod* 78, 433–439.
46. Rosas A, Bastir M (2002) Thin-plate spline analysis of allometry and sexual dimorphism in the human craniofacial complex. *Am J Phys Anthropol* 117, 236–245.
47. Ross C, Henneberg M (1995) Basicranial flexion, relative brain size, and facial kyphosis in *Homo sapiens* and some fossil hominids. *Am J Phys Anthropol* 98, 575–593.
48. Ross CF, Henneberg M, Ravosa MJ, et al. (2004) Curvilinear, geometric and phylogenetic modeling of basicranial flexion: is it adaptive, is it constrained? *J Hum Evol* 46, 185–213.
49. Ursi WJ, Trotman CA, McNamara JA Jr, et al. (1993) Sexual dimorphism in normal craniofacial growth. *Angle Orthod* 63, 47–56.
50. Viðarsdóttir US, O’Higgins P, Stringer C (2002) A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton. *J Anat* 201, 211–229.
51. Zelditch ML, Swiderski DL, Sheets DH, et al. (2004) *Geometric Morphometrics for Biologists. A Primer*. Oxford: Elsevier.



CHAPTER 4

Connecting the new with the old: modifying the combined application of Procrustes superimposition and principal component analysis, to allow for comparison with traditional lateral cephalometric variables

*This chapter is based on:
Wellens HL, Kuijpers-Jagtman AM.
Eur J Orthod. 2016;38(6):569-576.*



Abstract

Objective: The combination of generalized Procrustes superimposition (GPS) and principal component analysis (PCA) has been hypothesized to solve some of the problems plaguing traditional cephalometry. This study demonstrates how to establish the currently unclear relationship between the shape space defined by the first two principal components to the ANB angle, Wits appraisal, and GoGnSN angle, and to elucidate possible clinical applications thereof.

Methods: Digitized landmarks of 200 lateral cephalograms were subjected to GPS and PCA, after which the sample mean shape was deformed along/parallel to principal components (PC) 1 and 2, recording the ANB, Wits, and GoGnSN value at each location. Trajectories were then calculated through the PC1-PC2 space connecting locations with the same values. These were finally utilized to renormalize the PC1-PC2 space.

Results: The trajectories for the Wits appraisal were almost straight and parallel to PC1. Those for the ANB angle were angled approximately 20 degrees downward relative to PC1, with a more accentuated curvature. The GoGnSN curves were mildly angled relative to the PC2 axis, their curvature increasing slightly with increasing PC1 scores. By combining the aforementioned trajectories, it was possible to delineate the region of the PC1-PC2 shape space which would be regarded as normodivergent and skeletal Class I in traditional cephalometry. Geometric distortion could be avoided by assigning patients the ANB, Wits, or GoGnSN value of the sample mean shape, deformed to the patient's position within the PC1-PC2 plot.

Conclusion: The methodology successfully relates the shape space resulting from the GPS-PCA results with traditional cephalometric variables.

Introduction

Several recent studies have found the impact of cephalometrics on clinical practice to be rather limited (1–7). In fact, a 2013 meta-analysis by Rischen et al. (7) stated that ‘cephalograms are not routinely needed for treatment planning in Class II malocclusions’; conclusions mirrored in a similar study by Durão et al. (6) who, in view of the very small number of high quality cephalometric studies meeting their (stringent) inclusions criteria, concluded that ‘lateral cephalometric radiographs have been used without adequate scientific evidence’, and that ‘there is an urgent need to improve lateral cephalometry’s diagnostic efficiency and therapeutic efficacy’ (6).

This might explain why, in order to allocate patients to study groups, many researchers opt to combine several cephalometric variables (8–10), add dental and/or facial criteria (11–14), or even forego the former completely in favor of the latter (15–20). Intriguingly, several of the aforementioned studies are randomized controlled trials, which are considered to represent the highest standard among research designs (14–20). Two-dimensional lateral cephalometry is indeed burdened with many technical problems (21, 22), such as image enlargement, blurring, and structural doubling or shrouding (22–24), while ‘geometrical distortion’ may play a role as well (22). The latter has frequently been associated with the ANB angle (25) and Wits appraisal (26–28), whereby the ANB angle has been reported susceptible to changes in the relative antero-posterior position of point N (29), relative bimaxillary protrusion or retrusion (30), and changes in midfacial height (31), allowing patients with the same mandibulomaxillary relationships to exhibit differing ANB values. The same holds true for rotations of the mandibulomaxillary complex relative to the skull base (27, 28) and vertical dento-alveolar dimensional changes (32). The Wits appraisal was found to be highly sensitive to changes in the cant of the occlusal plane (33). It therefore seems geometrical distortion is linked to the inter-individually highly variable nature of the reference landmarks and planes included in the aforementioned cephalometric analyses.

This might be clarified further by considering the following analogy: when applying the ANB angle, orthodontists essentially attempt to triangulate intermaxillary relationships, much like land surveyors do. However, whereas the latter utilize external reference points of which the location (and elevation) is known exactly (e.g. ‘benchmarks’) (34), orthodontists implicitly assume that the patient’s reference structures are ‘located normally enough’ to ensure the validity of the measurements performed. The aforementioned analogy would also suggest there is little merit in attempting to solve lateral cephalometry’s problems by ‘moving around the surveyor’s tripod’ to a different patient-specific reference point or plane. In view of the aforementioned problems, some authors proposed applying Procrustes superimposition and principal component analysis

(PCA) to lateral cephalometry (35, 36). Generalized Procrustes superimposition (GPS) (37–40) is an iterative mathematical algorithm involving the subsequent centering, scaling, and rotation of digitized landmark configurations, minimizing the distance between corresponding landmarks using the least squares criterion (Figure 1).

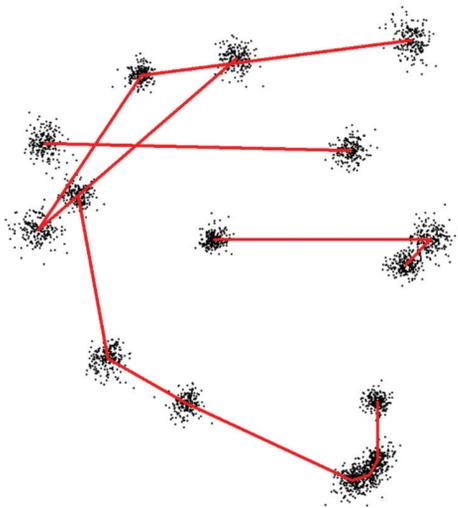


Figure 1. Generalized Procrustes superimposition of 16 skeletal landmarks ($n = 200$). GPS involves centering, scaling, and rotating the configurations in order to minimize the distance between the corresponding landmarks (using the least squares criterion), thus allowing for the assessment of shape. The sample mean shape is shown in red.

As a result, GPS allows for the assessment of shape differences (40–42), whereas PCA uncovers the directions (in multidimensional space) in which the superimposed configurations vary most (39, 41, 42). When applying GPS and PCA to a set of commonly used lateral cephalometric landmarks, the resulting first and second principal components (PCs) were found to predominantly describe variation in the vertical (dolichofacial versus brachyfacial morphology) and antero-posterior dimensions (Class II versus Class III), respectively (36, 43) (Figure 2). When plotting PC1 versus PC2, the resulting graph might be construed as a map, of which each point characterizes a particular patient's vertical and horizontal skeletal makeup in terms of PC scores on the x- (PC 1) and y-axis (PC 2) (Figure 3). The higher/lower an individual's PC score, the more this patient differs morphologically from the sample mean configuration, which is located at the origin of the axis system (e.g. the more high angle, low angle, retrognathic, or prognathic this patient is). Because 'inter-patient distance' in the PC1– PC2 shape space may be utilized as a measure of morphological similarity (39, 41) (patients located closer together are more similar morphologically), the underlying distribution of the PC scores may be used to establish cut-off points for cephalometric analysis: a logical approach would be to designate those patients belonging to the central portion of the distribution of PC1 scores [e.g. PC1 mean \pm 1 standard deviation (SD)] as being normodivergent, and of the PC2 scores as being skeletal Class I.

One of the main advantages of this population-driven approach is that the principal components it is based upon are derived from the co-ordinate data directly, irrespective of orthodontic preferences (or biases). This however implies that the first and second PCs might therefore include shape variance that is not related to the established (albeit ill-defined) orthodontic concepts to which they bear a striking resemblance: vertical growth pattern (PC1) and sagittal discrepancy (PC2) (Figure 2). Furthermore, the loss of a clear link between the GPS-PCA approach and the more familiar traditional cephalometric measures might be construed as a disadvantage: it is unclear how, if at all, both approaches might be related to one another. It might be beneficial if the advantages of the GPS-PCA approach could somehow be combined with the familiarity and utility of the traditional cephalometric measures, without reintroducing geometric distortion.

The aims of this proof-of-concept study therefore were to demonstrate a methodology for relating the shape space defined by the first two principal components to the 'traditional' cephalometric concepts of sagittal discrepancy (represented by the ANB angle and Wits appraisal) and vertical growth pattern (represented by the GoGnSN angle) while avoiding geometric distortion, and to illustrate how the currently proposed methodology might find clinical application.

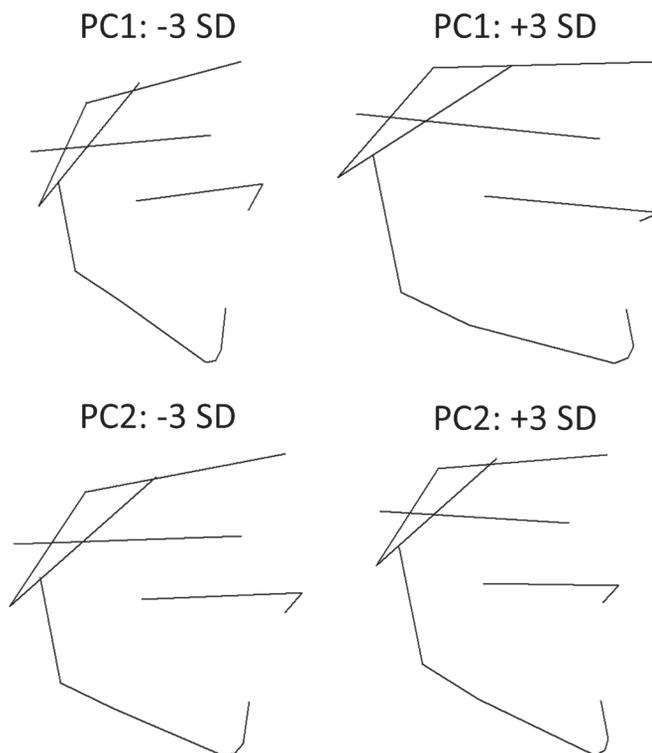


Figure 2. Deformation of the sample mean shape (depicted in red in Figure 1) of ± 3 SD along the first (upper left and right panes) and second principal components (PCs) (lower left and right panes). These represent 'directions of greatest shape variation' (in decreasing order), along which the sample mean shape may be deformed for visualization purposes. The first PC represents dolichofacial versus brachyfacial morphology, whereas the second one characterizes retrognathism versus prognathism.

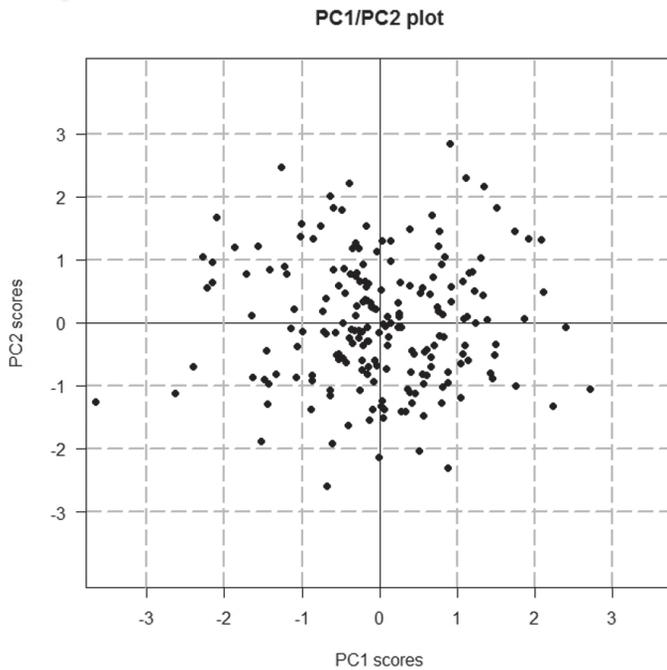


Figure 3. Plot of the first and second PCs (x-axis and y-axis, respectively), resulting from the PCA. The dots indicate where this sample's subjects are located within the PC1–PC2 space, whereby perpendicular projections onto the x- and y-axis represent the corresponding PC1 and PC2 scores, respectively. The sample mean shape is located at the origin of the axis system, while the dashed grey lines indicate the standard deviations. Referring to Figure 2, low PC1 scores (i.e. patients located more to the left in the plot) indicate a high angle growth pattern, whereas the reverse is true for high scores (i.e. brachyfacial morphology). Patients located higher in the PC1–PC2 plot (i.e. exhibiting higher PC2 scores) are more prognathic compared to patients with lower scores.

Methodology

Because we aimed to allocate study participants to the experimental groups based upon the underlying distribution of the PC scores, the required sample size was based on an estimation of the number of subjects present in the tails of a normally distributed sample. Because about 16 per cent of this distribution is located in each tail (more than 1 SD away from the mean), a sample size of about 200 patients was estimated to be required in order to obtain about 30 subjects in each tail. Two hundred consecutive lateral cephalometric radiographs (107 males, mean age: 12.8 years, SD: 2.2, range: 7.4–19.1; 93 females, mean age: 13.2 years, SD: 1.7, range 8.3–19.6) were therefore collected from the private practice of the first author (Table 1). The following inclusion criteria were applied: only pre-treatment radiographs, absence of craniofacial syndromes, only Caucasian patients, only radiographs taken in occlusion, and absence of gross movement artifacts. Patients had to be between 7 and 20 years old to be included in the sample. All images were collected using a Planmeca Proline XC (Planmeca Oy, Helsinki, Finland) by the first author, using appropriate settings and a standardized technique. The radiographs were then loaded in Viewbox (dHal software version 4.0.1.7, Kifissia, Greece), in order to identify the position of 16 skeletal landmarks (Figure 4). Cephalometric enlargement was compensated for during the digitizing process. The obtained coordinates were then exported to R (<http://www.r-project.org>) for further processing. The digitized skeletal

coordinates of the pooled sample were superimposed using GPS (Figure 1) (39, 41, 44), and stereometrically projected onto tangent space (39, 45), after which the sample mean shape was calculated. The GPS superimposed and projected landmark coordinates were then subjected to PCA (42, 46), rendering the principal component scores and their standard deviations (Figure 3).

In order to establish a relationship between the first two principal components on the one hand, and vertical growth pattern/sagittal discrepancy on the other hand, the sample mean shape was deformed from -3.5 to 3.5 SD, perpendicular or parallel to the x- (PC1) and y-axis (PC2) in 101×101 steps, recording the deformed sample mean shape's ANB angle, Wits appraisal (representing measures of sagittal discrepancy), and GoGnSN angle (as a proxy for sagittal discrepancy) in the process. The PC1–PC2 space (Figure 3) was thus sampled in terms of the three aforementioned traditional cephalometric measures at 10.201 discrete positions. Because the occlusal plane landmarks were excluded from the Procrustes analysis, their position in the deformed sample mean shape had to be extrapolated, by deforming the sample mean shape resulting from a GPS with occlusal plane landmarks to the same position in the PC1–PC2 space, and performing a thin plate spline (TPS) deformation (47) of the deformed configuration with occlusal plane landmarks on the one without them.

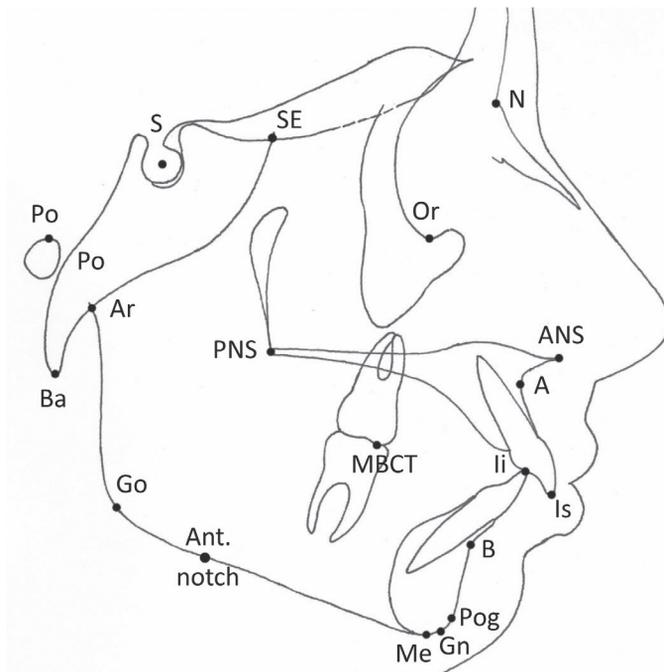


Figure 4. Digitized landmarks.

We then calculated trajectories through the PC1–PC2 space which would compensate for any observed changes in the Wits appraisal and ANB angle accompanying changing PC1 scores. These were determined by moving along the y-axis in 101 vertical steps between 3.5 and –3.5 SD PC2. At each vertical position, the space was sampled horizontally in 101 steps between –3.5 and 3.5 SD PC1. At each horizontal position, the corresponding deformed sample mean shape's ANB and Wits appraisal values were calculated. The latter two values were then compared to those exhibited by the deformed sample mean shape located at the same vertical level on the y-axis (e.g. same PC2 score). In case of differing values, we then calculated the vertical offset required to match the Wits appraisal or ANB angle of the deformed sample mean shape on the y-axis. The resulting trajectories therefore connected configurations exhibiting the same ANB or Wits values, but with varying GoGnSN angles.

Using the same methodology, similar trajectories were calculated to compensate for any changes in GoGnSN values associated with changing PC2 scores. Finally, we attempted to renormalize the PC1–PC2 space in order to obtain straight, orthogonal trajectories in the renormalized GoGnSN-ANB angle or GoGnSN-Wits appraisal space. In order to accomplish this, a 11×11 transformation matrix was defined, based upon the intersections of the compensation curves calculated earlier. Because the corresponding non-compensated coordinates were known as well, we were able to perform a TPS deformation (47), calculating the new patient coordinates based upon the matrices of compensated and non-compensated positions within the PC1–PC2 space.

Results

The sample's demographic data is presented in Tables 1 and 2. As evident from the heat map of the PC1–PC2 space in Figure 5, for the same value of PC2, low PC1 values (i.e. dolichofacial morphology) were associated with higher ANB values (yellow color), and higher PC1 values (i.e. brachyfacial morphology) with lower ANB values (in red). This can clearly be observed in Supplementary Animation 1. For the same PC2 value, high PC scores (prognathic morphology) were associated with smaller GoGnSN values, whereas lower scores (e.g. retrognathic shape) were linked to larger values (Supplementary Animation 2).

The diagonal curves in Figure 6A connect configurations exhibiting identical ANB values: when moving along them from left to right, the resulting deformed sample mean shape's GoGnSN angle changes from high to low, without concomitant change in the ANB value (Supplementary Animation 3). Contrary to the diagonal curves in Figure 6A, the corresponding trajectories for the Wits appraisal were found to be almost straight, and

virtually parallel to the PC1 axis (Figure 6B, Supplementary Animation 4). The ‘vertical’ curves in Figure 6C connect configurations exhibiting identical GoGnSN values: moving along these from top to bottom, the resulting configurations morph from Class III to Class II, with no accompanying change in the GoGnSN angle (Supplementary Animation 5). The thicker black curves in the three aforementioned figures delineate the regions of the PC1–PC2 plot which would be regarded as skeletal Class I (Figure 6A and 6B) and mesofacial (Figure 6C), respectively based upon the distribution of the underlying PC scores (PC score mean \pm 1 SD).

Figure 7A and 7B illustrate the TPS deformation grid (in red), constructed in order to renormalize the PC1–PC2 space. The corresponding original PC1–PC2 co-ordinate grid is depicted in grey. Supplementary Animation 6 provides a visual representation of the outer boundaries of the TPS deformation grid for the GoGnSNWits space. The aim of this renormalization was to obtain straight, orthogonal trajectories in the resulting, GoGnSN-Xdiff space.

The recalculated, post-TPS patient scores are depicted in Figure 8. Please note that the renormalized axes are no longer expressed in PC scores, but in the corresponding GoGnSN and ANB or Wits standard deviation values. Patients located on the same horizontal level exhibit very similar (but not necessarily identical, see below) ANB or Wits values, whereas patients aligned vertically will have very similar GoGnSN values. As was to be expected, the renormalized co-ordinate system is no longer oriented along the direction of maximum variation.

Table 1: Age distribution of the sample (n = 200).

	n	mean	sd	min	max
Male	107	12.8	2.2	7.4	19.1
Female	93	13.2	1.7	8.3	19.6
Pooled	200	13.0	2.0	7.4	19.6

Table 2: Distribution of the pooled sample’s ANB angle, Wits appraisal, and GoGnSN angle (n = 200).

	Mean	SD	Min	Max
ANB (°)	4.0	2.1	-2.8	8.6
Wits (mm)	3.0	3.3	-6.4	14.0
GoGnSN (°)	31.3	5.6	17.3	47.8

Figure 5. (RIGHT) Heat map of the PC1–PC2 space. The colours represent the ANB values found at 10.201 discrete positions within this space. Yellow and red indicate higher and lower ANB values, respectively.

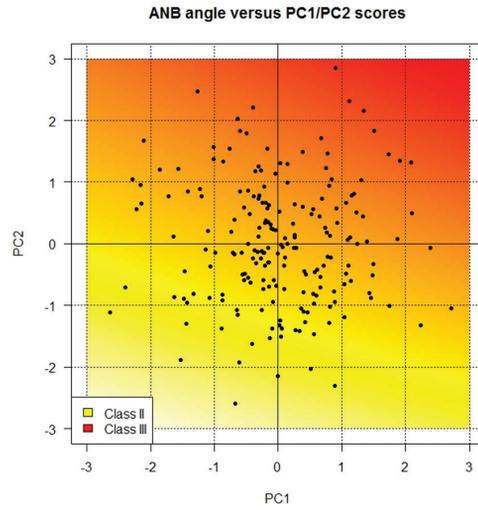
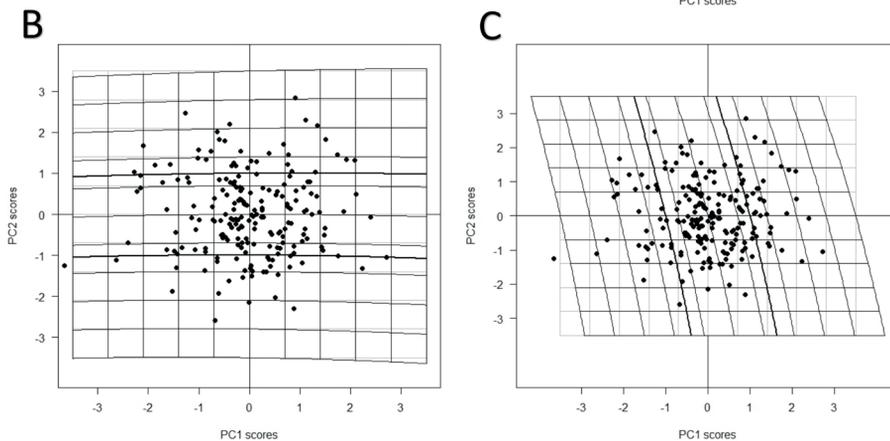


Figure 6. (A) PC1–PC2 plot, depicting the calculated trajectories within the PC1–PC2 space connecting locations with the same ANB measurements. Configurations located on the same trajectory therefore share the same ANB angle, albeit with differing degrees of facial divergence. The thick diagonals delineate the region of the PC1–PC2 plot which would be regarded as skeletal Class I (according to the ANB angle), based upon the distribution of the underlying PC scores (PC score mean \pm 1 SD). All curves are angled downward approximately 20 degrees relative to the PC1 axis. **(B)** PC1–PC2 plot, depicting the calculated trajectories connecting locations with the same Wits appraisal values. The thick diagonals delineate the region of the PC1–PC2 plot which would be regarded as skeletal Class I according to the Wits appraisal, based upon the distribution of the underlying PC scores (PC score mean \pm 1 SD). All curves are parallel to the PC1 axis and only very slightly curved. **(C)** PC1–PC2 plot, depicting the calculated trajectories connecting locations with the same GoGnSn measurements. The thick diagonals delineate the region of the PC1–PC2 plot which would be regarded as normodivergent, based upon the distribution of the underlying PC scores (PC score mean \pm 1 SD). The calculated trajectories are mildly angled relative to the PC2 axis, their curvature increasing slightly with increasing PC1 scores.



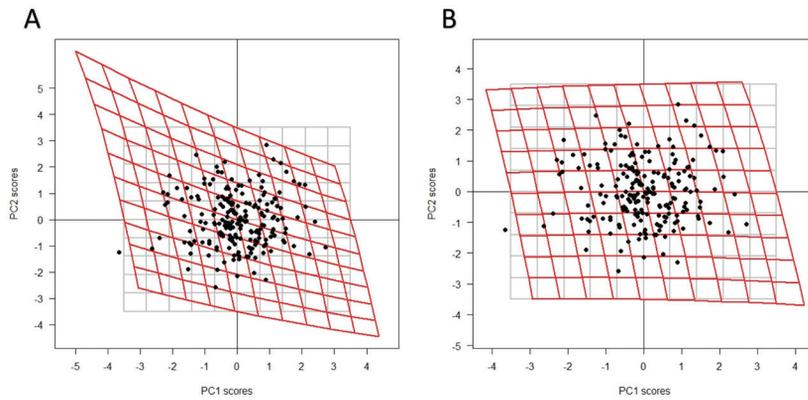


Figure 7. (A) Thin plate spline deformation grid (in red), constructed in order to renormalize the PC1–PC2 space in terms of the GoGnSN and ANB angle. The corresponding original PC1–PC2 coordinates are depicted in grey. TPS calculates a smooth deformation from the red towards the grey grid, and interpolates the resulting patient positions in the process. (B) Thin plate spline deformation grid (in red), constructed in order to renormalize the PC1–PC2 space in terms of the GoGnSN angle and Wits appraisal. The corresponding original PC1–PC2 coordinates are depicted in grey.

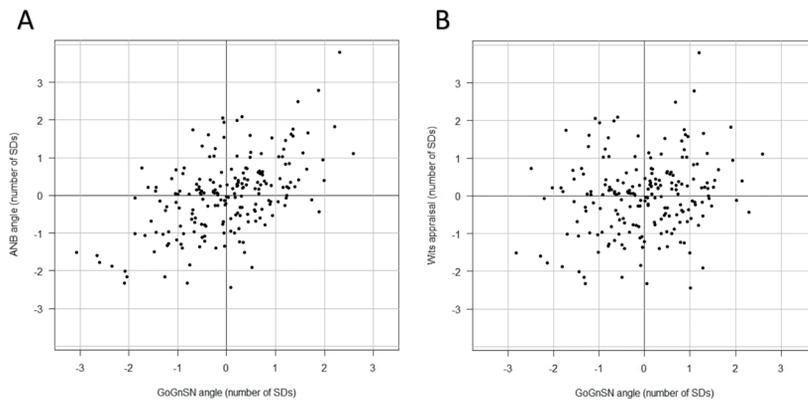


Figure 8. (A) Post-TPS patient coordinates. The new x and y-axes represents the GoGnSN and ANB angle values, respectively (in SD). (B) Post-TPS patient coordinates. The new x and y-axes represent the GoGnSN angle and Wits appraisal values, respectively (in SD).

Discussion

Contemporary cephalometric variables describe separate, discontinuous aspects of craniofacial morphology and typically do not lend themselves to straightforward graphical representation (21). This leaves the orthodontist with the intellectually challenging task of re-integrating the different variables into a cohesive mental picture, subsequent diagnosis and final treatment plan; a task which may be compounded further by diagnostic confusion, if the application of different cephalometric variables to the same morphological trait leads to differing or even contradictory diagnoses (8, 22, 48). This can clearly be observed in Figure 9, where the blue dots represent the present sample's patients for whom the diagnoses according to the ANB angle and Wits appraisal disagree (i.e. Class I/II or I/III, 47 patients), whereas the red dots represent patients with contradictory diagnostic outcomes (i.e. Class II/III, one patient).

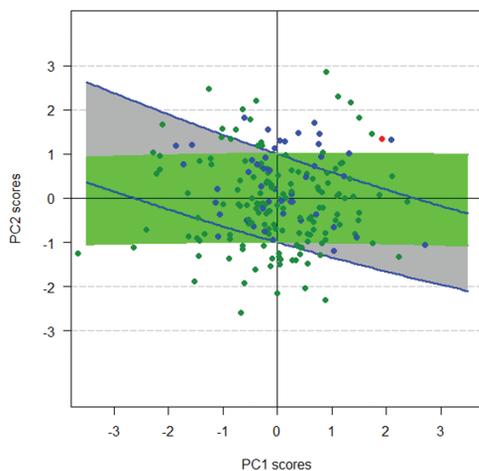


Figure 9. PC1–PC2 plot, depicting the concordance between the diagnoses according to the (traditional) ANB angle and Wits appraisal. Diagnostic agreement is indicated by a green dot (Class I/Class I), disagreement in blue (Class I/II, Class I/III), and contradiction in red (Class II/III). The green area represents the region of the PC1–PC2 plot which would be regarded as Class I according to the Wits appraisal. The corresponding region for the ANB angle is located in between the two blue curves. The grey areas depict the regions of the PC1–PC2 plot where patients should ideally be diagnosed differently without exception (Class I/II or Class I/III).

The combination of GPS and PCA offers a potential solution to this predicament, because the underlying distribution of the resulting PC2 scores allows for objective patient classification based upon their mandibulomaxillary relationships (Figures 2 and 3 and Supplementary Animation 2). The same holds true for the PC1 scores and vertical growth pattern (Figures 2 and 3 and Supplementary Animation 3), although some reservations apply (discussed further in the text). Our aim was to assess what might be learned, both clinically and theoretically, from establishing a relationship between the 'new' GPS/PCA approach and 'traditional' cephalometric variables, preferably without reintroducing geometric distortion in the process.

This was accomplished by deforming the sample mean shape to a large number of pre-determined positions within the PC1–PC2 space, and measuring the deformed sample mean shape's ANB angle, Wits appraisal, and GoGnSN angle in the process. The resulting PC1–PC2 plot may be used as an objective tool for comparing the performance of these cephalometric measures: Figure 9 depicts those regions of the PC1–PC2 plot which would be regarded as Class I according to the ANB angle and the Wits appraisal (determined by calculating trajectories within the PC1–PC2 plot connecting configurations with the same ANB angle and Wits appraisal and GoGnSN angle; Figure 6A–6C). Intriguingly, the 'Class I band' for the ANB angle (mean \pm 1 SD, in between the blue curves in Figure 9) was found to be rotated clockwise relative to that for the Wits appraisal (shaded in green, Figure 9), resulting in areas within the PC1–PC2 plot where patients should ideally be diagnosed differently without exception (e.g. Class I/II or Class I/III, areas shaded in grey in Figure 9). In other words, Figure 9 suggests that diagnostic confusion is unavoidable if both ANB and Wits are used for assessing mandibulomaxillary relationships (irrespective of which cut-off points are used), because they measure different morphological traits.

The fairly uniform distribution with which instances of diagnostic confusion seems to occur over the PC1–PC2 plot (Figure 9) might be explained by geometric distortion: individual variations in the location of the measurement's reference landmarks and planes tend to distort the corresponding cephalometric value (29–33). A potential approach to preventing geometric distortion might therefore be to not measure these variables directly, but instead assign patients the corresponding value of the sample mean shape, deformed to the patient's position within the PC1–PC2 plot (measurement 'by proxy'), thus applying the same 'ruler' to all patients. As such, it represents a generalization of the approach proposed earlier by Wellens (22). If this measurement methodology is adopted, Figure 6A and 6B suggest that the 'Wits analysis by proxy' might be a more useful measure than the corresponding ANB angle: Because the trajectories connecting configurations with the same Wits appraisal are almost straight, and virtually parallel to the PC1 axis (Figure 6B), it would seem the 'Wits analysis by proxy' is better aligned with the directions of greatest variation, compared to the corresponding ANB measurement.

As evident from Figure 7A and 7B, it is possible to renormalize the PC1–PC2 plot using a TPS deformation, such that the compensation lines and curves from Figure 5A and 5B straighten out within the newly defined co-ordinate system. The latter is of course a largely cosmetic operation, although it may facilitate the visual appraisal of the interpatient relationships in terms of the traditional cephalometric values. To insure the PCA only reflected skeletal variational patterns, we opted not to include the highly variable occlusal plane landmarks in the GPS. The position of these landmarks in the deformed sample mean shape therefore had to be interpolated using TPS deformation. Because pilot studies confirmed the reliability of this procedure, we felt it could be safely

adopted. Furthermore, it is always possible to apply the calculated transformations to the complete configurations (including dental landmarks), in order to visualize the occlusal plane/incisors in the resulting (skeletal) Procrustes superimposition.

Another possible criticism is that (currently) the methodology only takes into account the first two principal components. Although the remaining PCs indeed have an increasingly smaller influence on the patient's craniofacial shape (41), it may not necessarily be totally negligible. This would appear to pertain mostly to the GoGnSN angle, because the third and fourth PC seem to influence mainly the gonial angle, and will be investigated further in a follow-up investigation.

Another possible critique pertains to the apparent complexity of the methodology, which should nevertheless be relatively straightforward to implement in clinical practice: upon digitizing the lateral cephalogram in a predetermined order, a computer program would use pre-supplied data (the reference sample's post-GPS coordinates, matrix of principal components, and the 'by proxy' ANB, Wits, and GoGnSN 'normal zones') to calculate the patient's principal component scores, visualize his/her position in the PC1–PC2 plot, and calculate the accompanying vertical and sagittal 'by proxy' measurements directly, without having to repeat all calculations mentioned in the methodology. Apart from supplying 'distortion free' cephalometric values, the software might also provide surgical visual treatment objectives: by perpendicularly projecting a patient's position in the PC1–PC2 plot upon the x-axis in Figure 3, the corresponding configuration at that location may be calculated. This corresponds to the craniofacial shape the corresponding patient would exhibit if his/her first PC score were average (e.g. mesofacial). A similar approach may be used for projections on the y-axis.

Conclusion

The proposed methodology demonstrates how to establish the relationship between the first two principal components resulting from GPS/PCA, and conventional cephalometric variables such as the ANB angle, Wits appraisal, and GoGnSN angle. It also suggests how the latter measurements may be performed within this space without re-introducing geometric distortion, by assigning patients the corresponding value exhibited by the sample mean shape deformed to the patients position within the PC1–PC2 plot (measurement 'by proxy'). The Wits 'by proxy' measurements were found to be better aligned with the directions of maximum variation, compared to the corresponding ones for ANB angle.

Supplementary material

Supplementary material is available at European Journal of Orthodontics online. This includes movies visualizing the shape changes associated with movement from one position in the PC1-PC2 map to the next.

<https://academic-oup-com.ru.idm.oclc.org/ejo/article/38/6/569/2739002?searchresult=1#supplementary-data>

Acknowledgements

We would like to sincerely thank Professors Demetrios Halazonetis and Ellen A. BeGole for their extremely valuable advice, Professor Fred Bookstein for proofreading the manuscript, and the reviewers for their helpful comments.

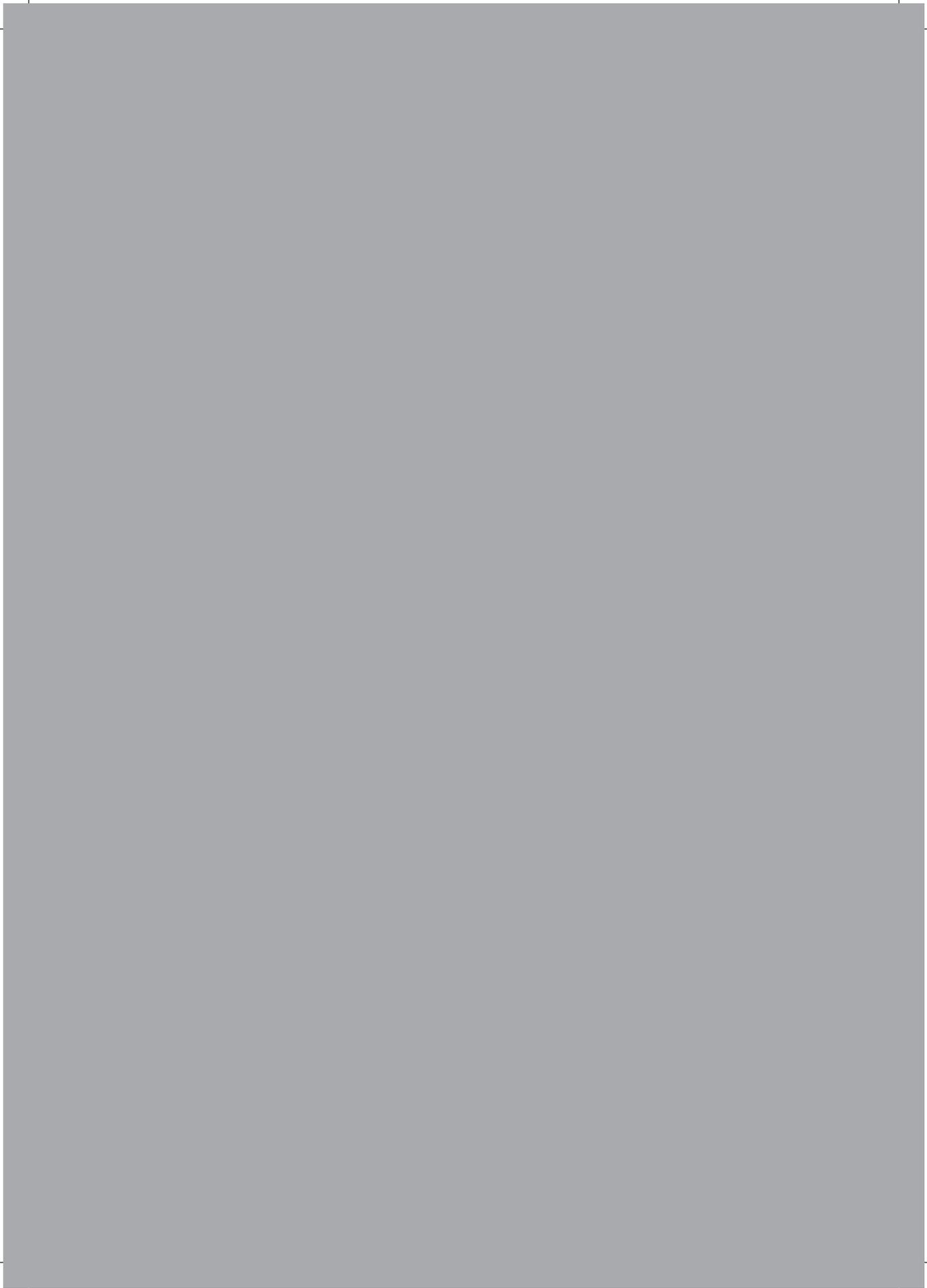
References

1. Atchison, K.A., Luke, L.S. and White, S.C. (1991) Contribution of pretreatment radiographs to orthodontists' decision making. *Oral Surgery, Oral Medicine, and Oral Pathology*, 71, 238–245.
2. Han, U.K., Vig, K.W.L., Weintraub, J.A., Vig, P.S. and Kowalski, C.J. (1991) Consistency of orthodontic treatment decisions relative to diagnostic records. *American Journal of Orthodontics and Dentofacial Orthopedics*, 100, 212–219.
3. Hansen, K. and Bondemark, L. (2001) The influence of lateral head radiographs in orthodontic diagnosis and treatment planning. *European Journal of Orthodontics*, 23, 452–453.
4. Nijkamp, P.G., Habets, L.L., Aartman, I.H. and Zentner A. (2008) The influence of cephalometrics on orthodontic treatment planning. *European Journal of Orthodontics*, 30, 630–635.
5. Devereux, L., Moles, D., Cunningham, S.J. and McKnight, M. (2011) How important are lateral cephalometric radiographs in orthodontic treatment planning? *American Journal of Orthodontics and Dentofacial Orthopedics*, 139, e175–e181.
6. Durão, A.R., Alqerban, A., Ferreira, A.P. and Jacobs, R. (2015) Influence of lateral cephalometric radiography in orthodontic diagnosis and treatment planning. *The Angle Orthodontist*, 85, 206–210.
7. Rischen, R.J., Breuning, K.H., Bronkhorst, E.M. and Kuijpers-Jagtman, A.M (2013) Records needed for orthodontic diagnosis and treatment planning: a systematic review. *PLoS ONE*, 8, e74186.
8. Kim, Y. (1978) Anteroposterior dysplasia indicator—adjunct to cephalometric differential-diagnosis. *American Journal of Orthodontics and Dentofacial Orthopedics*, 73, 619–633.
9. Ishikawa, H., Nakamura, S., Iwasaki, H. and Kitazawa, S. (2000) Seven parameters describing anteroposterior jaw relationships: postpubertal prediction accuracy and interchangeability. *American Journal of Orthodontics and Dentofacial Orthopedics*, 117, 714–720.
10. Bingmer, M., Özkan, V., Jo, J.M., Lee, K.J., Baik, H.S. and Schneider, G. (2010) A new concept for the cephalometric evaluation of craniofacial patterns (multiharmony). *European Journal of Orthodontics*, 32, 645–654.
11. Ge, Y.S., Liu, J., Chen, L., Han, J.L. and Guo, X. (2012) Dentofacial effects of two facemask therapies for maxillary protraction. *The Angle Orthodontist*, 82, 1083–1091.
12. Moreno Uribe, L.M., Vela, K.C., Kummet, C., Dawson, D.V. and Southard, T.E. (2013) Phenotypic diversity in white adults with moderate to severe Class III malocclusion. *American Journal of Orthodontics and Dentofacial Orthopedics*, 144, 32–42.
13. Moreno Uribe, L.M., Howe, S.C., Kummet, C., Vela, K.C., Dawson, D.V. and Southard, T.E. (2014) Phenotypic diversity in white adults with moderate to severe Class II malocclusion. *American Journal of Orthodontics and Dentofacial Orthopedics*, 145, 305–316.

14. Baysal, A. and Uysal, T. (2014) Dentoskeletal effects of Twin Block and Herbst appliances in patients with Class II division 1 mandibular retrognathia. *European Journal of Orthodontics*, 36, 164–172.
15. Tulloch, J.F., Phillips, C., Koch, G. and Proffit, W.R. (1997) The effect of early intervention on skeletal pattern in Class II malocclusion: a randomized clinical trial. *American Journal of Orthodontics and Dentofacial Orthopedics*, 111, 391–400.
16. Keeling, S.D., Wheeler, T.T., King, G.J., Garvan, C.W., Cohen, D.A., Cabassa, S., McGorray, S.P. and Taylor, M.G. (1998) Anteroposterior skeletal and dental changes after early Class II treatment with bionators and headgear. *American Journal of Orthodontics and Dentofacial Orthopedics*, 113, 40–50.
17. Wheeler, T.T., McGorray, S.P., Dolce, C., Taylor, M.G. and King, G.J. (2002) Effectiveness of early treatment of Class II malocclusion. *American Journal of Orthodontics and Dentofacial Orthopedics*, 121, 9–17.
18. Tulloch, J.F.C., Proffit, W.R. and Phillips, C. (2004) Outcomes in a 2-phase randomized clinical trial of early Class II treatment. *American Journal of Orthodontics and Dentofacial Orthopedics*, 125, 657–667.
19. O'Brien, K. et al. (2009) Early treatment for Class II Division 1 malocclusion with the Twin-block appliance: a multi-center, randomized, trial. *American Journal of Orthodontics and Dentofacial Orthopedics*, 135, 573–579.
20. Dolce, C., McGorray, S.P., Brazeau, L., King, G.J. and Wheeler, T.T. (2007) Timing of Class II treatment: skeletal changes comparing 1-phase and 2-phase treatment. *American Journal of Orthodontics and Dentofacial Orthopedics*, 132, 481–489.
21. Moyers, R.E. and Bookstein, F.L. (1979) The inappropriateness of conventional cephalometrics. *American Journal of Orthodontics*, 75, 599–617.
22. Wellens, H. (2009) Improving the concordance between various anteroposterior cephalometric measurements using Procrustes analysis. *European Journal of Orthodontics*, 31, 503–515.
23. Baumrind, S. and Frantz, R.C. (1971) The reliability of head film measurements. *American Journal of Orthodontics*, 60, 111–127.
24. Baumrind, S. and Frantz, R.C. (1971) The reliability of head film measurements. *American Journal of Orthodontics*, 60, 505–517.
25. Steiner, C.C. (1953) Cephalometrics for you and me. *American Journal of Orthodontics*, 39, 729–755.
26. Jacobson, A. (1975) The “Wits” appraisal of jaw disharmony. *American Journal of Orthodontics*, 67, 125–138.
27. Jacobson, A. (1976) Application of Wits appraisal. *American Journal of Orthodontics and Dentofacial Orthopedics*, 70, 179–189.
28. Jacobson, A. (1988) Update on the Wits appraisal. *The Angle Orthodontist*, 58, 205–219.
29. Taylor, C.M. (1969) Changes in the relationship of nasion, point A, and point B and the effect upon ANB. *American Journal of Orthodontics*, 56, 143–163.

30. Freeman, R.S. (1981) Adjusting A-N-B angles to reflect the effect of maxillary position. *The Angle Orthodontist*, 51, 162–171.
31. Binder, R.E. (1979) The geometry of cephalometrics. *Journal of Clinical Orthodontics*, 13, 258–263.
32. Hussels, W. and Nanda, R.S. (1984) Analysis of factors affecting angle ANB. *American Journal of Orthodontics*, 85, 411–423.
33. Roth, R. (1982) The ‘Wits’ appraisal—its skeletal and dento-alveolar background. *European Journal of Orthodontics*, 4, 21–28.
34. Benchmark (surveying) (2015). Wikipedia, the Free Encyclopedia, February 19, 2015. [http://en.wikipedia.org/w/index.php?title=Benchmark_\(surveying\)&oldid=647927071](http://en.wikipedia.org/w/index.php?title=Benchmark_(surveying)&oldid=647927071) (15 March 2015, date last accessed).
35. McIntyre, G.T. and Mossey, P.A. (2003) Size and shape measurement in contemporary cephalometrics. *European Journal of Orthodontics*, 25, 231–242.
36. Halazonetis, D.J. (2004) Morphometrics for cephalometric diagnosis. *American Journal of Orthodontics and Dentofacial Orthopedics*, 125, 571–581.
37. Goodall, C. (1991) Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 285–339.
38. Bookstein, F.L. (1997) Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1, 25–243.
39. Dryden, I.L. and Mardia, K.V. (1998) *Statistical Shape Analysis*. Wiley, Chichester, New York.
40. Kendall, D.G. (1977) The diffusion of shape. *Advances in Applied Probability*, 9, 428.
41. Zelditch, M.L., Swiderski, D.L. and Sheets, H.D. (2012) *Geometric Morphometrics for Biologists: A Primer*. 2nd edn. Academic Press, Amsterdam, the Netherlands.
42. Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441.
43. Wellens, H.L., Kuijpers-Jagtman, A.M. and Halazonetis, D.J. (2013) Geometric morphometric analysis of craniofacial variation, ontogeny and modularity in a cross-sectional sample of modern humans. *Journal of Anatomy*, 222, 397–409.
44. Gower, J.C. and Dijksterhuis, G.B. (2004) *Procrustes Problems*. Oxford University Press, Oxford, New York.
45. Claude, J. (2008) *Morphometrics with R*. 1st edn. Springer, New York.
46. Pearson, K. (1901) LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series* 6, 2, 559–572.
47. Bookstein, F.L. (1989) Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 567–585.
48. Demisch, A., Gebauer, U. and Zila, W. (1977) Comparison of three cephalometric measurements of sagittal jaw relationship: angle ANB, Wits appraisal and AB-occlusal angle. *Transactions of the European Orthodontic Society*, 2, 269–281.

Connecting the new with the old: modifying the combined application of Procrustes superimposition and principal component analysis, to allow for comparison with traditional lateral cephalometric variables



CHAPTER 5

ROC surface assessment of the ANB angle and Wits appraisal's diagnostic performance with a statistically derived 'gold standard': does normalizing measurements have any merit?

*This chapter is based on:
Wellens HLL, BeGole EA, Kuijpers-Jagtman AM.
Eur J Orthod. 2017 1;39(4):358-364.*



Abstract

Objective: To assess the ANB angle's and Wits appraisal's diagnostic performance using an extended version of Receiver Operating Curve (ROC) analysis, which renders ROC surfaces. These were calculated for both the conventional and normalized cephalometric tests (calculated by exchanging the patient's reference landmarks with those of the Procrustes superimposed sample mean shape). The required 'gold standard' was derived statistically, by applying generalized Procrustes superimposition (GPS) and principal component analysis (PCA) to the digitized landmarks, and ordering patients based upon their PC2 scores.

Methods: Digitized landmarks of 200 lateral cephalograms (107 males, mean age: 12.8 years, SD: 2.2, 93 females, mean age: 13.2 years, SD: 1.7) were subjected to GPS and PCA. Upon calculating the conventional and normalized ANB and Wits values, ROC surfaces were constructed by varying not just the cephalometric test's cut-off value within each ROC curve, but also the gold standard cut-off value over different ROC curves in 220 steps between -2 and 2 standard deviations along PC2. The volume under the resulting ROC surfaces (VUS) served as a measure of overall diagnostic performance. The statistical significance of the volume differences was determined using permutation tests (1000 rounds, with replacement).

Results: The diagnostic performance of the conventional ANB and Wits was remarkably similar for both Class I/II (81.1 and 80.75% VUS, respectively, $P > 0.05$). Normalizing the measurements improved all VUS highly significantly (91 and 87.2 per cent, respectively, $P < 0.001$).

Conclusion: The conventional ANB and Wits do not differ in their diagnostic performance. Normalizing the measurements does seem to have some merit.

Introduction

The orthodontic diagnostic toolset conventionally comprises both a clinical and radiological investigation; the latter usually consisting of a panoramic radiograph and/or peri-apical series, as well as a lateral (and anteroposterior) cephalogram. Although most contemporary orthodontic textbooks recommend the routine use of lateral cephalometry for diagnostic purposes (1, 2), its impact on clinical practice seems to be somewhat limited (3–7). A recent meta-analysis concluded that 'cephalograms are not routinely needed for treatment planning in Class II malocclusions' (8), while another suggested 'lateral cephalometric radiographs have been used without adequate scientific evidence', and that 'there is an urgent need to improve lateral cephalometry's diagnostic efficiency and therapeutic efficacy' (9). The efficacy of diagnostic imaging was defined by Fryback and Thornbury (10) using a six level hierarchical model, the first three of which pertain to the images' technical quality (level one), diagnostic accuracy (level two), and influence on the practitioner's diagnostic thinking (level three).

Since the addition of a lateral cephalogram has been found to cause few changes to treatment plans formulated without it (3–7), lateral cephalometry seems to score low in level three. This is usually attributed to the technical problems it is fraught with, such as image enlargement and structural blurring, doubling and shrouding (11), which would seem to impact mainly level one. Level two might be influenced by difficulties associated with choosing, precisely defining and pinpointing landmarks (11), while geometrical distortion might play a role as well. The latter seems to be linked to the highly variable nature of the cephalometric reference landmarks and planes (12), thereby allowing individuals with the same cephalometric values to exhibit markedly differing intermaxillary relationships, while the opposite may hold true as well (13–17).

One solution to this predicament might be to exchange the patient's highly variable reference framework with a fixed one, by superimposing a template on the digitized patient landmarks using Procrustes superimposition, and performing the measurements from the superimposed template's reference landmarks, instead of the patient's (12) (Fig. 1). Albeit unconventional, this approach significantly improved the correlation between the 'normalized' measurements, as compared to the conventional ones (12). Correlations however do not represent a measure of diagnostic performance.

Diagnostic performance is usually determined using Receiver Operating Characteristic curve analysis (ROC); a procedure which originated in signal analysis (18), but has since found widespread application in medicine (19, 20). ROC curve analysis plots the sensitivity (or true positive ratio) versus 1-specificity (or false-positive ratio) for a full range of possible values of the diagnostic test's cutoff value. The area under the resulting

curve serves as a measure of diagnostic performance: the larger the surface area under the curve (the closer the curve approaches the upper left corner of the graph), the more powerful the test.

A test characterized by a diagonal ROC curve (from lower left to upper right) has no discriminatory power whatsoever. Anything below 60 per cent area under the curve is usually designated 'very poor', whereas between 60 and 70 per cent UAC, tests are usually classified as 'poor', between 70 and 80 as 'fair', between 80 and 90 as 'good', and anything above as 'excellent'. ROC curve analysis is however dichotomous in nature, requiring clearly discernible health states in order to provide the black-or-white diagnostic result required to determine the test's diagnostic power (21). This would seem to align poorly with the continuous spectrum of facial variation present in the orthodontic patient population (22). Also, ROC curve analysis requires a gold standard (an ideally infallible 'reference test' which provides the correct answer to the diagnostic question) (21), which until recently did not seem to be available.

McIntyre and Mossey (23), Halazonetis (24) and later Akli et al. (25) proposed adopting a geometric morphometric approach to cephalometry, based upon the combined application of Procrustes superimposition and principal component analysis to previously digitized landmark coordinates; a methodology which is used ubiquitously in biology and anthropology for the analysis of shape (26, 27). Procrustes superimposition centres, scales and rotates landmark configurations to minimize the distance between the corresponding points using the least squares criterion (26, 27) (Fig. 2a, Supplementary Animation 1), while principal component analysis finds the directions in multivariate space along which the superimposed configurations vary most, in decreasing order (26, 27) (Fig. 2b, Supplementary Animation 2). In earlier studies, the first and second principal components (PCs; i.e. the major directions of variance), were found to predominantly describe variation in the vertical (dolichofacial versus brachyfacial morphology) and anteroposterior dimensions (Class II versus Class III), respectively (22, 24, 25, 28) (Fig. 2b, Supplementary Animation 2). A plot of PC1 versus PC2 may be therefore be used as a map, characterizing a patient's horizontal and vertical skeletal makeup, while also allowing for inter-patient comparison in terms of the same variables (Fig. 2b). Additionally, the distribution of the PC scores may be used to categorize patients: a logical approach would be to designate those patients belonging to the central portion of the PC1 score distribution (e.g. PC1 mean \pm 1 SD) as being normodivergent, and those in between PC2 mean \pm 1 SD as being skeletal Class I. Two recent publications provided some tools for delineating those regions of the PC1-PC2(-PC3) shape space containing patients which would be regarded as normo-, hypo- and hyper-divergent and skeletal Class I, II and III (25, 28).

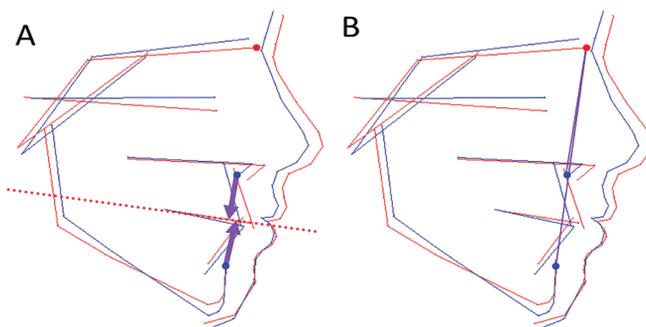


Figure 1. (a) The patient's configurations is shown in blue, the Procrustes superimposed sample mean shape in red. The normalized Wits appraisal is obtained by dropping perpendiculars from the patient's points A and B (in blue) onto the superimposed sample mean shape's occlusal plane (the dotted red line). (b) Similarly, the normalized ANB angle is obtained by measuring the angle between the patient's points A and B (in blue) and the superimposed sample mean shape's point N (in red).

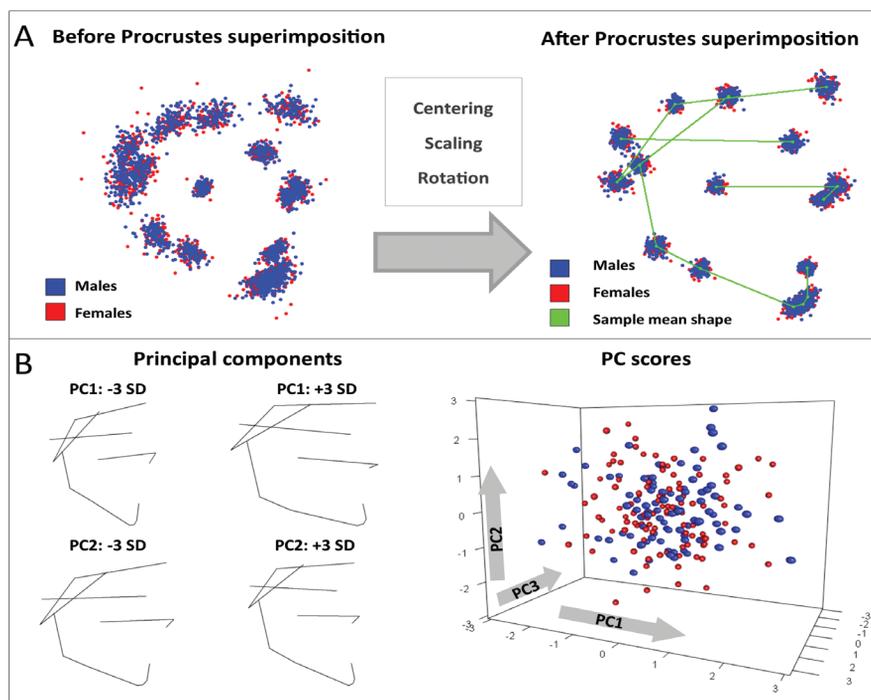


Figure 2. (a) On the left, the original coordinates of sixteen skeletal landmarks before Procrustes superimposition are shown. The right image shows the same landmarks after centring, scaling and rotating the configurations in order to minimize the squared distance between the corresponding landmarks. The sample mean shape is depicted in green.

Figure 2. (b) The first two principal components are shown on the left, by deforming the sample mean shape three standard deviations along the respective PCs. The right pane depicts the PC scores associated with the first three principal components. Each dot represents the value of a particular patient on principal components one, two and three. Since 16 landmarks were digitized, each patient thus has a 28-score long 'address' in multivariate space (4 degrees of freedom were lost to centring, scaling and rotation of the landmark configurations). Not all principal components do however represent biologically meaningful information: in this scenario, only the first five PCs are biologically 'interpretable'.

The aim of this investigation was to compare the diagnostic performance of the ANB angle and Wits appraisal using ROC analysis, whereby the 'gold standard' is derived statistically, by classifying patients based upon the distribution of the PC2 scores resulting from the combined application of Procrustes superimposition and principal component analysis (25, 28). Furthermore, we introduce in an extension of ROC analysis, whereby the gold standard cut-off is varied as well, resulting in ROC-surfaces instead of curves (29–32). Finally, we aimed to compare the diagnostic performance of the ANB angle and Wits appraisal measurements to their normalized counterparts (obtained by superimposing the sample mean shape on the patient's landmarks and measuring from the sample mean shape's reference structures).

Methodology

The methodology has been published in detail previously (28). Briefly, two hundred consecutive lateral cephalometric radiographs (107 males, mean age: 12.8 years, SD: 2.2, range: 7.4–19.1; 93 females, mean age: 13.2 years, SD: 1.7, range 8.3–19.6) were collected, using the following inclusion criteria: only pre-treatment radiographs, absence of craniofacial syndromes, only Caucasian patients, only radiographs taken in occlusion, and absence of gross movement artifacts. Patients had to be at least seven and no older than 20 years to be included in the sample.

The required sample size was calculated beforehand based on an estimation of the number of subjects present in the tails of a normally distributed sample. All images were collected using a Planmeca Proline XC (Planmeca Oy, Helsinki, Finland) by the first author, using appropriate settings and a standardized technique. The radiographs were then loaded in Viewbox (dHal software version 4.0.1.7, Kifissia, Greece), in order to identify the position of sixteen skeletal landmarks (Fig. 3). Cephalometric enlargement was compensated for during the digitizing process. The obtained coordinates were then exported to R (<http://www.r-project.org>) for further processing.

The digitized skeletal coordinates of the pooled sample were superimposed using generalized Procrustes superimposition (GPS, Fig. 2a, Supplementary Animation 1) (26, 27, 33, 34), and stereometrically projected onto tangent space (26, 27, 34), after which the male and female mean shapes were calculated. The significance of the morphological difference between them as well as their mean age difference, was tested using a 10 000 round permutation test. The GPS superimposed and projected landmark coordinates were then subjected to principal component analysis (26, 27, 34, 35), after which the principal component scores and their standard deviations were calculated (Fig. 2b,

Supplementary Animation 2). This allowed us to objectively classify patients in terms of their intermaxillary relationships based upon each patient's PC2 score.

We then calculated the corresponding (conventional) ANB angle and Wits appraisal values, as well as their normalized counterparts. The latter were determined by Procrustes-superimposing the (pooled) sample mean shape on the patient's landmarks, and measuring the ANB angle using the superimposed mean shape's point N as reference structure. Similarly, the Wits value was determined using the superimposed sample mean shape's occlusal plane, after rescaling to true size.

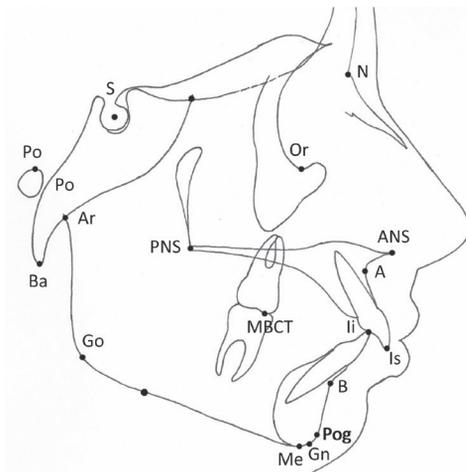


Figure 3. The sixteen digitized skeletal and dental landmarks: sella, nasion, porion, orbitale, anterior and posterior nasal spine, basion, articulare, points a and b, pogonion, gnathion, menton, antegonial notch, gonion, spenoethmoidale, mesiobuccal cusp tip of the upper first molar, upper and lower incisal edge. For their definitions, please refer to (12).

Next, ROC curves were constructed for the conventional and normalized ANB angle and Wits appraisal (ANBc/WitsC and ANBN/WitsN, respectively), by plotting sensitivity versus 1-specificity for the full range of possible cut-off values. This process was repeated 220 times while incrementally increasing the 'gold standard' cut-off value between minus two and two standard deviations on the PC2 axis. Every cycle's gold standard diagnosis was determined by comparing each patient's PC2 score to that cycle's gold standard cut-off value: patients with PC2 scores smaller than the gold standard PC2 score cut-off were designated 'Class II', and those with equal or larger PC2 scores 'Class I'. When placing the resulting 220 ROC curves side-by-side, a ROC surface was created (i.e. a three-dimensional mesh; Fig. 4, Supplementary Webpage 3), the volume under which would add up to one for a perfect test (i.e. dimensions $1 \times 1 \times 1$), and 50 per cent for an indeterminate test.

After calculating the ROC surface volumes of the classic and normalized measures, the statistical significance between them was calculated by randomly permuting the

220 corresponding ROC curves between the two measures under investigation with replacement (1000 rounds), and calculating the volume of the resulting randomly permuted ROC surfaces. Since 1000 randomized ROC surfaces were generated, 499 500 volume differences were thus calculated, the number of which exceeding the original one determined the significance of the difference. The significance level was set at 5 per cent. The digitizing error was determined in previous investigations involving the same material, and was found to be non-significant (22, 28).

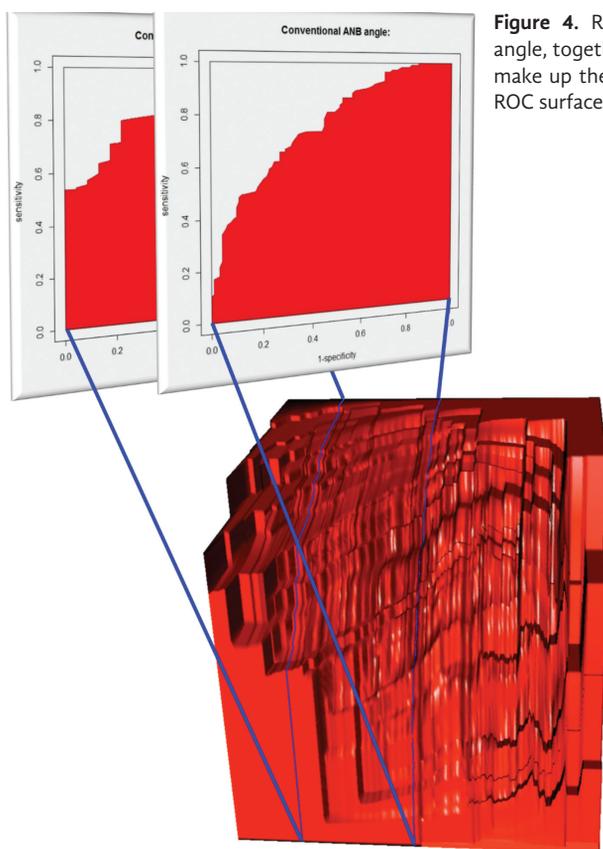


Figure 4. ROC surface for the conventional ANB angle, together with two of the 220 ROC curves that make up the ROC surface. Their position within the ROC surface is depicted in purple.

Results

Neither the male/female shape difference (Fig. 4), nor the age difference between them was statistically significant ($P = 0.1926$ and 0.1818 , respectively, 10,000 permutation rounds). Both groups were therefore pooled for further analysis. Table 1 lists the diagnostic performance of the conventional and normalized ANB and Wits, expressed as the volume under the ROC surface (expressed in percentages). The conventional ANB angle and Wits appraisal performed remarkably similar, at about 80 per cent volume under the ROC surface (Table 1) (Fig. 5). The difference between them (0.34 per cent, Table 2) was not significant ($P = 0.402$, Table 2). Normalizing the measurements increased the volumes for both Wits and ANB, although the latter improved about 10 per cent (Fig. 6), as opposed to 7 per cent for the former (WitsN: 87.15 per cent, ANBN: 91.04 per cent, Table 1). The difference between them was highly significant ($P < 0.001$, Table 2). All pairwise comparisons were highly significant, except that between the conventional ANB and Wits (Table 2).

Table 1: Volume under the surface values for the conventional and normalized ANB and Wits measurements.

	VUS
ANBC	81.092
ANBN	91.037
WitsC	80.746
WitsN	87.148

VUS: Volume under surface (percentage).

Table 2: Results of the permutation test in order to compare the VUS measurements for the conventional and normalized ANB and Wits.

Permutation test: (1000 rounds)	Original vol. diff.	Randomly permuted volume difference:				<i>P</i>
		Mean (%)	SD (%)	Max. (%)	> Orig.	
ANBc/ANBn	9.95	0.48	0.36	2.60	0	<0.001
ANBc/WitsC	0.34	0.33	0.25	1.84	201007	0.402
ANBn/WitsN	3.89	0.34	0.26	2.07	0	<0.001
ANBc/WitsN	6.06	0.42	0.31	2.26	0	<0.001
ANBn/WitsC	10.29	0.48	0.36	2.54	0	<0.001
WitsC/WitsN	6.40	0.37	0.28	2.08	0	<0.001

Orig. vol. diff. : Original volume difference (%).

>Orig. : The number of iterations in which the volume difference exceeded the original.

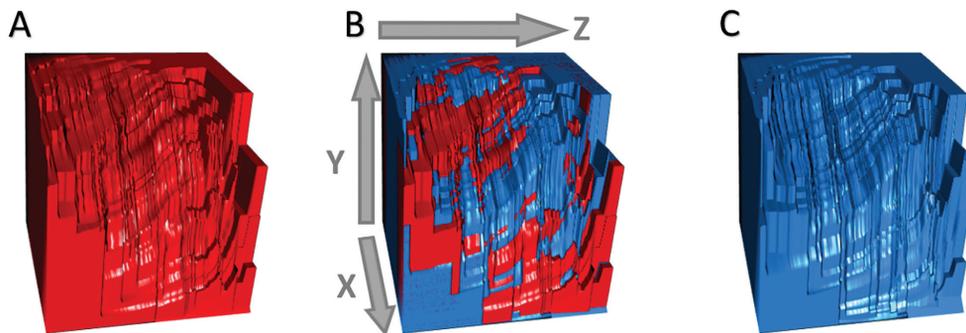


Figure 5. (a and c) Resulting ROC surfaces for the conventional ANB angle and Wits appraisal, respectively. (b) Superimposition of the ANBc and WitsC ROC surfaces. The x-axis represents 1-specificity, the y-axis the sensitivity, and the z-axis the gold-standard cut-off in between -2 and 2 SD along PC2 (220 steps). The difference between both ROC surfaces was non-significant ($P > 0.05$, Table 2).

Discussion

To date, surprisingly few studies are available about the diagnostic performance of currently available lateral cephalometric tests (8, 9). Whereas conventionally the results of newly introduced tests were often correlated to those of existing ones to assess or compare performance, more recent studies increasingly rely on ROC analysis for this purpose. In the absence of a true gold standard for cephalometric diagnosis, some authors have chosen to classify their study subjects based upon occlusion (36, 37) or existing cephalometric analyses, applied either singly (38) or combined (39). Other studies have included profile assessments (40, 41), while one study applied a Delphi approach to establish their gold standard (41). Akli et al. (25) proposed combining Procrustes superposition and principal component analysis, and using the underlying distribution of PC1, 2 and 3 scores to provide a more objective, statistically based classification methodology based upon craniofacial morphology (i.e. craniofacial shape; Fig. 2). As such, this methodology might serve as a 'gold standard' in ROC analysis for the assessment of vertical growth pattern and mandibulomaxillary discrepancy.

Traditional ROC analysis is somewhat limited by its dichotomous nature: by calculating the probability that a test will correctly identify the diseased and healthy patient in a pair, it provides the diagnostic power of a rather blunt, 'yes-or-no' test to whether disease is present, therefore applying strictly to two class problems (19–21). Although extensions to the ROC methodology to accommodate three-class and multi-class situations have recently appeared in the literature (30–32), these do not seem to apply to orthodontic cephalometry either: the absence of a 'disease state' in the true sense generally precludes the formulation of straightforward, clear-cut, and universally applicable cut-off points for the different cephalometric classes. Instead, the continuous spectrum of

craniofacial variability present in the orthodontic population begs the diagnostically more sophisticated question of whether patients are relatively more or less prognathic/retrognathic instead of just Class I, II or III, not just at all possible cut-off points of the pertaining cephalometric measure, 'but also of the gold standard' (i.e. it basically represents an 'infinite class problem').

This study therefore proposed a modification of the recently published extensions to the ROC methodology (29–32), by varying the cut-off points of the cephalometric measure under investigation 'within in each ROC curve', as well as the cut-off points of the gold standard 'over different ROC curves'. In doing so, the traditional three-class diagnostic problem (i.e. Class I, II or III) is reduced to a 2-class one (since the 'less Class II/more Class II' diagnostic question is equivalent to the 'more Class III/less Class III' question), albeit applied over a broad range of gold-standard cut-off values. When placed next to one another, the combined ROC curves generate a ROC surface, the volume under which serves as a more sophisticated measure of overall diagnostic performance.

When applying this approach to the conventional and normalized ANB angle and Wits appraisal, the conventional measurements were found to perform strikingly similar, at about 80 per cent volume under the surface ($P = 0.40$, Table 2). In an earlier publication (28), those regions of the PC1-PC2 plot which would be regarded as skeletal Class I according to both measurements were identified, and found to be distinctly different in shape, whereby the Wits appraisal's region shape was almost identical to that defined by the underlying distribution of the PC2 scores. We might therefore expect the Wits appraisal to slightly outperform the ANB angle, which was found to not be the case. Tables 1 and 2 do seem to suggest that normalizing the measurements has some merit: the volumes under the ROC surfaces of the normalized measurements were found to be significantly higher compared to those of the conventional measurements. The observed improvements (Table 2) lend credence to the hypothesis that cephalometric diagnostic confusion may be explained, in part, by the high inter-individual variability of the reference landmarks and planes used in orthodontic cephalometric tests (12, 28). It is interesting to note that normalizing the ANB angle led to a larger improvement compared to the Wits appraisal (9.9 versus 6.4 %, respectively, Table 2). This might be explained by a slightly more pronounced susceptibility of the Wits appraisal to changes in the cant of the occlusal plane (17), for which the normalizing procedure is not able to compensate fully.

The proposed approach differs from previously published methodology (29–32): in medical settings it is hardly ever possible to vary all the different classes' gold standard cut-off to the extent possible in cephalometrics, since there usually are few if any degrees to 'being ill'. This allowed us to rephrase the three-class diagnostic question into

a 2-class one, resulting in a somewhat differently shaped ROC surface. There naturally is a practical limit on the number of cut-off points at which the test can be evaluated: it is of little use to evaluate the test at many more levels as there are patients in the sample. Also, we decided to assess the diagnostic performance in between 2 standard deviations above and below the mean PC2 score, due to the dwindling number of patients above/below this limit.

From the clinical point of view, the most sophisticated methodology currently available for assessing intermaxillary relationships would seem to be the use of PC2 scores (22, 24, 25), which unfortunately requires the availability of a relatively large database of ethnicity-specific patient coordinates in order to scrutinize craniofacial variability (i.e. GPS and PCA). Another potential problem is the abstract nature of the PC2 scores (due to the lack familiarity to the clinical orthodontist). This might be circumvented simply by assigning patients the accompanying cephalometric value of the sample mean shape, deformed to each patient's position in the PC1–PC2 map, as proposed previously (measurement by proxy) (28). Since this procedure applies the same 'ruler' to all patients, it also prevents geometric distortion, although it again requires the availability of a coordinate database. Notwithstanding the trivial nature of providing such databases in anonymized form, the normalization procedure offers improved diagnostic performance in the absence thereof: only the sixteen coordinates of the sample mean shape are required to calculate the normalized values. The latter also represents values that are more familiar to most orthodontists.

Conclusion

The ANB angle and Wits appraisal were found to perform very similarly at about 80 percent area under the surface. Normalizing the ANB and Wits improved all VUS highly significantly, with almost 10 and 6.6 percent, respectively.

Supplementary material

Supplementary data (movies clarifying the shape changes associated with positional changes within the PC1-PC2 plot, as well as an interactive webpage containing a 3D representation of a ROC surface) are available at European Journal of Orthodontics online. <https://doi.org/10.1093/ejo/cjx002>

Conflict of interest

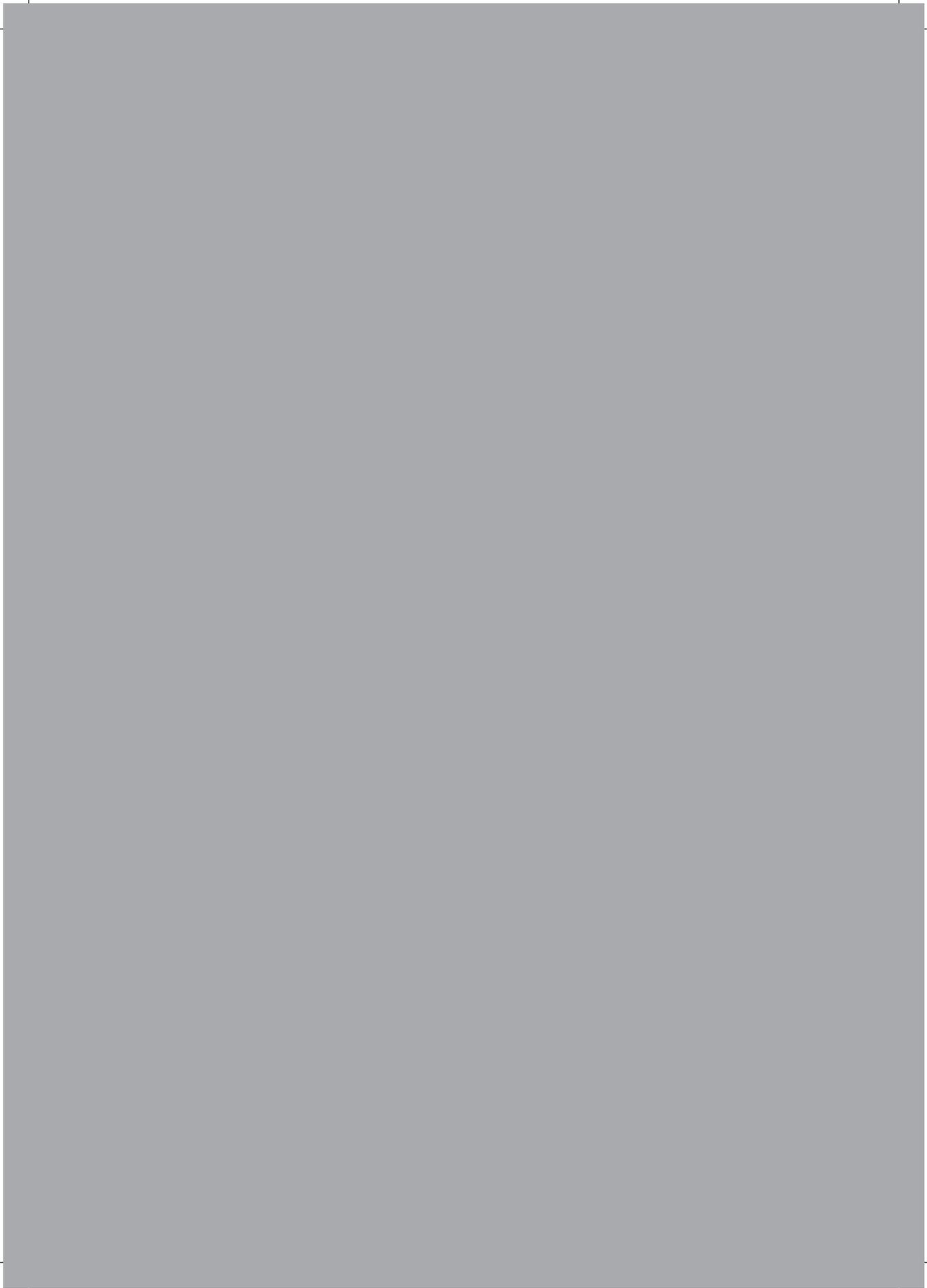
None to declare.

References

1. Graber, T., Vanarsdall, R. and Vig, K. (2005) *Orthodontics: Current Principles and Techniques*. Mosby, Philadelphia, Pasadena, 4th edn.
2. Proffit, W.R., Fields, H.F., Jr and Sarver, D.M. (2006) *Contemporary Orthodontics*. Elsevier Health Sciences, St. Louis, Missouri.
3. Atchison, K.A., Luke, L.S. and White, S.C. (1991) Contribution of pretreatment radiographs to orthodontists' decision making. *Oral surgery, oral medicine, and oral pathology*, 71, 238–245.
4. Han, U.K., Vig, K.W., Weintraub, J.A., Vig, P.S. and Kowalski, C.J. (1991) Consistency of orthodontic treatment decisions relative to diagnostic records. *American Journal of Orthodontics and Dentofacial Orthopedics*, 100, 212–219.
5. Hansen, K. and Bondemark, L. (2001) The influence of lateral head radiographs in orthodontic diagnosis and treatment planning. *European Journal of Orthodontics*, 23, 452–453.
6. Nijkamp, P.G., Habets, L.L., Aartman, I.H. and Zentner, A. (2008) The influence of cephalometrics on orthodontic treatment planning. *European Journal of Orthodontics*, 30, 630–635.
7. Devereux, L., Moles, D., Cunningham, S.J. and McKnight, M. (2011) How important are lateral cephalometric radiographs in orthodontic treatment planning? *American Journal of Orthodontics and Dentofacial Orthopedics*, 139, e175–e181.
8. Durão, A.R., Alqerban, A., Ferreira, A.P. and Jacobs, R. (2015) Influence of lateral cephalometric radiography in orthodontic diagnosis and treatment planning., *The Angle orthodontist*, 85, 206–210.
9. Rischen, R.J., Breuning, K.H., Bronkhorst, E.M. and Kuijpers-Jagtman, A.M. (2013) Records needed for orthodontic diagnosis and treatment planning: a systematic review., *PLoS ONE*, 8, e74186.
10. Fryback, D.G. and Thornbury, J.R. (1991) The efficacy of diagnostic imaging. *Medical Decision Making*, 11, 88–94.
11. Baumrind, S. and Frantz, R.C. (1971) The reliability of head film measurements. *American Journal of Orthodontics*, 60, 111–127.
12. Wellens, H. (2009) Improving the concordance between various anteroposterior cephalometric measurements using Procrustes analysis. *European Journal of Orthodontics*, 31, 503–515.
13. Taylor, C.M. (1969) Changes in the relationship of nasion, point A, and point B and the effect upon ANB. *American Journal of Orthodontics*, 56, 143–163.
14. Freeman, R.S. (1981) Adjusting A-N-B angles to reflect the effect of maxillary position. *The Angle orthodontist*, 51, 162–171.
15. Binder, R.E. (1979) The geometry of cephalometrics. *Journal of Clinical Orthodontics: JCO*, 13, 258–263.
16. Hussels, W. and Nanda, R.S. (1984) Analysis of factors affecting angle ANB. *American Journal of Orthodontics*, 85, 411–423.

17. Roth, R. (1982) The 'Wits' appraisal - its skeletal and dento-alveolar background. *European Journal of Orthodontics*, 4, 21–28.
18. Swets, J.A., Green, D.M., Getty, D.J. and Swets, J.B. (1978) Signal detection and identification at successive stages of observation. *Perception & Psychophysics*, 23, 275–289.
19. Metz, C.E. (1978) Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8, 283–298.
20. Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36.
21. Swets, J.A. (1979) ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, 14, 109–121.
22. Wellens, H.L., Kuijpers-Jagtman, A.M. and Halazonetis, D.J. (2013) Geometric morphometric analysis of craniofacial variation, ontogeny and modularity in a cross-sectional sample of modern humans. *Journal of Anatomy*, 222, 397–409.
23. McIntyre, G.T. and Mossey, P.A. (2003) Size and shape measurement in contemporary cephalometrics. *European Journal of Orthodontics*, 25, 231–242.
24. Halazonetis, D.J. (2004) Morphometrics for cephalometric diagnosis. *American Journal of Orthodontics and Dentofacial Orthopedics*, 125, 571–581.
25. Akli, E., Marinaki, L. and Halazonetis, D.J. (2015) Selecting subjects with high craniofacial shape homogeneity for clinical trials. *American Journal of Orthodontics and Dentofacial Orthopedics*, 148, 1026–1035.
26. Dryden, I.L. and Mardia, K.V. (1998) *Statistical Shape Analysis*. Wiley, Chichester, New York.
27. Zelditch, M.L., Swiderski, D.L. and Sheets, H.D. (2012). *Geometric Morphometrics for Biologists: A Primer*. Academic Press, Amsterdam, The Netherlands, 2nd edn.
28. Wellens, H.L.L. and Kuijpers-Jagtman, A.M. (2016) Connecting the new with the old: modifying the combined application of Procrustes superimposition and principal component analysis, to allow for comparison with traditional lateral cephalometric variables. *European Journal of Orthodontics*, 38, 569–576.
29. Mossman, D. (1999) Three-way ROCs. *Medical Decision Making*, 19, 78–89.
30. Dreiseitl, S., Ohno-Machado, L. and Binder, M. (2000) Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making*, 20, 323–331.
31. Nakas, C.T. and Yiannoutsos, C.T. (2004) Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine*, 23, 3437–3449.
32. Li, J. and Fine, J.P. (2008) ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics*, 9, 566–576.
33. Gower, J.C. and Dijksterhuis, G.B. (2004) *Procrustes Problems*. Oxford University Press, New York.
34. Claude, J. (2008) *Morphometrics* with R. Springer, New York, 1st edn.
35. Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441.

36. Han, U.K. and Kim, Y.H. (1998) Determination of Class II and Class III skeletal patterns: receiver operating characteristic (ROC) analysis on various cephalometric measurements. *American Journal of Orthodontics and Dentofacial Orthopedics*, 113, 538–545.
37. Freudenthaler, J.W., Celar, A.G. and Schneider, B. (2000) Overbite depth and anteroposterior dysplasia indicators: the relationship between occlusal and skeletal patterns using the receiver operating characteristic (ROC) analysis. *European Journal of Orthodontics*, 22, 75–83.
38. Kumar, S., Valiathan, A., Gautam, P., Chakravarthy, K. and Jayaswal, P. (2012) An evaluation of the Pi analysis in the assessment of anteroposterior jaw relationship. *Journal of Orthodontics*, 39, 262–269.
39. Neela, P.K., Mascarenhas, R. and Husain, A. (2009) A new sagittal dysplasia indicator: the YEN angle. *World Journal of Orthodontics*, 10, 147–151.
40. Baik, C.Y. and Ververidou, M. (2004) A new approach of assessing sagittal discrepancies: the Beta angle. *American Journal of Orthodontics and Dentofacial Orthopedics*, 126, 100–105.
41. Anderson, G., Fields, H.W., Beck, F.M., Chacon, G. and Vig K.W. (2006) Development of cephalometric norms using a unified facial and dental approach. *Angle Orthodontist*, 76, 612–618.



CHAPTER 6

General discussion and future perspectives

6

7.1 Introduction

Notwithstanding the impressive volume of literature pertaining to lateral cephalometry (as of May 2018, a simple PubMed search for “Cephalometry” returns over 25,000 results), it seems orthodontists still struggle to efficiently distill diagnostic data from these types of radiographs. Two recent meta-analyses failed to clearly establish the usefulness of lateral cephalometry in orthodontics, concluding that lateral cephalograms are not routinely needed for orthodontic treatment planning (1,2).

Many explanations have been offered to explain the difficulties associated with analyzing lateral cephalograms, which can broadly be categorized as technical problems, such as x-ray beam divergence, doubling of bilateral anatomical features, structural shrouding, and head rotation in the cephalostat, which together compound the task of accurately and reliably identifying landmarks. On the other hand, analytical aspects pertain to, or are associated with the choice and definition of (reference) landmarks, as well as the fixed or floating, age, gender and/or ethnicity dependent value of the associated normative values. A third, and often overlooked category is geometrical distortion, which might be defined as ‘the geometric phenomenon allowing patients with seemingly identical mandibulo-maxillary relationships to exhibit markedly different cephalometric values’, and has frequently been associated with measures of sagittal discrepancy, such as the ANB angle and the Wits appraisal (e.g. the influence of the vertical or anteroposterior position of point N, relative bimaxillary protrusion, retrusion, or rotation and facial height on the ANB angle, influence of occlusal plane rotation and dento-alveolar height on Wits appraisal).

The central premise of this thesis is that the role of geometrical distortion in lateral cephalometry can be understood more clearly by considering the somewhat vexing analogy between the latter and a so-called ‘cadastral land survey’ (i.e. the land-surveyor analogy). Land surveyors also make extensive use of angular and linear measurements to determine, among others, property boundaries. However, when attempting to apply some of our traditional approaches for assessing sagittal intermaxillary relationships to the cadastral survey (the ANB angle and Wits appraisal), it becomes intuitively clear that inter-individual variation in the location of the associated reference landmarks and planes essentially forms the basis of geometrical distortion. The absence of any tools to avoid positional uncertainty of the reference structures essentially precludes any definitive statement as to the validity of the resulting measurements.

Despite its deceptive simplicity, the land-surveyor analogy allows some surprisingly profound insights. For one, the analogy suggests that geometric distortion might constitute one of lateral cephalometry’s more fundamental problems: what exactly

is the relevance of landmarking error in lateral cephalometry if the appropriateness of the measurement's reference landmarks is doubtful to begin with? Secondly, the traditional approach to problem-solving in cephalometrics, simply moving around the 'cephalometric measuring-tripod' to a different set of landmarks or planes to perform linear or angular measurements from, would seem to serve little purpose other than shifting positional uncertainty from one set of reference landmarks to the next. Land surveyors cleverly avoid this measurement conundrum by employing a fixed, external reference frame, consisting of benchmarks (i.e. points of which the location has been determined highly accurately in three spatial dimensions). This begs the question whether lateral cephalometry could potentially benefit from imitating the land-surveyors' approach, by performing the measurements from a fixed reference frame instead of an inter-individually variable one.

Several questions come to mind when considering the studies presented in the previous chapters. While attempting to address these questions, potential future research directions for the currently presented work will be addressed as well.

7.2 Relevance of the proposed methodology

A first, fairly obvious question pertains to the relevance of the currently presented work, particularly in light of the recent addition of three-dimensional cone beam computed tomography (3D CBCT) to the orthodontic diagnostic armamentarium. It is perfectly reasonable to wonder why one should still bother investigating a rather old, strictly planar imaging technique when an exciting three-dimensional alternative is available (3,4).

The answer to this question is twofold. First, there are good reasons to believe that (in Europe) lateral cephalometry will, at least in the near future, remain the imaging method of choice for assessing craniofacial growth. Although continuous technological advancements allow incremental reductions in the radiation exposure associated with 3D CBCT imaging (5,6), similar dose reductions apply to two-dimensional imaging as well. By applying the ALARA principle (As Low As Reasonably Achievable), the radiation protection guidelines published by the European Union (7) still argue against the routine application of 3D CBCT in orthodontic diagnosis and treatment planning. A selective approach, evaluated on a case-by-case basis (e.g. craniofacial syndromes, ectopic eruption problems, facial asymmetry, etc.), and with appropriate (i.e. limited) field-of-view and associated beam energy is advocated instead (7). The recently developed Dutch guideline on radiographs for orthodontic diagnosis and treatment planning in normal orthodontic patients takes the same restrictive stand (8). Furthermore, the

dose reductions currently achieved in (ultra-low dose) 3D CBCT, albeit promising, come with a clear ‘cost’ as (at present) the dosage reduction may limit the quality of the images that can be obtained (5). Intriguingly, there seems to be somewhat of a geographical divide in the approach to radiation hygiene on both sides of the Atlantic (4,9): the North-American literature appears to adopt a slightly more relaxed attitude in this regard, where 3D CBCT is sometimes viewed as the de-facto (i.e. standard) imaging modality (5,6).

There is no debate about the advantages of 3D CBCT, which is not burdened by many of the complications of lateral cephalometry covered extensively in chapter two, such as cephalometric enlargement, structural shrouding and/or superimposition and head positioning-induced errors (4,10). The accuracy of 3D CBCT with regard to linear measurements has been clearly established, as well as its’ usefulness for diagnosing canine impaction and supernumerary teeth, endodontic and periodontal problems, root resorption, condylar pathology and craniofacial asymmetries, as well as airway and even sinus problems (3,4,9,11). Furthermore, 3D CBCT allows assessing the space available for placing temporary anchorage devices (11), and has proven instrumental in demonstrating the treatment changes caused by bone anchored expansion and protraction appliances or surgery, via intra-individual (skull base) superimposition of the reconstructions before and after treatment (12). Nevertheless, there are only a few studies that show that the availability of a CBCT contributed significantly to a change in the orthodontic treatment plan (13–16).

Despite all these advantages, it is nonetheless quite intriguing to note how the orthodontic literature seems to mostly, if not completely ignore the failure of cone beam computed tomography to ease the challenge of extracting valid cephalometric measurements from the three-dimensional image. Geometric distortion, in all of its variations (10), applies to 3D CBCT to at least the same extent as it does to conventional lateral cephalometry. In fact, one might argue that the three-dimensional character of the former offers an additional dimension for geometric distortion to express itself in. The 3D CBCT version of the land surveyor hypothesis requires taking the vertical dimension into consideration as well, further compounding an already convoluted problem.

The currently proposed methods for analyzing 3D CBCT datasets usually are simple extensions of techniques introduced earlier in two-dimensional cephalometry (17–21). Moreover, some proposed techniques even completely pass by the three-dimensional aspect of the data, by reconstructing a two-dimensional lateral or frontal cephalogram from the 3D CBCT image and identifying landmarks two-dimensionally (19,22,23). Again referring to the land-surveyor analogy, it is of little consequence in this regard if the proposed methods employ seemingly “new” landmarks or planes (20,24,25),

identifiable only on a 3D view of the skull, to perform the measurements from (i.e. “moving around of the cephalometric tripod”). From this perspective, it makes perfect sense to first attempt solving the comparatively easier two-dimensional problem as expressed in lateral cephalometrics, to then attempt extending the solution, if any, to the third dimension. The latter would be an obvious target for future research.

7.3 Skeletal versus soft-tissue based cephalometric analysis?

Another potential criticism pertains to the skeletal nature of the proposed methodology. Contrary to the mostly skeletally oriented analyses proposed shortly after the introduction of cephalometry in orthodontics (26,27) more recent approaches seem to place greater emphasis on the appraisal of the esthetic qualities of the patient’s profile, both in frontal and lateral view (28–33). Unfortunately, the science underpinning the currently proposed methods for appraising facial esthetics is often quite thin (34). In adolescents, few of the proposed angles, distances and ratios seem to have a significant relationship with facial esthetics (35,36). Furthermore, although assessing which of the profile’s components contribute to, or detract from the patient’s facial esthetics is undoubtedly important, orthodontists and maxillofacial surgeons typically only have direct control over the supporting structures of the soft tissues (i.e. the dental and skeletal structures). Additionally, the correlation between the soft tissue and its underlying skeletal/dental structures has been shown to be rather poor (37), suggesting their shape patterns may exhibit considerable independent variation. Combined, this makes the task of predicting the soft tissue response to treatment notoriously difficult.

Nevertheless, it would be interesting to extend the methodology as proposed in this thesis to include the soft tissues, for instance by performing a joint Procrustes superimposition of the hard and soft tissues. The latter approach is not usually reported in the geometric morphometric literature since from a mathematical/methodological point of view, the low correlation between these structures’ patterns of variation implies that it makes much more sense to describe them independently. In this way, the potentially larger variation in one is prevented from ‘contaminating’ the other’s. One might of course argue that orthodontic diagnosis represents a distinct usage scenario, in which the interrelationship between the variational patterns of the facial hard and soft tissues is of specific interest, however moderate their correlation. From this point of view, it might make sense to perform such a joint superimposition, although it would clearly impact the way the hard and soft tissues variational patterns are interpreted.

As mentioned above, another approach would be to perform independent Procrustes superimpositions of the hard and soft tissues (37,38). Albeit informative, this approach

has the distinct disadvantage of losing the 'direct connection' between both: compared to a joint superimposition, separate superimpositions obfuscate the analysis and interpretation of how a change in one might impact the other. A comparison of both approaches in a future project could provide more insight into this dilemma.

7.4 Sliding semi-landmarks?

Most anatomical structures are relatively sparse in terms of clearly defined landmarks, such as intersections or extremities of bony sutures (e.g. Nasion) or sharply delineated extremities (e.g. anterior and posterior nasal spine). Instead they are characterized predominantly by rather 'featureless' smooth curves or surfaces (39–41). Conventional cephalometrics deals with this situation by defining some landmarks based upon the spatial orientation of the digitized landmark relative to a 'horizontal' and/or 'vertical' direction (the orientation of which is often determined quite arbitrarily) (42–45). An excellent example is the gnathion landmark, which is usually defined as "the most anterior-inferior point on the anterior surface of the mandibular symphysis". The antero-posterior direction can however be defined in a number of ways (the porion-orbitale line, an arbitrary angle to the SN-line, perpendicularly to the true-vertical line, natural head-position, etc.) (39,42–45).

This approach is clearly not compatible with the geometric morphometric workflow: aside from the obvious drawbacks associated with an arbitrary choice of an axis system (39), orientation is treated as a nuisance factor in the Procrustes algorithm, and is thus removed from the equation (46–48). Therefore, the landmarks used in a geometric morphometric project should ideally be orientation-independent. Since the result of the Procrustes algorithm depends on locations of the landmarks, a position-dependent placement of the landmarks may result in malposition of these landmarks after the (generalized) Procrustes rotation. The original 'horizontal' may no longer be such after the rotation since the 'horizontal direction', whatever its definition, is determined globally, instead of individually. The geometric morphometric framework provides a solution to this predicament in the form of semi-landmarks which (on top of the usual translation, scaling and rotation of the Procrustes algorithm) are allowed to slide along the pertaining curves and surfaces until they match the positions of a reference configuration's corresponding points (40,49,50). In the process, either the bending energy (51) or the perpendicular Procrustes distance to the curve/surface is minimized, as part of an additional step of the iterative generalized Procrustes algorithm (49). This ensures the landmarks are homologous among patients, since the latter is no longer defined by the (sliding) landmarks themselves, but by the curves that define them (52).

Nevertheless, a conscious decision was taken not to apply sliding semi-landmarks in this thesis. Although the above-mentioned critiques with respect to combining the conventional lateral cephalometric landmarking scheme with modern geometric morphometric methods undoubtedly hold true, they seem to be of little relevance to the specific usage scenario of the methodology presented in this thesis: the diagnosis of individual patients. From a clinical/diagnostic point of view, we are for instance less likely to be interested in fully appraising the anatomic details of the mandibular outline all the way from the chin to the condylar process, which would be an excellent scenario for the application of sliding semi-landmarks (53). Instead we are more likely interested in the broader anatomical relationships, such as the antero-posterior and vertical position of the mandible relative to the maxilla and the skull base. From this point of view (i.e. in view of the relative lack of the orientation-dependent landmarks in the proposed methodology), it makes less difference if the mandibular chin point or gonial angle landmarks have been slid to minimize bending energy or perpendicular Procrustes distance.

Nevertheless, it is perfectly possible to integrate sliding semi-landmarks in the currently proposed methodology, by employing a hybrid approach. Since the proposed methodology aims to provide a fixed reference frame for cephalometric analysis, it is not recommended to perform a generalized Procrustes analysis, adding the patient to the reference sample, as this might influence the resulting mean shape (albeit very subtle, depending on the sample size). Instead, an ordinary Procrustes analysis is performed of said patient on the sample's mean shape. The sliding semi-landmarks (located for instance on the outline of the symphysis, mandibular angle, etc.) are then slid relative to the corresponding landmarks on the sample mean shape. Since the configuration is potentially no longer in Procrustes superimposition after sliding the semi-landmarks, the procedure is repeated (alternating Procrustes superimposition and landmark sliding) until the resulting change in Procrustes distance is below a certain threshold. This approach may provide some benefits in terms of reducing digitizing error, since the pertaining landmark locations are less orientation dependent, their exact position being determined solely by the sliding procedure. A project clarifying this approach and the combination with the subsequent steps of the analysis as laid out in this thesis should be planned for future research.

7.5 Relevance, from a geometric morphometric methodological perspective

The traditional, generalized Procrustes superimposition (GPS) and principal component analysis (PCA) driven geometric morphometric workflow has been criticized recently for its' apparent lack of biological foundation (22,23,28). One comment pertained to the strictly mathematical manner along which PCA decomposes the major directions

of shape variation, without any consideration for the anatomical or developmental questions driving the morphological investigation. With regard to GPA, there is the proven susceptibility to the relative distribution of the landmark points, not just in terms of the rather hypothetical ‘Pinocchio effect’ brought about by extreme displacements of individual landmarks (54), but also in the influence of the landmark distribution on the resulting covariance structure of the superimposed data (39) More specifically, Procrustes analysis treats similar shape variation among larger and smaller structures within the same shape differently.

Bookstein’s critiques (55,56) pertain specifically to the ill-advised use of the GPS/PCA workflow as a de facto ‘fishing expedition’ for any salient directions of morphological variation of the organism under investigation, without the need for the biologist to develop any working hypotheses (55–57). This approach, together with the aforementioned methodological weaknesses of the GPS/PCA workflow, may indeed allow misleading variational patterns to be erroneously construed as true biological ‘signals’. The message of Bookstein’s critical appraisal of the GPS/PCA workflow in fact was to stress the importance of ‘good scientific practice’: the formulation of the initial working hypothesis, to be confirmed or dismissed via differing investigational approaches, ideally with varying experimental conditions (57). In any case, the hypothesis under investigation should drive the investigation, as opposed to a battery of biology-agnostic mathematical algorithms.

The current situation of lateral cephalometry, however, paints a different picture from the one described above. For one, orthodontists are acutely aware of the craniofacial growth patterns under investigation, as articulated so skillfully by Sassouni almost 60 years ago (58). Furthermore, the shape variations described by the first two principal components in chapter 3 align extremely well with these preconceptions regarding the most important directions of human facial variation. As pointed out by Halazonetis (59), simply rotating the PC1-PC2 plot and/or changing the sign on the PC1 and/or 2 axis results in a graph strikingly similar to the one generated by Sassouni (58) when deforming the sample mean shape a few standard deviations in the plus and minus direction along both resulting axes. As such, the geometric morphometric workflow provides an excellent tool to mathematically concretize the orthodontists’ preconceptions regarding craniofacial shape variation, while avoiding many of the aforementioned pitfalls characterizing the traditional craniometric approach (e.g. separating size from shape, avoiding the choice of a ‘true horizontal’ or reference landmarks/planes, etc.). Nevertheless, it would be of interest to investigate how the alternative geometric morphometric methodology recently proposed by Bookstein (56) that allows the researcher more control over the directions of interest in shape space, might be integrated in the currently proposed methodology.

7.6 Worked example

To illustrate how some of the concepts introduced in the previous chapters and the above discussion might be applied in clinical practice, we will analyze the outcome of a combined maxillofacial-orthodontic treatment from a lateral cephalometric perspective, using a simultaneous superimposition of both hard and soft tissues, as suggested in the discussion.

This almost 15 years old male patient presented with a skeletal Class II, brachyfacial growth pattern, characterized by a somewhat obtuse nasolabial angle, a retruded mandible and lower lip, with a pronounced labiomental fold and equally pronounced chin button (Fig. 1). There was little incisor display at rest or during smiling (Figs. 1 and 2). Dentally, a Class II molar and canine occlusion was found, which was about half a premolar width on one side, and almost full cusp on the other, with slight spacing in both jaws (Fig. 2).



Figure 1. Patient's pre-treatment extra-oral views



Figure 2. Patient's pre-treatment intra-oral views

Even in the absence of a cephalometric analysis, the brachyfacial growth pattern and mandibular (dento-alveolar) retrusion were obvious in the lateral cephalogram (Fig 3,A). Despite the slightly obtuse nasolabial angle, the upper incisors exhibited a pronounced labial inclination. The orthopantomogram was mostly unremarkable, with the exception of the absence of third molars in the upper jaw (Fig. 3B). The patient's main complaint was the overjet, and to a lesser extent the spacing in the upper jaw. From an orthodontic point of view, the main concern was preventing further deterioration of the patient's profile due to flattening of the upper lip as a consequence of upper incisor retraction, if treatment were to be executed using (functional) orthodontic treatment alone.



Figure 3, A. Patient's lateral cephalometric view. Even in the absence of a cephalometric analysis, the brachyfacial growth pattern and mandibular (dento-alveolar) retrusion are obvious. Despite the labial inclination of the upper and lower incisors, the upper lip seems to lack support, slightly slanting posteriorly.



Figure 3, B. The patient's orthopantomogram, which seems largely inconspicuous, with the possible exception of the absent upper third molars.

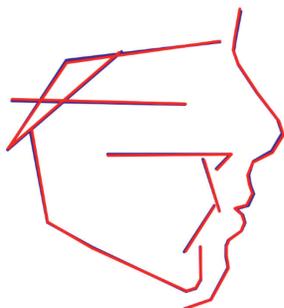
Since the aim was to analyze and compare this patient's pre- and post-treatment lateral cephalograms using both skeletal and soft tissue landmarks, some preparations were required in order to be able to proceed: the combined hard and soft-tissue morphological variability of the two hundred patients included in chapter three's study had to be ascertained first. Applying the concepts laid out in chapter three, the pertaining landmark configurations were superimposed using Procrustes analysis, after which the male and female mean shapes were calculated and the shape difference between them tested for significance using a 10.000 round permutation test (without replacement, significance level set at $p = 0.05$, please refer to chapter three for details).

Furthermore, the sample was divided into younger and older individuals based upon to the overall median (13.11 years), and the corresponding mean shapes were again contrasted in terms of shape differences using a permutation test. One might argue that this represents a somewhat crude assessment of age-related shape differences, since there is no clearly defined age gap between the two resulting groups, but for diagnostic purposes, this will do fine. The two previous tests reveal whether separate "templates" (i.e. mean shapes) should be used for male versus female patients, and younger versus older ones. The results were clearly non-significant for both, indicating that the sample could be pooled for calculating the overall mean shape (Fig. 4, A and B).

Next, the Procrustes superimposed configurations were subjected to principal component analysis, to uncover the directions of major shape variation. Which PCs are of particular interest, was determined by visualizing them, calculating the percentage of variation

explained by each, and by determining the number of biologically interpretable PCs (i.e. those that describe biologically meaningful information, as opposed to measurement noise). Nine PCs were found to be biologically interpretable, using Perez-Neto's random average under permutation test (60). Eight of these are depicted in Fig. 5.

Male/female mean shape



Younger versus older mean shape

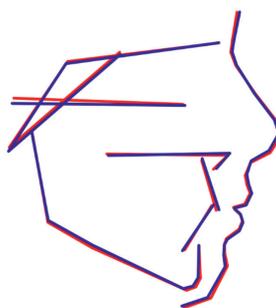


Figure 4, A. Male (in blue) versus female (in red) mean shapes. The shape difference was non-significant at $p = 0.1594$. For clarity's sake, the pooled mean shape is not shown, since it would basically be indistinguishable from the male and female ones. **Figure 4, B.** Younger (in blue) – older (in red) mean shapes. These also did not differ significantly in terms of their shape ($p = 0.423$). Intriguingly the older (red) mean shape exhibits an ever so slightly larger nose and chin, without similar protrusion at the lips; facial features which have frequently been associated with the final stages of facial growth

Interestingly, the two first PCs again seemed to mostly describe those directions of variations of particular interest to our profession: dolichofacial versus brachyfacial morphology for PC1, and relative maxillomandibular protrusion and retrusion for PC2. Soft tissue variation seemed to dominate the next two PCs, which describe variations in nasal size (i.e. larger or smaller nose) combined with lip protrusion/retrusion for PC3, and nasal dorsal length variations, combined with a more anterior/posterior nasal bridge, as well as palatal plane cant changes for PC4. These last two PCs appeared to confirm the conclusions of Halazonetis, that the correlation between soft tissue variation and that of the underlying hard tissues is low (38), as few and very small concomitant skeletal changes are observable in these PCs. It is also visually obvious that subsequent PCs describe less and less of the remaining variation, as the observable morphological differences become smaller and smaller with higher PC counts.

With the information thus collected, we were able to return our attention to the patient under investigation. Having determined that the first two PCs describe vertical and anteroposterior morphological characteristics of interest, which in combination explained 37.8 per cent of total shape variation (PC1 explains 22.7 % and PC2 16 %, Table

l), a next logical next step was to visualize the Procrustes superposition of the patient's landmarks onto the sample mean shape, and to augment this information with the patient's position in the PC1-PC2 score plot. From Fig. 5, we can ascertain that patients located more to the right in this PC1-PC2 plot (i.e. exhibiting higher PC1 scores) are more brachyfacial, while those located more to the left are more dolichofacial. Vertically, patients with higher PC2 scores are more "retrognathic", while those positioned lower in the plot ore more "prognathic". The Procrustes superposition can provide further insights if a thin plate spline deformation plot is added, which represents the smooth deformation of the sample mean shape onto the patients coordinate data. The resulting grid lines highlight the most important locations of shape differences: the larger the difference, the more locally deformed the grid becomes.

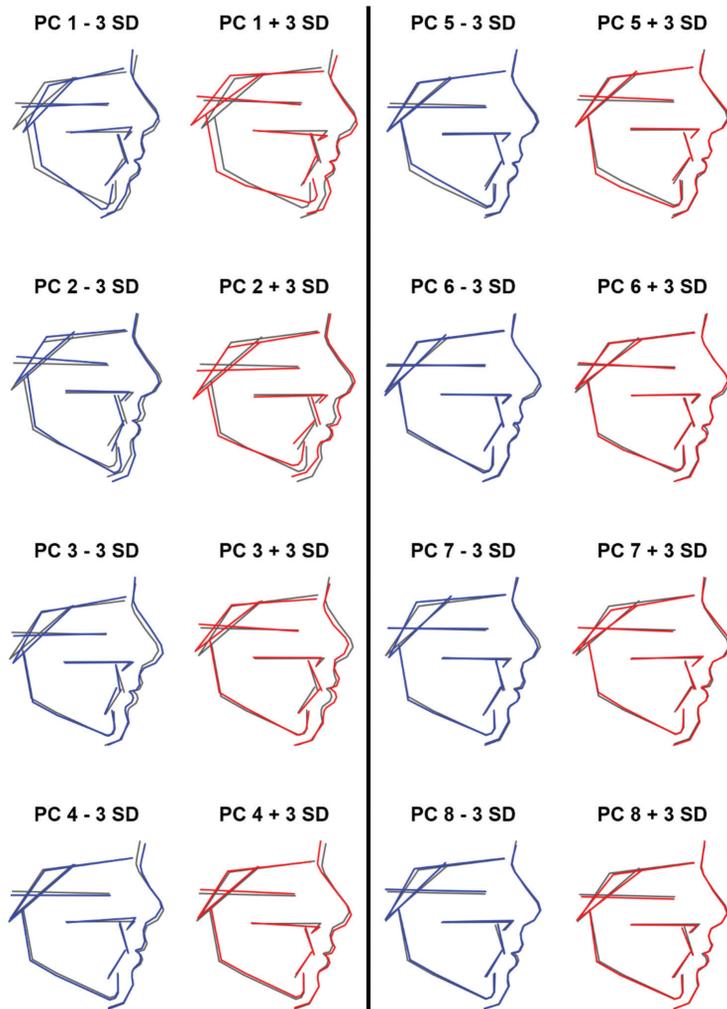


Figure 5. Principal components 1 through 8. Mean shape (in grey) plus (in red) or minus (in blue) 3 SDs .

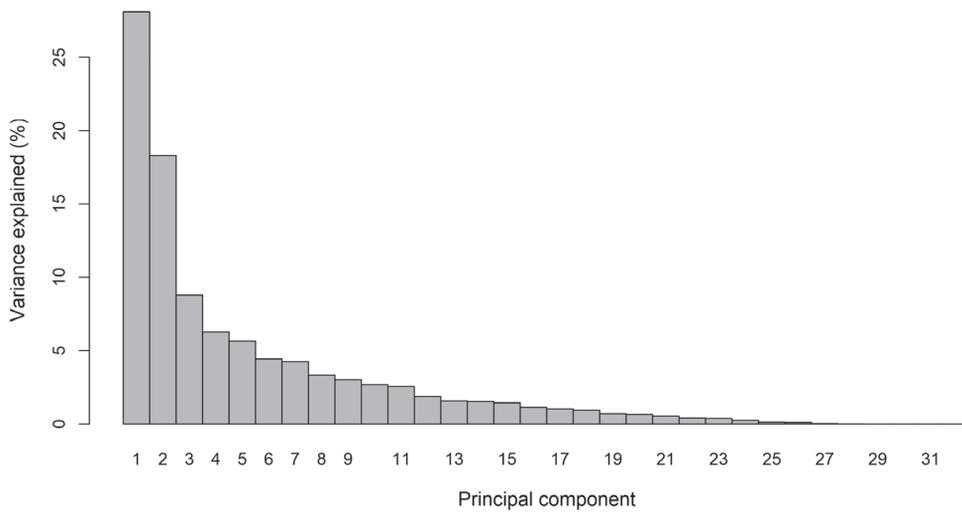


Figure 6. Scree plot, visualizing the percentage of total variability explained by each PC. It is obvious that the first few PCs explain the majority of the total variation in shape. The inflection in the “curve” is sometimes used to determine the number of biologically interpretable PCs. Likewise, those PCs explaining more than five percent of total variation are sometimes selected as such. It is however preferable to apply more robust methods, such as those referred to in the text above and in chapter three, to determine the number of biologically interpretable PCs.

Table 1: Percentage of variability explained by each PC for PCs 1 through 10.

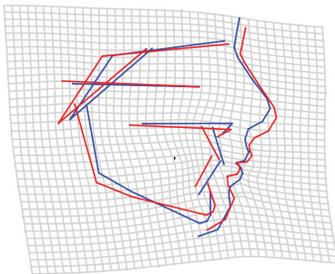
PC	% of variability explained	cumulative sum
1	22.7	22.7
2	16.0	38.7
3	10.8	49.5
4	5.7	55.2
5	4.4	59.5
6	3.6	63.1
7	3.2	66.3
8	2.8	69.1
9	2.4	71.4
10	2.3	73.8

The left pane in Fig. 7 suggests the patient (in red) has a pronounced brachyfacial growth pattern compared to the sample mean shape (in blue). This is confirmed by the PC1-PC2 plot (the right pane of Fig. 7), which depicts our patient (red circle) almost 3 SDs away from the mean in terms of his PC1 score; in fact, no one in the sample of 200 patients described in chapter three was more brachyfacial. As such, a particular distance from

the mean (beyond 2 to 2.5 SDs in terms of PC1 and/or PC2 score, for example) might perhaps constitute a more objective measure for determining the need for surgery. Also from the left pane in Fig. 7, the relative anteroposterior position of the patient's jaws seems fairly normal, which more or less corresponds with the patient's PC2 score just beyond +1 SD (i.e. slightly retrognathic).

The superposition also suggests that flattening of the upper lip due to upper incisor retraction (which are clearly labially inclined) might be unfavorable for the profile due to the relative prominence of the nose, and that bringing the mandible forward surgically to correct the bite might lead to excessive chin prominence. Orthodontic treatment consisted of aligning, decompensating and coordinating both arches with full fixed straightwire appliances. Care was taken to protect upper lip support. A bilateral sagittal split osteotomy was subsequently performed in the mandible and a Lefort-I in the maxilla. A chin osteotomy was performed as well, to prevent excessive chin prominence. The orthodontic treatment was completed in about 24 months. The patient's post-treatment facial and intra-oral views are depicted in Figs. 8 and 9.

Patient/Sample mean shape



PC1 versus PC2 scores

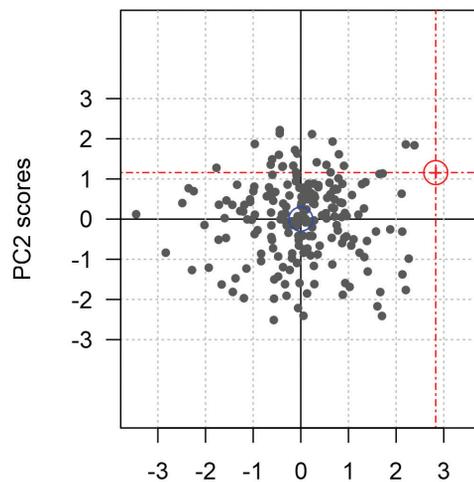


Figure 7. The left pane depicts the patient's pre-treatment tracing (in red) superimposed on the sample mean shape (in blue), with the accompanying thin spline deformation plot. The patient's position in the PC1-PC2 plot is displayed in the right pane.

A pleasing, harmonious profile was obtained after treatment, although the chin still appeared somewhat prominent. The patient's vertical relationships improved considerably, while the relative prominence of the nose decreased. Dentally, a solid Class I occlusion was obtained, while the overjet and overbite were fully corrected. A consonant smile line was obtained, with sufficient incisor display on smiling



Figure 8. Patient's post-treatment facial views.

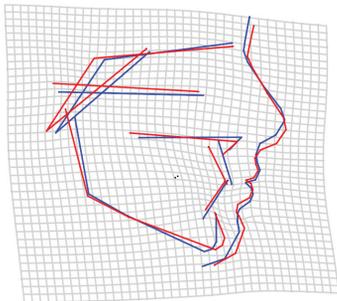


Figure 9. Patient's post-treatment intra-oral views.

The information accumulated before can now be put to use for ascertaining the morphological changes that were brought about by the combined orthodontic-surgical treatment. There are several ways to go about this. One way is to reproduce Fig. 7, now using the superimposed post-treatment landmarks, while also visualizing the patient's new, post-treatment position in the PC1-PC2 plot (Fig. 10).

The resulting superimposition on the sample mean shape visually confirms many of the points raised above. Comparing Figs. 7 and 10, it is quite obvious the patient's post-treatment vertical relationships more closely resemble those of the mean shape. The decrease in the relative prominence of the nose is also readily discernible, as is the remaining slight prominence of the chin. The thin plate spline deformation plot after treatment is more relaxed (i.e. less warped) post-treatment. It is interesting to note that a seemingly successful treatment does not necessarily imply that all morphological variables were normalized: the PC1-PC2 plot seems to indicate the patient changed less than one SD in terms of PC1 (became slightly less brachyfacial), while also crossing the zero-mark into Class III territory (evidenced by the slightly Class III PC2 score). From the superimposition in Fig. 10, it nonetheless appears the post-treatment dento-alveolar relationships in this patient are fairly normal. This can however be understood by realizing that in this example, the mandibular chin region is characterized by three points (Pogonion, Gnathion and Menton), while the dento-alveolar region only has

Patient/Sample mean shape



PC1 versus PC2 scores

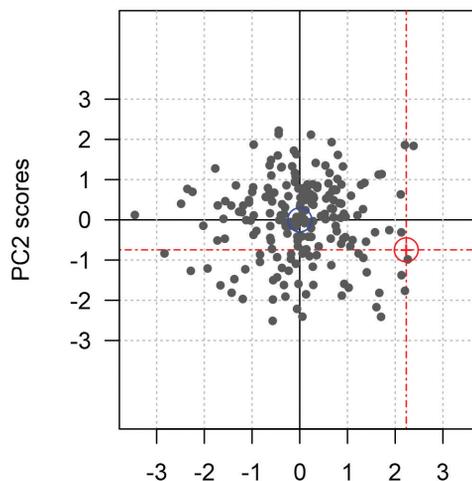


Figure 10. Superimposition of post-treatment tracing (in red) on the sample mean shape (in blue) with the accompanying thin plate spline deformation plot. The right pane depicts the patient's new position in the PC1-PC2 plot.

point A. The net result is that the chin point is more heavily “weighted” in the Procrustes superimposition, compared to the dento-alveolar region. The PC1-PC2 plot therefore merely confirms the chin point became more prominent. If a more dento-alveolarly oriented analysis is required, the number of chin point landmarks should be lowered.

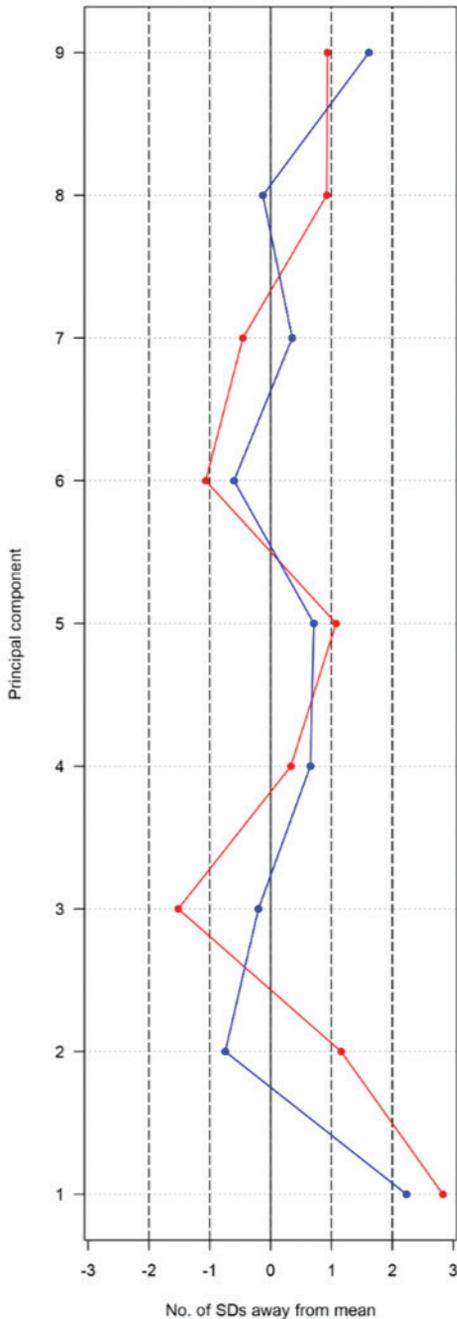


Figure 11. Patient’s PC scores, recorded before (in red) and after treatment (in blue). The largest morphological change brought about by this combined ortho-surgical treatment seems to pertain to PC2 (intermaxillary sagittal relationships), which changed slightly less than two SDs in the direction of relative mandibular prognatism.

Another interesting way to visualize the treatment changes in view of the morphological analysis presented above, is to depict PC scores from before to after treatment for all biologically interpretable PCs (9, in this example). The observed changes in PC scores can then be related to Fig. 5 for interpreting how facial shape was impacted by treatment.

Interestingly, five PCs were more than one SD away from the mean before treatment. This changed to two after treatment. Also, of interest, it seems the change in most PC scores was rather modest (in this particular patient), notwithstanding the surgical aspect of treatment.

This example is of course limited in scope, but serves to provide a glimpse of what can be accomplished by applying and further developing some of the concepts in this thesis.

7.7 Final remarks

The attentive reader will have noticed this PhD thesis (probably understandably) makes little attempt to provide any final, fundamental answers to the dizzyingly varied array of problems facing lateral cephalometry in orthodontics; instead it elucidates a rationale for analyzing these problems from a geometrical point of view. Furthermore, it provides some of the tools required to attempt solving these problems and testing the resulting solution's efficacy. Above anything else, the work should illustrate how there are no easy solutions to the cephalometric measurement predicament.

Quite to the contrary, it suggests the more final answers to our questions in this regard should be sought in a scientific field many orthodontists are unfamiliar with and like to avoid at all cost: mathematics. Just as there has turned out to be a requirement for individuals with training in biostatistics to improve our handling and analysis of scientific data, there seems to be an equally pertinent need for individuals with a thorough mathematical background, who genuinely understand our diagnostic needs and requirements, and are able to provide tailor-made, mathematically sound solutions for them. Until then, the techniques laid out in this PhD project may provide some useful tools for improving the diagnostic yield of future cephalometric analyses.

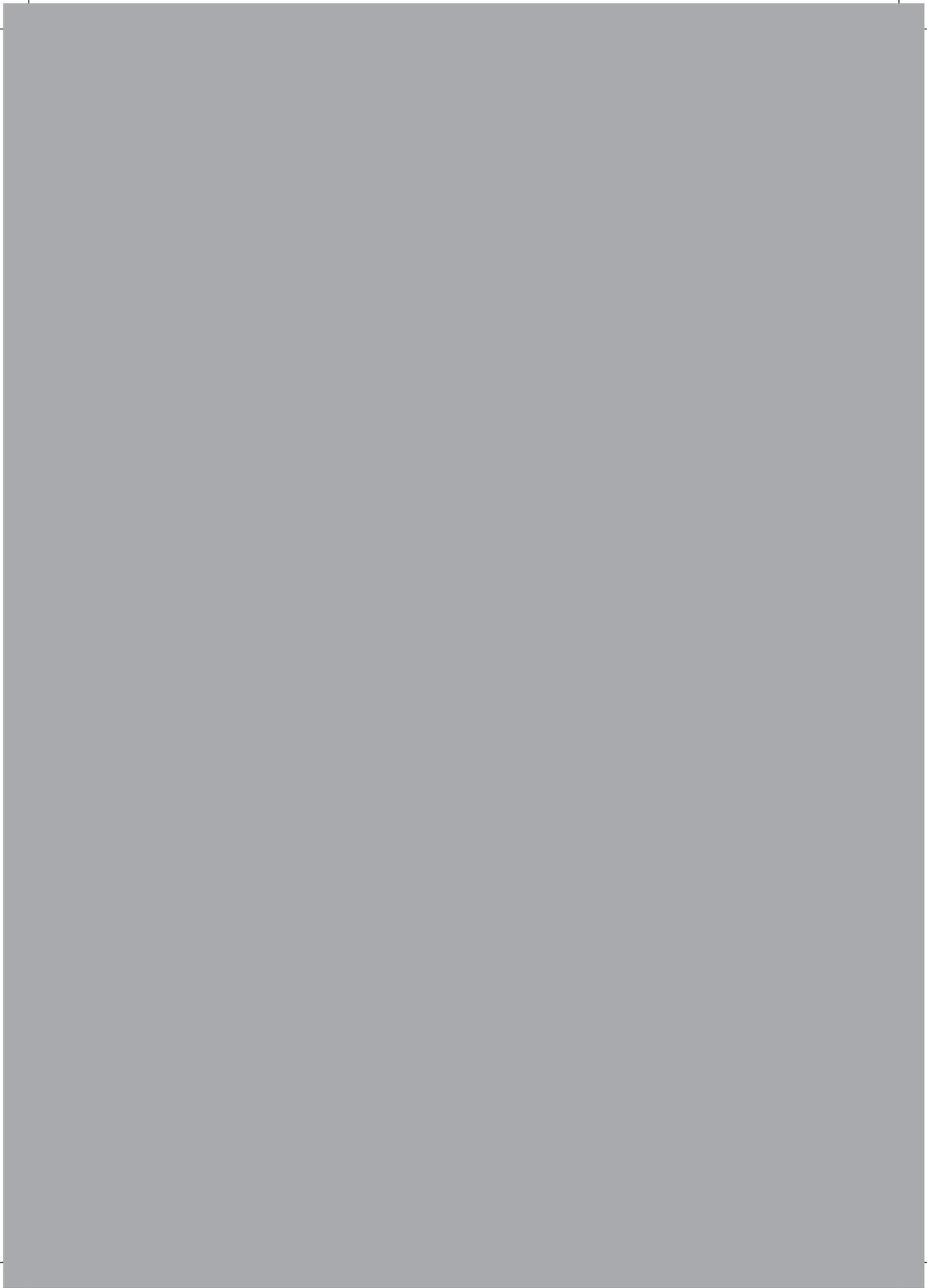
References

1. Rischen RJ, Breuning KH, Bronkhorst EM, Kuijpers-Jagtman AM. Records Needed for Orthodontic Diagnosis and Treatment Planning: A Systematic Review. PLoS ONE [Internet]. November 12th 2013 [cited march 4th 2018];8(11). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3827061/>
2. Durão AR, Alqerban A, Ferreira AP, Jacobs R. Influence of lateral cephalometric radiography in orthodontic diagnosis and treatment planning. *Angle Orthod.* 2015;85(2):206–10.
3. Kapila S, Conley RS, Harrell WE. The current status of cone beam computed tomography imaging in orthodontics. *Dentomaxillofacial Radiol.* 2011;40(1):24–34.
4. Larson BE. Cone-beam computed tomography is the imaging technique of choice for comprehensive orthodontic assessment. *Am J Orthod Dentofac Orthop.* 2012;141(4):402-10.
5. Liljeholm R, Kadesjö N, Benchimol D, Hellén-Halme K, Shi X-Q. Cone-beam computed tomography with ultra-low dose protocols for pre-implant radiographic assessment: An in vitro study. *Eur J Oral Implantol.* 2017;10(3):351–9.
6. EzEldeen M, Stratis A, Coucke W, Codari M, Politis C, Jacobs R. As Low Dose as Sufficient Quality: Optimization of Cone-beam Computed Tomographic Scanning Protocol for Tooth Autotransplantation Planning and Follow-up in Children. *J Endod.* 2017;43(2):210–7.
7. radiation_protection_172.pdf [Internet]. [cited februari 13th 2018]. Available at: http://www.sedentext.eu/files/radiation_protection_172.pdf
8. Nederlandse Vereniging van Orthodontisten. Richtlijn Orthodontie Röntgen. 2018;[in press].
9. Halazonetis DJ. Cone-beam computed tomography is not the imaging technique of choice for comprehensive orthodontic assessment. *Am J Orthod Dentofac Orthop.* 2012;141(4):403-11.
10. Wellens H. Improving the concordance between various anteroposterior cephalometric measurements using Procrustes analysis. *Eur J Orthod.* 2009;31(5):503–15.
11. Kapila SD, Nervina JM. CBCT in orthodontics: assessment of treatment outcomes and indications for its use. *Dento Maxillo Facial Radiol.* 2015;44(1):20140282.
12. Cevidanes LHC, Oliveira AEF, Grauer D, Styner M, Proffit WR. Clinical application of 3D imaging for assessment of treatment outcomes. *Semin Orthod.* 2011;17(1):72–80.
13. Bjerklin K, Bondemark L. Management of ectopic maxillary canines: variations among orthodontists. *Angle Orthod.* 2008;78(5):852–9.
14. Haney E, Gansky SA, Lee JS, Johnson E, Maki K, Miller AJ, e.a. Comparative analysis of traditional radiographs and cone-beam computed tomography volumetric images in the diagnosis and treatment planning of maxillary impacted canines. *Am J Orthod Dentofac Orthop* 2010;137(5):590–7.
15. Alqerban A, Willems G, Bernaerts C, Vangastel J, Politis C, Jacobs R. Orthodontic treatment planning for impacted maxillary canines using conventional records versus 3D CBCT. *Eur J Orthod.* 2014;36(6):698–707.

16. Eslami E, Barkhordar H, Abramovitch K, Kim J, Masoud MI. Cone-beam computed tomography vs conventional radiography in visualization of maxillary impacted-canine localization: A systematic review of comparative studies. *Am J Orthod Dentofac Orthop.* 2017;151(2):248–58.
17. Park S-H, Yu H-S, Kim K-D, Lee K-J, Baik H-S. A proposal for a new analysis of craniofacial morphology by 3-dimensional computed tomography. *Am J Orthod Dentofac Orthop.* 2006;129(5):600.e23-34.
18. Wong RWK, Chau ACM, Hägg U. 3D CBCT McNamara's cephalometric analysis in an adult southern Chinese population. *Int J Oral Maxillofac Surg.* 2011;40(9):920–5.
19. Rodriguez-Cardenas YA, Arriola-Guillen LE, Flores-Mir C. Björk-Jarabak cephalometric analysis on CBCT synthesized cephalograms with different dentofacial sagittal skeletal patterns. *Dent Press J Orthod.* 2014;19(6):46–53.
20. Lee M, Kanavakis G, Miner RM. Newly defined landmarks for a three-dimensionally based cephalometric analysis: a retrospective cone-beam computed tomography scan review. *Angle Orthod.* 2015;85(1):3–10.
21. Swennen G. *3D Virtual Treatment Planning of Orthognathic Surgery: A Step-by-Step Approach for Orthodontists and Surgeons.* Springer; 2016.
22. Moshiri M, Scarfe WC, Hilgers ML, Scheetz JP, Silveira AM, Farman AG. Accuracy of linear measurements from imaging plate and lateral cephalometric images derived from cone-beam computed tomography. *Am J Orthod Dentofac Orthop.* 2007;132(4):550–60.
23. Kumar V, Ludlow JB, Mol A, Cevidanes L. Comparison of conventional and cone beam CT synthesized cephalograms. *Dento Maxillo Facial Radiol.* 2007;36(5):263–9.
24. Pittayapat P, Jacobs R, Bornstein MM, Odri GA, Kwon MS, Lambrichts I, e.a. A new mandible-specific landmark reference system for three-dimensional cephalometry using cone-beam computed tomography. *Eur J Orthod.* 2016;38(6):563–8.
25. Pittayapat P, Jacobs R, Bornstein MM, Odri GA, Lambrichts I, Willems G, e.a. Three-dimensional Frankfort horizontal plane for 3D cephalometry: a comparative assessment of conventional versus novel landmarks and horizontal planes. *Eur J Orthod.* 2018;40(3):239-248.
26. Cecil C. Steiner. Cephalometrics for you and me. *Am J Orthod.* 1953;29(10):729–55.
27. Jacobson A. The “Wits” appraisal of jaw disharmony. *Am J Orthod.* 1975;67(2):125–38.
28. Legan HL, Burstone CJ. Soft tissue cephalometric analysis for orthognathic surgery. *J Oral Surg Am Dent Assoc* 1965. 1980;38(10):744–751.
29. Ricketts RM. Divine proportion in facial esthetics. *Clin Plast Surg.* 1982;9(4):401–22.
30. Holdaway RA. A soft-tissue cephalometric analysis and its use in orthodontic treatment planning. Part I. *Am J Orthod.* 1983;84(1):1–28.
31. Holdaway RA. A soft-tissue cephalometric analysis and its use in orthodontic treatment planning. Part II. *Am J Orthod.* 1984;85(4):279–93.
32. Arnett GW, Bergman RT. Facial keys to orthodontic diagnosis and treatment planning. Part I. *Am J Orthod Dentofac Orthop.* 1993;103(4):299–312.

33. Arnett GW, Jelic JS, Kim J, Cummings DR, Beress A, Worley Jr CM, et al. Soft tissue cephalometric analysis: diagnosis and treatment planning of dentofacial deformity. *Am J Orthod Dentofacial Orthop.* 1999;116(3):239–253.
34. Kiekens RMA, Maltha JC, van 't Hof MA, Kuijpers-Jagtman AM. A measuring system for facial aesthetics in Caucasian adolescents: reproducibility and validity. *Eur J Orthod.* 2005;27(6):579–84.
35. Kiekens RMA, Kuijpers-Jagtman AM, van 't Hof MA, van 't Hof BE, Maltha JC. Putative golden proportions as predictors of facial esthetics in adolescents. *Am J Orthod Dentofac Orthop.* 2008;134(4):480–3.
36. Kiekens RMA, Kuijpers-Jagtman AM, van 't Hof MA, van 't Hof BE, Straatman H, Maltha JC. Facial esthetics in adolescents and its relationship to “ideal” ratios and angles. *Am J Orthod Dentofac Orthop.* 2008;133(2):188.e1-8.
37. Halazonetis DJ. Morphometric correlation between facial soft-tissue profile shape and skeletal pattern in children and adolescents. *Am J Orthod Dentofac Orthop* 2007;132(4):450–7.
38. Halazonetis DJ. Morphometric evaluation of soft-tissue profile shape. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod.* 2007;131(4):481–9.
39. Moyers RE, Bookstein FL. The inappropriateness of conventional cephalometrics. *Am J Orthod.* 1979;75(6):599–617.
40. Bookstein FL. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med Image Anal.* 1997;1(3):225–243.
41. Bookstein FL. *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge University Press; 1997.
42. Weber DW, Fallis DW, Packer MD. Three-dimensional reproducibility of natural head position. *Am J Orthod Dentofac Orthop.* 2013;143(5):738–44.
43. Barbera AL, Sampson WJ, Townsend GC. Variation in natural head position and establishing corrected head position. *Homo Int Z Vgl Forsch Am Menschen.* 2014;65(3):187–200.
44. Zebeib AM, Naini FB. Variability of the inclination of anatomic horizontal reference planes of the craniofacial complex in relation to the true horizontal line in orthognathic patients. *Am J Orthod Dentofac Orthop.* 2014;146(6):740–7.
45. Santos RMGD, De Martino JM, Haiter Neto F, Passeri LA. Influence of different setups of the Frankfort horizontal plane on 3-dimensional cephalometric measurements. *Am J Orthod Dentofac Orthop.* 2017;152(2):242–9.
46. Goodall C. Procrustes methods in the statistical analysis of shape. *J R Stat Soc Ser B Methodol.* 1991;285–339.
47. Dryden IL, Mardia KV. *Statistical Shape Analysis.* Wiley; 1998.
48. Zelditch ML, Swiderski DL, Sheets HD. *Geometric Morphometrics for Biologists: A Primer.* Academic Press; 2012.
49. Ivan Perez S, Bernal V, Gonzalez PN. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *J Anat.* 2006;208(6):769–84.

50. Gunz P, Mitteroecker P. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix Ital J Mammal*. 2013;24(1):103–9.
51. Bookstein FL. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Trans Pattern Anal Mach Intell*. 1989;11(6):567–85.
52. Bookstein FL, Streissguth AP, Sampson PD, Connor PD, Barr HM. Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. *NeuroImage*. 2002;15(1):233–51.
53. Coquerelle M, Bookstein FL, Braga J, Halazonetis DJ, Weber GW, Mitteroecker P. Sexual dimorphism of the human mandible and its association with dental development. *Am J Phys Anthropol*. 2011;145(2):192–202.
54. Rohlf FJ, Slice D. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Syst Zool*. 1990;39(1):40–59.
55. Bookstein FL. The Inappropriate Symmetries of Multivariate Statistical Analysis in Geometric Morphometrics. *Evol Biol*. 2016;43(3):277–313.
56. Bookstein FL. A method of factor analysis for shape coordinates. *Am J Phys Anthropol*. 2017;164(2):221–45.
57. Bookstein FL. *Measuring and Reasoning: Numerical Inference in the Sciences*. Cambridge University Press; 2014.
58. Sassouni V, Nanda S. Analysis of dentofacial vertical proportions. *Am J Orthod*. 1964;50(11):801–823.
59. Halazonetis DJ. Morphometrics for cephalometric diagnosis. *Am J Orthod Dentofacial Orthop*. 2004;125(5):571–81.
60. Peres-Neto PR, Jackson DA, Somers KM. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Comput Stat Data Anal*. 2005;49(4):974–997.



CHAPTER 7

Summary / Samenvatting



Summary

In **chapter 1** the background and rationale of this PhD study are explained. Orthodontists still struggle to efficiently distill diagnostic data from cephalometric images and to demonstrate the usefulness of lateral cephalometry in orthodontics. Many explanations have been offered to explain the difficulties associated with analyzing lateral cephalograms, which can broadly be categorized as technical problems, which together complicate accurate and reliable identification of landmarks, and analytical problems associated with the choice and definition of (reference) landmarks and normative values. Another, often overlooked category is geometrical distortion, which might be defined as ‘the geometric phenomenon allowing patients with seemingly identical mandibulo-maxillary relationships to exhibit markedly different cephalometric values.

The central premise of this thesis is that the role of geometrical distortion in lateral cephalometry can be understood more clearly by considering the land-surveyor analogy. Land surveyors also make extensive use of angular and linear measurements to determine, among others, property boundaries. However, when attempting to apply some of our traditional approaches for assessing sagittal intermaxillary relationships to the cadastral survey (the ANB angle and Wits appraisal), it becomes intuitively clear that inter-individual variation in the location of the associated reference landmarks and planes essentially forms the basis of geometrical distortion. The absence of any tools to avoid positional uncertainty of the reference structures essentially precludes any definitive statement as to the validity of the resulting measurements.

The land-surveyor analogy suggests that geometric distortion might constitute one of lateral cephalometry’s more fundamental problems: what exactly is the relevance of landmarking error in lateral cephalometry if the appropriateness of the measurement’s reference landmarks is doubtful to begin with? Secondly, the traditional approach to problem-solving in cephalometrics, simply moving around the ‘cephalometric measuring-tripod’ to a different set of landmarks or planes to perform linear or angular measurements from, would seem to serve little purpose other than shifting positional uncertainty from one set of reference landmarks to the next. Land surveyors cleverly avoid this measurement conundrum by employing a fixed, external reference frame, consisting of benchmarks (i.e. points of which the location has been determined highly accurately in three spatial dimensions). This begs the question whether lateral cephalometry could potentially benefit from imitating the land-surveyors’ approach, by performing the measurements from a fixed reference frame instead of an inter-individual variable one.

Chapter 2 provides a relatively simple test of the potential merits of the proposed method by comparing the traditional cephalometric measurements to those obtained when exchanging the patient's reference points and/or planes with those of a Procrustes superimposed template (the 12-year male-female Bolton template): the 'normalized' measurements. The conventional and normalized values were calculated in 71 patients (26 males: mean age 13.1 years, SD 1.1 years; 45 females: mean age of 14.6 years, SD 8.2 years). The measurements involved were the ANB angle and Wits appraisal, the individualized ANB angle according to Hussels and Nanda, Järvinen's floating norm, the APDI (antero-posterior dysplasia index, introduced by Kim and Vietas), perpendicular projections of points A and B onto Hall-Scott's maxillo-mandibular bisector, similar projections on palatal plane (as proposed by Ferrazzani), on Frankfort horizontal plane (introduced by Chang), and on the SN-line (Taylor), as well as and Downs' AB plane angle. A considerable increase was observed in the correlation between the "normalized" measurements, in comparison to the conventional counterparts. As an example, the correlation between the conventional ANB angle and Wits appraisal was a moderate $\rho=0.624$, compared to $\rho=0.972$ for the normalized measurements. The increased correspondence between the normalized analyses improved the chances of both tests agreeing on the patient's sagittal discrepancy. Albeit no true measure of diagnostic performance, the improved correlations thus decreased the possibility of differing, or even totally opposing diagnostic outcomes resulting from their application to non-borderline patients. The proposed methodology therefore seemed to merit further investigation.

The decision to use the Bolton 12-year male-female averaged template in determining the normalized measurements was quite arbitrary and begs the question whether a more population-specific reference frame could be developed, suitable for North-European patients. Further questions arose: should a different reference frame be applied for male and female patients, or for adults versus children? This required scrutinizing and characterizing the patterns of craniofacial variation of the target population, as described in **chapter 3**.

One hundred and seventy eight orthodontic patients (79 male, and 99 female) were collected between the ages of 7.5 and 40 years old. Sixteen skeletal landmarks were digitized in each patient, after which the resulting configurations were subjected to generalized Procrustes superimposition. The male and female subgroups were tested for differences in mean shapes and ontogenetic trajectories. The latter pertains to changes in shape, resulting from changes in size. Size, in this context, serves as a proxy for 'growth and development'. Shape variability was characterized using principal component analysis, applied to the Procrustes superimposed landmark configurations. Furthermore, six different scenarios for craniofacial modularity were tested.

The results showed that there were no significant differences in the male and female Procrustes mean shapes ($p=0.33$), although males were on average found to be significantly larger ($p<0.001$). Mild sexual ontogenetic allometric divergence was noted, although the spherical scatter of the male-female point clouds limited the ability to draw definitive conclusions, probably as a result of the cross-sectional nature of the patient sample, without clearly separated age classes. The same sphericity likewise obscured shape differences between older and younger individuals. When controlling for the effects of allometry, the male-female shape difference became statistically highly significant, albeit clinically still very subtle and probably insignificant. Principal component analysis indicated that of the four retained biologically interpretable components, the two most important sources of variability were vertical shape variation (i.e. dolichofacial vs. brachyfacial growth patterns) and sagittal relationships (maxillary prognathism vs. mandibular retrognathism, and vice versa). Additionally, the presence of an anterior and posterior craniofacial columnar module was confirmed, separated by the pterygomaxillary plane, as proposed by Enlow. These modules can be further subdivided into four sub-modules, involving the posterior skull base, the ethmomaxillary complex, a pharyngeal module, and the anterior part of the jaws. In conclusion, this study provided a population specific reference frame (the pooled sample mean shape) and quantified the associated shape variation. It also provided evidence that, at least for diagnostic purposes, the use of a single, pooled reference frame would seem to make sense. Chapters two and three suggested that the geometric morphometric framework, involving generalized Procrustes superimposition (GPS) and principal component analysis (PCA), might be helpful in solving some of lateral cephalometry's more fundamental problems (i.e. geometric distortion). It was not quite clear however how our traditional cephalometric measures, such as ANB angle, Wits appraisal, or GoGnSN angle, relate to this new tool.

Chapter 4 aimed to demonstrate how the hitherto unclear relationship between the shape space defined by the first two principal components (resulting from the PCA) and the aforementioned traditional cephalometric measures may be established, and to elucidate possible clinical applications thereof. In the process, it was hoped the proposed methodology would provide further support for the land surveyor analogy, by quantifying, and demonstrating visually how the traditional lateral measures represent compound (and often convoluted) measures of craniofacial shape. Two hundred lateral cephalograms were digitized, after which the resulting landmark configurations were subjected to GPS and PCA. The sample mean shape was then deformed along/parallel to principal components (PCs) 1 and 2, recording the resulting ANB, Wits, and GoGnSN value at each location. This allowed calculating trajectories through the PC1–PC2 space connecting locations with identical values. These were finally utilized to renormalize the PC1–PC2 space.

Intriguingly, the resulting Wits appraisal trajectories were almost straight and parallel to PC1. Those for the ANB angle were angled approximately 20 degrees downward relative to PC1, with a more accentuated curvature. The GoGnSN curves were mildly angled relative to the PC2 axis, their curvature increasing slightly with increasing PC1 scores. The trajectories' curvature and slope, and the changing nature thereof over the PC1-PC2 plane, provides further evidence of often quite complex nature of the craniofacial traits measured by the traditional cephalometric measures. By combining the aforementioned trajectories, it was possible to delineate the region of the PC1–PC2 shape space which would be regarded as normodivergent and skeletal Class I according to traditional lateral cephalometry and to contrast this to those defined by the GPS-PCA approach. Geometric distortion could be avoided by assigning patients the ANB, Wits, or GoGnSN value of the sample mean shape, deformed to the patient's position within the PC1–PC2 plot.

As mentioned above, correlation calculations clearly do not represent a true measure of diagnostic performance. Although encouraging, the results in chapter one therefore provided only indirect evidence of the proposed method's merits. Diagnostic performance is usually established using Receiver Operating Characteristic (ROC) curve analysis, which plots the sensitivity (or true positive ratio) versus 1-specificity (or false-positive ratio) for a full range of possible values of the diagnostic test's cut-off value. The area under the resulting curve serves as a measure of diagnostic performance: the larger the surface area under the curve (the closer the curve approaches the upper left corner of the graph), the more powerful the test.

One of the biggest hurdles in the application of ROC curve analysis in lateral cephalometry has always been the fact that it requires a gold standard, providing the correct answer to the diagnostic question. Until recently, the latter simply did not seem exist; a problem for which the introduction of the geometric morphometric framework in orthodontics might have provided a convenient solution, as evident from chapter three. Another potential problem is ROC curve analysis' dichotomous nature, requiring clearly discernible health states in order to provide the black-or-white diagnostic result required to determine the test's diagnostic power, which would seem to align poorly with the continuous spectrum of facial variation present in the (orthodontic patient) population. Additionally, as evident from the curved, sloped trajectories presented in chapter four, the application of a single, static gold-standard cut-off value for the metric under investigation would seem to make little sense. A floating-norm approach to these cut-off points would seem more appropriate.

The aim of **chapter 5** therefore was to assess the diagnostic performance of both the conventional and normalized version of the ANB angle and Wits appraisal using an extended version of Receiver Operating Curve (ROC) analysis which renders ROC

surfaces, instead of curves. The required 'gold standard' was derived statistically, by applying generalized Procrustes superimposition (GPS) and principal component analysis (PCA) to the digitized landmarks, and ordering patients based upon their PC2 scores. The patient sample of chapter four was revisited, consisting of 200 lateral cephalograms (107 males, mean age: 12.8 years, SD: 2.2, 93 females, mean age: 13.2 years, SD: 1.7), which were subjected to GPS and PCA. Upon calculating the conventional and normalized ANB and Wits values, ROC surfaces were constructed by varying not just the cephalometric test's cut-off value within each ROC curve, but also the gold standard cut-off value over different ROC curves in 220 steps between minus two and two standard deviations along PC2. The volume under the resulting ROC surfaces (VUS) served as a measure of overall diagnostic performance. The statistical significance of the volume differences was determined using permutation tests (1000 rounds, with replacement).

Intriguingly, the diagnostic performance of the conventional ANB and Wits was remarkably similar (81.1 and 80.75% VUS, respectively, $p > 0.05$). Normalizing the measurements improved all VUS highly significantly (91 and 87.2 %, respectively, $p < 0.001$). A potentially confusing consequence of changing the gold standard cut-off value as well in ROC surface analysis, is that that in doing so, the conventional ROC curve analysis' three Class problem (Class I, II or III) is effectively turned into a two Class one (less Class II-more Class III, or more Class II-less Class III). Hence only one ROC curve is reported per measure in ROC surface analyses (which combined create a surface) instead of the usual two for ROC curve analysis (one for Class II/I and one for Class I/III). The conclusion of chapter five was that the conventional ANB and Wits measures of sagittal discrepancy do not differ in their diagnostic performance. Normalizing the measurements did seem to have some merit since the improvements were significant, albeit perhaps not spectacular. The latter may be explained quite simply, due to the fact that the first two principal components explained a considerable amount, but still only part of the variability in craniofacial shape.

Chapter 6 discusses some methodological issues, and elucidates future perspectives.

Samenvatting

In **hoofdstuk 1** worden de achtergrond en denkwijze uit de doeken gedaan, die de grondslag vormen voor dit onderzoek. Orthodontisten worstelen er nog steeds mee om op een efficiënte manier gegevens te distilleren uit röntgencefalometrische beelden, en het nut van laterale cefalometrie in de orthodontie staat niet vast. Er zijn verschillende verklaringen voor de moeilijkheden die de analyse van de röntgenschedelprofielfoto (RSP) met zich mee brengt, die in grote lijnen gerubriceerd kunnen worden als enerzijds problemen van technische aard, die het accuraat en betrouwbaar identificeren van de anatomische meetpunten bemoeilijken, en anderzijds analytische problemen, die verband houden met de keuze en omschrijving van de (referentie) punten en normwaarden. Een derde categorie, die ogenschijnlijk vaak over het hoofd wordt gezien, is geometrische vervorming. Deze laatste zou men kunnen definiëren als “het fenomeen waarbij patiënten met een ogenschijnlijk identieke sagittale verhouding tussen de onder- en bovenkaak, toch zeer verschillende (relevante) cefalometrische waarden kunnen vertonen”. De rode draad in dit proefschrift is dat de rol die geometrische vervorming speelt in laterale röntgencefalometrie, beter begrepen kan worden door de verrassende analogieën in acht te nemen die er lijken te bestaan met een zogenaamde ‘cadastale landmeting’ (de landmeter-analogie). Landmeters maken evenzeer overvloedig gebruik van afstands- en hoekmetingen om (onder andere) nauwkeurig de begrenzing van eigendommen te lokaliseren. Wanneer we echter trachten sommige traditionele technieken voor het beoordelen van de intermaxillaire sagittale verhouding (zoals de ANB hoek of de Wits-meting) toe te passen in het kader van landmeting, wordt intuïtief duidelijk dat interindividuele variatie in de locatie van de referentiemeetpunten en -vlakken aan de basis ligt van geometrische vervorming. Het ontbreken van technieken om de onzekerheid betreffende de positie van de referentiestructuren te vermijden, verhindert het trekken van definitieve conclusies betreffende de validiteit van de metingen.

De landmeter-analogie suggereert dat geometrische vervorming tot de meer fundamentele problemen van laterale röntgencefalometrie behoort. Wat is immers het werkelijke belang van meetfouten bij het digitaliseren van röntgenfoto's, indien er twijfels zouden kunnen bestaan omtrent de validiteit van de referentiepunten en -vlakken die bij deze meting worden gebruikt? Ten tweede blijkt hieruit dat de traditionele probleem-oplossende strategie in de laterale röntgencefalometrie, namelijk het eenvoudigweg verplaatsen van het ‘cefalometrische meetstatief’ naar een andere reeks referentiemeetpunten en -vlakken, weinig zoden aan de dijk zet. Men verplaatst er enkel de positionele onzekerheid mee van de vorige verzameling referentiestructuren naar de volgende. Landmeters vermijden dit meetprobleem handig door gebruik te maken van een extern referentiekader bestaande uit “benchmarks”: meetpunten waarvan de positie vooraf met hoge nauwkeurigheid werd bepaald in drie dimensies. Daarbij komt

de vraag op of laterale röntgencefalometrie gebaat zou kunnen zijn bij het imiteren van deze aanpak, door gebruik te maken van een vast referentiekader, in plaats van ééntje dat interindividueel varieert.

Hoofdstuk 2 beschrijft een relatief eenvoudige test voor het mogelijke potentieel van de voorgestelde methodologie, waarbij de traditionele cefalometrische metingen vergeleken worden met die welke verkregen worden wanneer de referentie punten en vlakken van de patiënt vervangen worden door deze van een gesuperponeerd sjabloon (de '12 year male-female Bolton template'): de zgn. genormaliseerde metingen. De traditionele en genormaliseerde waarden werden berekend in 71 patiënten (26 mannen, gemiddelde leeftijd 13.1 jaar, SD 1.1 jaar; 45 vrouwen, gemiddelde leeftijd 14.6 jaar, SD 8.2 jaar). De betrokken metingen waren de ANB hoek en Wits meting, de geïndividualiseerde ANB hoek volgens Hussels en Nanda, Järvinen's glijdende norm, de APDI (antero-posterieure dysplasie index, geïntroduceerd door Kim en Vietas), loodrechte projecties op Hall-Scott's maxillo-mandibulaire bissectrice, gelijkaardige projecties op het palatale vlak (zoals voorgesteld door Ferrazzani), op het Frankfort horizontale vlak (voorgesteld door Chang) en op de SN-lijn (Taylor) alsook Downs' AB plane angle. Er werd een aanzienlijke toename genoteerd in de correlatie tussen de genormaliseerde waarden, in vergelijking met de overeenkomstige traditionele metingen. Ter illustratie bedroeg de correlatie tussen de traditionele ANB hoek en Wits meting een matige $p=0.624$, in vergelijking met $p=0.972$ voor de genormaliseerde waarden. De toegenomen correlatie verhoogde daarbij de kans dat beide metingen het eens waren omtrent de sagittale discrepantie van de patiënt. Hoewel het geen werkelijke meting betrof van diagnostische performantie, verminderde de verbeterde correlatie op die manier de kans dat verschillende of zelfs contradictorische diagnostische resultaten voortkwamen uit hun toepassing op non-bordeline patiënten. Gezien het voorgaande leek het derhalve gerechtvaardigd de methodologie aan verder onderzoek te onderwerpen.

De beslissing om de Bolton 12-jarige unisex template te benutten als vast referentiekader voor de bepaling van de genormaliseerde metingen in het vorige onderzoek was vrij arbitrair, waarbij zich de vraag stelt of het niet mogelijk is een populatie-specifieker referentiekader te ontwikkelen, toepasbaar op Noord-Europese patiënten. Bovendien kan men zich afvragen of het wel wenselijk of gerechtvaardigd is éénzelfde template te gebruiken voor mannen en vrouwen, of voor volwassenen en kinderen. Om een antwoord te bieden op deze vragen bleek het noodzakelijk de craniofaciale variatie binnen de doelpopulatie uitvoerig te typeren, zoals beschreven in **hoofdstuk 3**.

Bij 178 patiënten (79 mannen en 99 vrouwen) tussen de 7.5 en 40 jaar oud, werden zestien skelettale meetpunten gedigitaliseerd, waarna de resulterende configuraties onderworpen werden aan Procrustes superpositie. De mannelijke en vrouwelijke subgroepen werden

vervolgens getest op verschillen in hun gemiddelde vorm (tenzij uitdrukkelijk anders vermeld hebben we het, wanneer in de hierna volgende tekst gerefereerd wordt naar “vorm” steeds over shape, en niet over form) en de zogenaamde ontogenetische paden (ontogenetic trajectories), waarbij deze laatste betrekking hebben op veranderingen in vorm die gepaard gaan met veranderingen in grootte (grootte wordt in deze context overigens gebruikt als een vrij zwak substituuut voor “groei en ontwikkeling”). De variatie in vorm werd geanalyseerd aan de hand van hoofdcomponentenanalyse (principal component analysis), toegepast op de Procrustes gesuperponeerde configuraties van meetpunten. Verder werden ook zes verschillende scenario’s getest voor modulariteit (modularity). De resultaten lieten zien dat er geen significante verschillen bestaan in de gemiddelde gelaatsvorm van mannen en vrouwen ($p=0.33$), hoewel mannen gemiddeld significant groter bleken te zijn ($p<0.001$). Niettegenstaande er een milde divergentie vastgesteld werd tussen de mannelijke en vrouwelijke ontogenetische paden (ontogenetic trajectories), was het moeilijk hieromtrent definitieve conclusies te trekken door de eerder sferische verdeling van de beide puntenwolken. Deze verdeling was waarschijnlijk een rechtstreeks gevolg van het cross-sectionele karakter van de steekproef, zonder duidelijk gescheiden leeftijdscategorieën. Dezelfde sferische verdeling heeft waarschijnlijk bijgedragen aan het maskeren van vormverschillen tussen jongere en oudere patiënten. Wanneer gecontroleerd werd voor de effecten van allometrie, werden de man-vrouw vormverschillen plots hoog significant, hoewel nog steeds zeer subtiel en waarschijnlijk niet relevant.

De hoofdcomponenten-analyse bracht aan het licht dat van de vier behouden hoofdassen, de twee belangrijkste bronnen van variabiliteit enerzijds brachyfaciale versus dolichofaciale groeipatronen bleken te zijn (m.a.w. “verticale” variatie), en anderzijds relatieve mandibulaire prognatie en retrognatie (“horizontale” variatie). Verder werd de aanwezigheid bevestigd van een anterieure en posterieure “columnaire module” gescheiden door het pterygomaxillaire vlak, zoals voorgesteld door Enlow. Deze modules konden verder onderverdeeld worden in vier sub-modules, waaronder een posterieure module (die de posterieure schedelbasis bevat), het ethmo-maxillaire complex, een pharyngeale module, en het anterieure gedeelte van de kaken. Samenvattend werd in hoofdstuk drie een populatie-specifiek referentiekader voorgesteld (de gemiddelde vorm van de samengevoegde steekproef). Tevens werd de hiermee geassocieerde vormvariatie gekwantificeerd. Hoofdstuk drie levert ook bewijzen op voor de stelling dat, ten minste voor diagnostische doeleinden, het gebruik van een enkel, unisex referentiekader gerechtvaardigd is.

In hoofdstukken 2 en 3 werd gesuggereerd dat het geometrisch-morfometrisch kader, dat onder meer gebaseerd is op Procrustes superpositie en hoofdcomponentenanalyse, van nut zou kunnen zijn bij het oplossen van sommige van de meer fundamentele

problemen met betrekking tot laterale röntgencefalometrie (met name geometrische vervorming). Het is echter onduidelijk hoe sommige van de traditionele cefalometrische variabelen, zoals de ANB hoek, Wits meting, en de GoGnSN hoek, zich verhouden tot deze methodologie.

Hoofdstuk 4 had dan ook tot doel aan te tonen hoe de (voorheen onbekende) relatie tussen de vormruimte (shape space) gedefinieerd door de eerste twee hoofdcomponenten resulterend uit de hoofdcomponentenanalyse, en de eerder vermelde cefalometrische variabelen kan worden bepaald. Tevens was het de bedoeling mogelijke klinische toepassingen te illustreren. De hoop was dat de technieken voorgesteld in hoofdstuk 4 de landmeter-analogie verder wetenschappelijk zouden onderbouwen door te kwantificeren en visualiseren in hoeverre onze traditionele cefalometrische variabelen samengestelde en vaak erg gecompliceerde aspecten van craniofaciale vorm meten.

Tweehonderd RSPs werden gedigitaliseerd, waarna de resulterende meetpunt-configuraties onderworpen werden aan Procrustes superpositie en hoofdcomponenten-analyse. De gemiddelde vorm van de steekproef (sample mean shape) werd vervolgens vervormd langs (parallel met) de eerste en tweede hoofdcomponenten (principal components PCs 1 en 2), waarbij de resulterende ANB hoek, Wits meting en GoGnSN hoek werd bepaald op iedere locatie. Dit liet ons toe trajecten te berekenen door de PC1-PC2 ruimte die locaties verbinden met identieke cefalometrische waarden. Deze werden uiteindelijk gebruikt om de PC1-PC2 ruimte te renormaliseren.

Intrigerend genoeg bleken de trajecten voor de Wits meting vrijwel rechtlijnig, en parallel te lopen met de eerste hoofdcomponent (PC1). De curves voor de ANB hoek vertoonden daarentegen een dalende hoek van ongeveer 20 graden ten opzichte van PC1 as, met een meer geaccentueerde kromming. Deze bleken voor de GoGnSN hoek licht gekanteld ten opzichte van de PC2 as, waarbij de kromming licht toenam met stijgende PC1 waarden. De eerder vermelde kromming van de berekende trajecten, alsook hun hoek ten opzichte van de hoofdcomponent-assen, vormt verder bewijs voor de vaak erg complexe aard van de craniofaciale kenmerken die door de traditionele cefalometrische variabelen worden gemeten. Door het combineren van de eerder vermelde trajecten bleek het tevens mogelijk die zones van het PC1-PC2 vlak af te lijnen die als normodivergent en Klasse I worden beschouwd door de traditionele cefalometrie, en deze te vergelijken met de overeenkomstige zones gedefinieerd door de nieuw voorgestelde methodologie. Geometrische vervorming zou mogelijk vermeden kunnen worden door de gemiddelde vorm van de steekproef (sample mean shape) te vervormen naar de positie van de patiënt in het PC1-PC2 vlak, en de resulterende waarde toe te kennen aan deze patiënt.

Zoals eerder werd vermeld, vormen correlatieberekeningen geen adequate maatstaf voor diagnostische performantie. De bemoedigende resultaten van hoofdstuk één vormden derhalve slechts een indirect bewijs voor de mogelijke toegevoegde waarde geboden door de voorgestelde methodologie. Diagnostische performantie wordt doorgaans bepaald aan de hand van Receiver Operating Characteristic (ROC) curve analysis (“ROC analyse”), waarbij de gevoeligheid (of “terecht positieve ratio”) uitgezet wordt ten opzichte van de specificiteit (of “foutief positieve ratio”) over een volledig reeks mogelijke cut-off waarden voor de diagnostische test. De oppervlakte onder de resulterende curve vormt de maatstaf voor diagnostische performantie: hoe groter deze oppervlakte (hoe dichter deze curve de linker- bovenhoek van de grafiek benadert), hoe performanter de test.

Eén van de grootste hinderpalen bij de toepassing van ROC analyse in laterale cefalometrie is traditioneel steeds de vereiste beschikbaarheid gebleken van een gouden standaard, die het correcte antwoord biedt op de diagnostische vraag. Tot voor kort leek deze namelijk eenvoudigweg niet te bestaan; een probleem waarvoor de introductie van geometrische morfometrie in de orthodontie mogelijk een geschikte oplossing biedt, zoals uiteengezet in hoofdstuk drie. Een ander potentieel probleem is het strikt tweeledig karakter van ROC analyse, waarbij scherp afgelijnde of eenduidige gezondheidstoestanden vereist zijn om de zwart-witte diagnostische resultaten aan te leveren waarmee de diagnostische performantie bepaald kan worden. Dit laatste lijkt echter moeilijk te rijmen met het eerder continue spectrum van craniofaciale variatie dat aangetroffen wordt in de (orthodontische patiënten-) populatie. Het gebogen, stijgende of dalende verloop van de berekende trajecten in hoofdstuk 4 suggereert bovendien dat het gebruik van een enkele, statische cut-off waarde in de laterale cefalometrie weinig steek houdt. Het gebruik van een zogenaamde “glijdende-schaal” benadering met betrekking tot deze cut-off waarden lijkt derhalve zinvoller.

Het doel van **hoofdstuk 5** was derhalve om de diagnostische performantie te onderzoeken van zowel de traditionele als de genormaliseerde versies van de ANB hoek en Wits meting, aan de hand van een uitgebreide versie van ROC analyse waarbij een ROC oppervlak wordt gegeneerd, in plaats van een curve. De vereiste gouden standaard werd hierbij statistisch bepaald, door het onderwerpen van de gedigitaliseerde configuraties van meetpunten aan Procrustes superpositie en hoofdcomponentenanalyse, en het sorteren van patiënten op basis van hun PC2 score. De steekproef uit hoofdstuk 4, bestaande uit 200 laterale schedelprofielopnames, werd herbekeken (107 mannen, gemiddelde leeftijd: 12.8 jaar, SD: 2.2, 93 vrouwen, gemiddelde leeftijd: 13.2 jaar, SD: 1.7) en onderworpen aan Procrustes superpositie en hoofdcomponentenanalyse.

Na het berekenen van de conventionele en genormaliseerde ANB en Wits waarden, werden ROC oppervlakken berekend, waarbij niet alleen de cut-off waarde van de test werd gevarieerd binnen elke ROC curve, maar tevens de cut-off waarde van de gouden standaard, over de verschillende ROC curves (tussen plus en minus 2 standaard deviaties langs PC2, in 220 stappen). Het volume onder de resulterende ROC oppervlak (Volume Under the Surface, of VUS) fungeerde daarbij als maatstaf voor diagnostische performantie. De statistische significantie van het volumeverschil tussen de verschillende ROC oppervlakken werd berekend aan de hand van permutatietests (1000 permutaties, met replacement of hergebruik van de variabelen). Intrigerend genoeg bleek de diagnostische performantie van de conventionele ANB hoek en Wits meting zeer vergelijkbaar (81.1 en 80.75% VUS respectievelijk, $p > 0.05$). Het normaliseren van de meetwaarden verhoogde de VUS significant (91 en 87.2 % respectievelijk, $p < 0.001$).

Een mogelijk erg verwarrende consequentie van het bijkomend variëren van de cut-off waarde van de gouden standaard in ROC surface analysis is dat daarbij het 3 klassen-probleem van de klassieke ROC curve analyse (Klasse I, II of III), effectief omgezet wordt in een twee-klasse probleem (minder Klasse II - meer Klasse III, of het omgekeerde). Daarom wordt bij ROC surface analysis slechts één resultaat getoond per test, in plaats van de gebruikelijke twee bij ROC curve analyse (één voor Klasse II/I en één voor Klasse I/III). De conclusie van hoofdstuk 5 was dat de conventionele ANB hoek en Wits waarde niet verschillen in hun diagnostische performantie. Het normaliseren van de meetwaarden leek zinvol, vermits de resultaten significant bleken, hoewel mogelijk niet spectaculair. Dit laatste kan eenvoudig verklaard worden door het feit dat hoewel de eerste twee hoofdcomponenten een aanzienlijk deel verklaren van de variatie, dit nog steeds slechts een deel betreft van de totale craniofaciale variatie.

In **hoofdstuk 6** worden een aantal methodologische kwesties verder besproken, en wordt tevens getracht te illustreren hoe dit onderzoek zich in de toekomst verder zou kunnen ontwikkelen.



Acknowledgements
Curriculum Vitae
List of publications
Awards and honors

Acknowledgements

A project such as this one, spanning the greater part of a decade from inception to completion, could never have been carried through without the unwavering support of a sizable number of people.

The support crew:

It cannot be emphasized enough how instrumental such support is when research is involved. The countless hours spent reading, planning, collecting data, investigating, coding and writing, place a significant burden on one's family, particularly when attempting to juggle and balance these tasks with family life and a bustling private practice. As special witnesses to this lengthy PhD project, my wife and children were able to share the joys and thrills brought about by a successfully finished project, a manuscript appearing online or in print, meeting new people through collaboration, talks, or research-related correspondence, learning, creating, etc. But they also witnessed the aggravation/frustration caused by investigative dead ends, disappointing results, intractable coding errors or bewildered reviewers. Most of all, they were inevitably sometimes left to pick up the slack, brought about by dad researching ... things. My family's patience, understanding and support however encouraged me to see this project through, for which I am eternally grateful.

The instigator:

Many years ago, I started my journey in orthodontic research with a project pertaining to arch form. While preparing to analyze the accumulated data, I came into contact with Professor Ellen A. BeGole, who graciously provided me with some manuscripts to which I had trouble getting access to. She also spontaneously offered to proofread my work, and gave me some invaluable pointers with regard to the statistics. We continued regularly corresponding by e-mail, which eventually developed into a years-long close friendship and multiple visits to her home in Wheaton, Chicago. I planned on impressing my friend and mentor during the public defense of the PhD project she inspired me to undertake ages ago. Heartbreakingly, not long after I invited her over, Professor BeGole unexpectedly passed away. Now it seems there is little I can do for my dear friend, other than to honor her memory. I am truly thankful for the small part I was allowed to play in the life of this remarkable lady.

Ellen Ann BeGole was born in Detroit in 1934. Her father was an engineer and inventor, who held multiple patents. Perhaps not surprisingly, his daughter exhibited a similar knack for mathematics. During the 1950's, women were still generally expected to pursue 'less challenging' careers, if any. Never one to choose the easy road in life, Ellen decided to dedicate hers to mathematics, obtaining a bachelor's degree from the University of

Michigan in 1955, and her master's from Columbia University in 1969. At that time, she thus became one of the few women in the US with a degree in mathematics, on track to becoming an even rarer breed, by setting her heart on an academic career in biostatistics. Six consecutive years of research earned her a PhD from the University of Pittsburgh in 1974. She settled at the University of Florida for two years (where she played the carillon in her spare time), after which she finally found her home at the University of Illinois at Chicago. Here, she worked as a biostatistician from 1975 up to her retirement in 2000, and even beyond as Professor Emeritus.

Prof. Dr. Ellen A. BeGole.

166



An adventurous spirit, Ellen enjoyed travelling, visiting new places and making new friends along the way. A long road trip through Europe proved excellent for this purpose. She would not be fazed by the lack of company, nor by the anything but straightforward nature of travelling at the time. But her appetite for adventure and her interest in other cultures really surfaced when she won a Fulbright scholarship in 1998, allowing her to teach as at the King Abdul Aziz University, Jeddah, Saudi Arabia for five months as a visiting Professor; a time she remembered as one of the true highlights in her career. She would return to the middle east in April 2000, this time teaching at Kuwait university. Her general interest in people was also readily apparent during her time at the UIC, where she mentored an impressive number of orthodontic graduates throughout their training

and into PhD projects. She enjoyed interacting with these promising young individuals of diverse origin, and made a point of knowing everyone personally and expressing a genuine interest in their hopes and worries. It would not be an exaggeration to state she was somewhat of a mother figure to many of them, who was always there in case of trouble or just as well to simply catch up. She also had a very keen sense of humor, who in her spare time enjoyed reading, woodworking and assembling intricately finished clocks.

During her long and productive career, she (co-)authored more than 150 papers and manuscripts, the last one as recent as 2017 (at 82 years old), demonstrating that at any time in her career, her mind was like her pencils: always sharp. One of her well-known book contributions was the 'Statistics for the Orthodontist' chapter in 'Orthodontics: Current Principles and Techniques' (2nd, 3rd and 4th edition). She was also a gifted speaker, which spurred the DePaul Geographical Society to invite her to talk about her experiences in the Middle East for their 40th anniversary year (Desert Odyssey: An American Woman in Saudi Arabia, February 10, 2001), while her presentation on 'Evidence based Orthodontics' for the Illinois Orthodontic Alumni Association in 2004 earned her a standing ovation. Scientifically, she was very forward-thinking, as evidenced by the ubiquitous use of computers, (cubic) splines, and principal component analysis in her early research projects, all of which play prominent roles in the current state-of-the-art statistical study of shape. She also contributed more than her share to the orthodontic community by reviewing an incomprehensible number of manuscripts for (amongst others) the AJO-DO, the Angle Orthodontist, the World Journal and Orthodontics and the Korean Journal of Orthodontics. Her career is a shining testament to what can be achieved in life when it is lived in the way Ellen liked her science: unbiased, but also level-headed, open-eyed, incurably-curious, and warm-hearted. She was a truly unique and wonderful individual. She is survived by her daughter Jane and her granddaughter Melanie, who seems to have inherited her grandmother's mathematical genes, recently graduating as a physicist with a major in astronomy.

When thinking about those people who were influential in orthodontics, we often remember those who spend lots of time in the limelight, the "rock stars of orthodontics". It is easy to overlook those diligently toiling away in the background to enable others this moment of scientific glory: the ones teaching study design and data analytic techniques, making sure appropriate conclusions are drawn from the experimental data, and advising journals on the applicability of the statistics used in submitted manuscripts. We owe much of the progress that has been made in orthodontics with respect to the improvement of scientific standards to these mathematically trained individuals. We owe it to them, these less vocal 'giants of orthodontics', to never forget... .

This In Memoriam appeared in the June 2018 issue of the AJO-DO.

The enablers:

I owe a special thanks to both my Promotor, Prof. Dr. A.M. Kuijpers-Jagtman, and my friend and co-author, Prof. Dr. D. Halazonetis. I am very grateful to Professor Kuijpers-Jagtman's for her wide-eyed, research domain-spanning, unbiased view on science in general, and the people behind it in particular. Her uncharacteristically open-minded, level-headed approach to science, and her complacency-averse personality, helped to make this PhD project a most memorable and pleasurable learning experience. I thank her for giving me this opportunity, and for her help and support along the years. Intriguingly, many of the same qualities with respect to world-view, research and personality can be attributed to my co-author Prof. Halazonetis, without whom the pertaining manuscript would never have turned out the way it has. One of the most intelligent orthodontists on earth (sometimes annoyingly so), I regard the back-and-forth correspondence as part of the preparations for our mutual publication as the most pleasant, interesting and fun times of the entire project, for which I am genuinely grateful. I sometimes wish I had more time and opportunity to pick his interesting brain (in a purely non-pathologist sort of way). But who knows what promises the future holds in terms of joint projects.

Curriculum Vitae

Hans Wellens was born in 1973 in Wilrijk, Belgium. In 1996, he obtained his DDS from Ghent University, with honors.

A certified orthodontist with a full-time two-chair private practice in Sint-Michiels, Bruges, Belgium, he started his doctoral project at the Department of Dentistry, section of Orthodontics and Craniofacial Biology at Radboud University Medical Centre Nijmegen, the Netherlands in 2009, focusing on the applications of Procrustes analysis in orthodontic diagnosis and treatment planning.

He is a reviewer for the American Journal of Orthodontics and Dentofacial Orthopedics, and the European Journal of Orthodontics. Aside from geometric morphometrics, he has an active interest in stereophotogrammetry, having developed a custom built, 14 DSLR camera setup for three-dimensional facial capture.

List of publications

Publications related to this thesis

- Wellens H. "Improving the concordance between various anteroposterior cephalometric measurements using Procrustes analysis", *European Journal of Orthodontics*, 31 (2009): 503-15.
- Wellens HL, Kuijpers-Jagtman AM, Halazonetis DJ. "Geometric morphometric analysis of craniofacial variation, ontogeny and modularity in a cross-sectional sample of modern humans", *Journal of Anatomy*, 222 (2013): 397-409.
- Wellens HL, Kuijpers-Jagtman AM. "Connecting the new with the old: modifying the combined application of Procrustes superimposition and principal component analysis, to allow for comparison with traditional lateral cephalometric variables", *European Journal of Orthodontics*, 38 (2016): 569-576.
- Wellens HLL, BeGole EA, Kuijpers-Jagtman AM. "ROC surface assessment of the ANB angle and Wits appraisal's diagnostic performance with a statistically derived 'gold standard': does normalizing measurements have any merit?", *European Journal of Orthodontics*, 39 (2017):358-364.

Other peer reviewed publications:

- Wellens H. "Applicability of mathematical curve-fitting procedures to late mixed dentition patients with crowding: a clinical-experimental evaluation", *American Journal of Orthodontics and Dentofacial Orthopedics*, 131 (2007), 160.e17-25.
- Wellens H. "A clinical experimental simulation of changes in intercanine width associated with the correction of crowding: a pilot study", *European Journal of Orthodontics*, 29 (2007):632-638.

Abstracts and presentations at scientific meetings:

- 2008: Oral presentation at the "3-D Virtual Diagnosis and Treatment Planning of Cranio-Maxillo-Facial Deformity" congress, Bruges, with "Improving the concordance between the ANB angle and Wits appraisal using Procrustes analysis".
- 2008: Invited lecture, University of Chicago at Illinois: "Assessing sagittal discrepancy: finding a cure for the diagnosis" (1,5 hours).

- 2014: Invited lecture, University of Chicago at Illinois: “New Vistas in Cephalometrics” (3 hours).
- 2016: Oral presentation at the American association of Orthodontics conference, Orlando, Florida: “Connecting the New with the Old: Modifying the Combined Application of Procrustes Superimposition and Principal Component Analysis, to Allow for Comparison with Traditional Lateral Cephalometric Variables”.
- 2016: Oral presentation at the Mediterranean Orthodontic congress, Cyprus: “Assessing the ANB angle and Wits appraisal`s diagnostic performance using true gold standard ROC curve analysis: does normalizing measurements have any merit?”.

Awards and honors:

- 2003: Winner of the scientific contest of the ‘Tandartsenkrant’ (‘Dentist News’) with the article “Ectopische eruptie van de hoektand: diagnose en interceptie. Een ‘literature-based’ stapsgewijze benadering met behulp van flowcharts.” (appeared januari 2004).
- 2004: Scientific laureate for the VVT (“Society of Flemish dentists”). Awarded a research grant of 12.400 Euro for project “applicability of mathematical curve-fitting procedures to late mixed dentition patients with crowding”.
- 2016: Top reviewer for the American Journal of Orthodontics and Dentofacial Orthopedics

