치의학박사학위논문

Developing a universally applicable prediction model for soft tissue changes after orthognathic surgery based on the sparse partial least squares method

광범위 적용 가능한 sparse partial least squares 기반
턱교정 수술 후 연조직 예측 알고리즘의 개발

2015년 2월

서울대학교 대학원
치의과학과 치과교정학 전공

서 희 연

치의학박사학위논문

Developing a universally applicable prediction model for soft tissue changes after orthognathic surgery based on the sparse partial least squares method

지도교수 이 신 재

이 논문을 치의학박사 학위논문으로 제출함

2014년 11월

서울대학교 대학원

치의과학과 치과교정학 전공

서 희 연

서희연의 치의학박사 학위논문으로 인준함

2014 년 12 월

위 원 장 (인)

부위원장 (인)

위 원 (인)

위 원 (인)

위 원 (인)

- ABSTRACT -

Developing a universally applicable prediction model for soft tissue changes after orthognathic surgery based on the sparse partial least squares method

Hee-Yeon Suh, BS, DDS

Department of Orthodontics, Graduate School, Seoul National University

*(Directed by Professor **Shin-Jae Lee**, DDS, MSD, PhD, PhD)*

Introduction: The aims of the present study are 1) to develop a model for an accurate soft tissue prediction that can be applied to various modes of surgical correction: Class II correction, Class III correction, the mandibular surgery and/or the maxillary surgery, and additional genioplasty; 2) to minimize the number of variables incorporated via SPLS method in order to build a parsimonious and interpretable prediction model.

Materials and methods: The subjects of this study consisted of 318 patients who had undergone the combined surgical-orthodontic correction of skeletal Class II or skeletal Class III malocclusions. The prediction model was composed of 232 independent and 64 dependent variables. Two prediction methods, the PLS method and the SPLS method were compared. In this study, to evaluate the predictive performance, test errors were calculated in absolute

values through the leave-one-out cross-validation method. We promoted sparsity by SPLS method to build a parsimonious and more interpretable prediction model. Leave-one-out cross-validation method was used to determine the optimal SPLS model, i.e., a model selection.

Results: Prediction models by the SPLS method showed prediction performance comparable to prediction model by the PLS method. Since there were no significant differences in prediction performance depending on surgical movement or gender of the subjects, PLS and SPLS models built in present study are universally applicable. The SPLS showed no significant difference in prediction performance until the mean number of selected independent variables was reduced to 34.

Conclusions: It was our observation that both PLS and SPLS methods resulted in accurate prediction of soft tissue change associated with various types of surgical intervention. This study showed that it was possible to build a parsimonious and interpretable prediction model by the SPLS method. The SPLS method was found to be a powerful and objective tool for simplification. Based on our findings, we propose that the SPLS method can provide an improved algorithm in predicting surgical outcomes after orthognathic surgeries.

Key Words: Class II malocclusion, Class III malocclusion, orthognathic surgery, profile prediction, partial least squares method, sparse partial least squares method

Student Number: 2011-30658

국문초록

광범위 적용 가능한 sparse partial least squares 기반 턱교정수술 후 연조직 예측 알고리즘의 개발

서 희 연

서울대학교 대학원 치의과학과 치과교정학 전공

(지도교수: 이 신 재)

연구 목적: 본 연구의 목적은 1) 턱교정수술의 종류에 관계없이 정확하게 수술 후 연조직 변화를 예측할 수 있는 예측 모형을 개발하고, 2) sparse partial least squares (SPLS) 방법을 통해 경제적이고 해석이 보다 용이한 예측 모형을 수립하는 것이다.

재료 및 방법: 본 연구는 서울대학교치과병원에서 교정 및 턱교정수술을 받은 318 명의 골격성 II 급 및 골격성 III 급 부정교합자를 대상으로 하였다. 총 232 개의 독립변수와 64 개의 종속변수(32 개의 술 후 연조직 계측점 좌표)로 예측모형을 수립하였다. PLS 모형과 SPLS 모형의 예측 정확도를 비교하였다. 예측 정확도 비교를 위하여 leave-one-out cross-validation 을 통해 test error 를 절대 값으로 계산하였다. Sparsity 를 부과하는 SPLS 방법을 사용하여 경제적이고 해석이 보다 용이한 예측 모형을 만들고자 하였다. Leave-one-out cross-validation 으로 최적의 SPLS 모형을 선택하였다.

결과: SPLS 방법에 기반한 예측 모형의 예측 정확도는 PLS 기반 모형의 예측 정확도와 유사하였다. 수립된 예측 모형들은 수술의 방향이나 종류, 성별에

관계없이 높은 예측 정확도를 보였다. 이는 본 연구에서 수립된 예측 모형을 수술의 종류에 관계없이 광범위하게 적용할 수 있음을 의미한다. SPLS 기반 예측모형에서 sparsity 를 부과하여 평균적으로 선택되는 독립변수가 34 개로 감소하여도 모든 변수를 투입한 PLS 모형에 비하여 예측 정확도에 있어 임상적으로 유의한 차이를 보이지 않았다.

결론: PLS 와 SPLS 방법 모두 수술의 종류에 관계없이 높은 예측 정확도를 보였으며 SPLS 방법을 통해 정확하면서도 보다 해석이 용이하고 경제적인 턱교정수술 환자의 연조직 예측 모형을 수립할 수 있었다.

주요어: II 급 부정교합, III 급 부정교합, 턱교정수술, 연조직 변화 예측, partial least squares 방법, sparse partial least squares 방법

학번: 2011-30658

Developing a universally applicable prediction model for soft tissue changes after orthognathic surgery based on the sparse partial least squares method

Hee-Yeon Suh, BS, DDS

Department of Orthodontics, Graduate School, Seoul National University

*(Directed by Professor **Shin-Jae Lee**, DDS, MSD, PhD, PhD)*

광범위 적용 가능한 sparse partial least squares 기반 턱교정 수술 후 연조직 예측 알고리즘의 개발

서울대학교 대학원 치의과학과 치과교정학 전공

(지도교수: 이 신 재)

서 희 연

CONTENTS

Ⅰ. INTRODUCTION

Orthodontic literature is replete with studies regarding the association between surgical skeletal repositioning and the consequential soft tissue response. Predicting postoperative facial profile is an essential step in the treatment planning of combined surgical-orthodontic treatment.[1-4] Defining a patient's projected post-surgery soft tissue profile used to be the starting point of planning the surgical-orthodontic treatment. This step also increases a patient's understanding and acceptance of the recommended treatment.[5,6]

Computer programs attempting to predict the soft tissue changes after surgery have been greatly improved by graphics and user interfaces.[7] All orthognathic surgical simulation software programs are based on either preprogrammed soft- to- hard-tissue movement ratios derived from studies that reported the mean ratios of soft-tissue to hard-tissue movements,[8] or ordinary least squares (OLS) equations that represent the database for manual and computerized surgical predictions.[1,2] However, during surgical-orthodontic treatment, the soft tissue profile does not directly follow the surgical changes in the underlying bony structures.

Statistics is an utmost essential tool in finding clinically significant evidences that enable clinicians and researchers to conduct proper decision-making. Post-surgery soft tissue response is associated with a number of factors, including skeletal configuration (e.g., the amount of surgical skeletal movement horizontally and vertically), soft tissue drapes, and patients' characteristics. In addition, some or many of these factors can correlate to each other. For example, the horizontal skeletal anatomy is closely related to vertical repositioning, and so is the vertical skeletal configuration to horizontal repositioning. Therefore, the characterization of soft tissue responses is highly intricate and results in statistical challenges. First, the

characterization often relies on a large number of variables with relatively small sample size, which causes high-dimensionality low sample size (HDLSS). Second, there would be a high correlation structure between and/or within variables. Third, variables may have mixed effects (i.e., fixed and random effects). Conventional statistical analyses cannot properly deal with these challenges, and hence prediction of post-surgery outcome requires a sophisticated methodology that involves multiple predictor and multiple response variables simultaneously.[9,10]

Likewise, modern biomedical datasets often involve an ill-posed problem due to HDLSS and/or multi-collinearity. The partial least squares (PLS) method has been used as an alternative approach to solve this ill-posed problem instead of an ill-conditioned ordinary least squares (OLS) regression model.[11,12]

The PLS method is a comparatively new way of constructing prediction equations. Its application to various scientific and biologic disciplines from chemical engineering to brain image analysis is becoming increasingly widespread.[13,14]

Applying the PLS method is advantageous when the number of variables are many and the variables are highly correlated. The merit of the PLS method is its capability of taking correlation structures into account, not only between the predictor- and response variables, but also controlling for the correlation within the predictor variables and/or the response variables. The first study applied multivariate PLS method in dental research was published in 2012.[9] In the study, the authors applied the PLS method to predict soft tissue profile after mandibular setback surgeries. PLS method demonstrated considerably more accurate predictions than the conventional ordinary least squares (OLS) method.[9] The conventional OLS method was found not accurate enough when there were a multitude of correlated variables.[9] Among the variables considered when predicting the soft tissue response to surgery are the patient's age[1,3,15-18];

gender[3,17-20]; time after surgery[21,22]; and pre-surgical soft tissue characteristics including tissue thickness measured at various landmarks.[4,18,23-27] These various factors can be considered in the PLS method through orthogonal linear combinations that can extract a small number of significant components that are combinations of the original variables.[28] In addition, the improved accuracy of the PLS method is likely due to the fact that the soft tissue response at a specific point is highly dependent on its adjacent soft tissue response, *i.e.*, the interdependency of soft tissue points upon each other.[10] The aforementioned investigation, however, had only been performed for mandibular setback surgeries alone.

Studies after Suh (2012) also reported that PLS method is more accurate in predicting soft tissue change after various surgeries, including one-jaw surgeries, two-jaw surgeries and genioplasties, correcting Class II or Class III skeletal discrepancies.[10,29]

Furthermore, a prospective study with a new set of data was performed. In the study, the authors tested the validity of the soft tissue prediction method developed for Class II surgery patients. The multivariate PLS regression again returned more accurate prediction results than the conventional method did in Class II surgery patients.

Previously, the subjects were homogenous in terms of surgical interventions since a homogenous sample was required when investigating the soft tissue response to a specific orthognathic surgery.[9] Including an additional surgery has a great influence on the soft tissue profile change. Most papers have been reporting results of one specific maxillofacial surgical procedure in the sense that the more surgical procedures one adds, the more complex soft tissue prediction becomes.[1,30] There has been an increase in the use of bimaxillary surgery because it is increasingly recognized to produce more stable results than single-jaw mandibular procedures in Class III correction.[31,32] Prediction programs for bimaxillary surgery were less predictable than

for 1-jaw surgery.[5,7,30,33] It would be difficult to exactly determine the changes in the soft tissue profile that are specific to the mandibular setback surgery for other orthognathic surgical procedures, such as Le Fort I osteotomy and/or genioplasty, have been included. Prior studies derived different prediction equations for Class II or Class III skeletal discrepancies. This was typical since the directions of surgical movement differ for each type of skeletal discrepancy. Given this, having one universally applicable prediction model will be appealing. With such model, all patients' change of soft tissue can be accurately predicted regardless of the kind of surgery they underwent.

The PLS method is known to yield an accurate soft tissue prediction model after surgery. Unfortunately, prediction equation based on the PLS method is highly complicated that it is almost impossible to interpret. To identify factors that might influence the soft tissue response more, previous research depicted a loading plot for the anteroposterior lower lip response. The loading pattern showed that several predictor variables had higher values of influence than others.[29] However, when the number of variables is increased, it may be difficult to interpret the loading plot and identify which landmarks have significant impact on soft tissue response.

In accurate prediction model by the PLS method, we used as much variables as possible to be accurate. It was also because we didn't have scientific evidence of which variables are more important. The purpose of previous studies was improved accuracy. However, to be more useful clinically, it would be practical to have minimum number of variables.

Sparse PLS can solve these problems by promoting sparsity during the process of variable selection. The number of variables incorporated in predicting the treatment result can be dramatically reduced. Partial least squares (PLS) regression is an alternative to classical regression for handling multi-collinearity. The core of PLS regression is a dimension reduction

technique that operates under the assumption of a basic latent decomposition of a response matrix and a predictor matrix.[34] In Chun and Keles (2009), investigators developed a sparse PLS regression that aims to promote sparsity by imposing a penalty onto the direction vector of PLS.[34]

The aims of the present study are 1) to develop a model for an accurate soft tissue prediction that can be applied to various modes of surgical correction: Class II correction, Class III correction, the mandibular surgery and/or the maxillary surgery, and additional genioplasty; 2) to minimize the number of variables incorporated via SPLS method in order to build a parsimonious and interpretable prediction model.

II. REVIEW OF LITERATURE

1. Notations, acronyms, and terms

Notations in this paper follow the definitions have been suggested by Geladi and Kowalski (1986). All vectors will be column vectors, the corresponding row vectors will be designated as transposed vectors. Data matrices are denoted by upper case bold letters (e.g., **X**). The identity matrix is denoted by **I**. Column vectors are denoted by lower case bold letters (e.g., **x**). Matrix or vector transposition is denoted by an uppercase superscript T (e.g., $\mathbf{X}^T$). Two bold letters placed next to each other imply matrix or vector multiplication unless otherwise mentioned. The number of rows, columns, or sub-matrices is denoted by a lowercase italic letter and a given row, column, or sub-matrix is denoted by a lowercase italic letter (e.g., *i*). Predictor variables are stored in an *i* by *j* matrix denoted **X** whose variable is denoted $\mathbf{x}_{i,j}$ and where the rows are observations and the columns are variables. For convenience, **APPENDIX TABLE** also lists our main notations, acronyms, and terms.

2. Limitation of the conventional regression model

The classical regression approach estimates the unknown parameters of the equation by ordinary least squares (OLS) method. The prediction equation is written as $\mathbf{Y} = \mathbf{X}\mathbf{B}_{OLS} + \mathbf{E}$, where **E** is an $n \times k$ matrix of residual for **Y** containing the errors which are assumed to be uncorrelated, to be normal and to have the same variance. $\mathbf{B}_{OLS}$ is a $p \times k$ matrix solution of least squares, coefficients $\mathbf{B}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ by multivariate Gauss-Markov theorem.[35] Gauss-Markov theorem shows that the least squares estimate $\widehat{\mathbf{B}}$ is a good choice when the errors are

uncorrelated, normal and have the same variance. The latter equation gives a hint towards the most frequent problem in OLS: the inverse of $\mathbf{X}^T\mathbf{X}$ may not exist.[36]

When the factors are few in number, are not significantly correlated, and have a well-understood relationship to the responses, then conventional analysis, like OLS can be a good way to turn data into information. However, the OLS method assumes that all the predictor variables are independent, which is not the case especially for the numerous dental and facial variables in **X** and **Y** matrices. In practice, this condition is never exactly met. In multiple regressions, linear or nonlinear, collinearities among the predictor variables x_j sometimes cause severe problems.[37] The estimated coefficients $\hat{\beta}_j$, can be very unstable, and thereby far from their target values. In particular, this makes predictions by the regression model to be poor.[38] Each skeletal landmark in an individual subjects is located side by side and moves together during the surgical repositioning, i.e., surgery influences all variables in **X** matrix. Hence these variables all go up, down, forward and/or backward when the operation is being performed. These correlations among the **X** variables make the OLS solution, which requires the inversion of $\mathbf{X}^T\mathbf{X}$, almost infinite. The estimated coefficients $\hat{\beta}_j$, in turn, give the predictive model bad prediction properties. Moreover, the number of predictor variables that can affect soft tissue response, p, is much larger than the number of observations which also makes the effective rank of **X** much smaller than p. Additionally, when the number of predictors exceeds the number of observations, the likely result will be a model that fits the training dataset perfectly but that will fail to predict test data well. This phenomenon is termed *over-fitting*.[39]

3. Feasible approaches to solve multi-collinearity problem and large p small n situation - Multivariate regression methods using latent variables

To solve the problems of multi-collinearity and large *p* small *n* situation, two multivariate regression methods using latent variables were introduced in statistics: principal component regression (PCR) and partial least squares (PLS) regression. PCR and PLS can handle problems with any number of predictor variables. They are applicable even when the number of predictor variables is much greater than the number of observations. Both can yield more accurate prediction equations than conventional method does, not only in multi-collinear and $n \ll p$ situation, but also when the standard assumptions of regression analysis are satisfied. Researchers use latent variable methods in order to 'focus' the information of large data into a few underlying phenomena (also called latent variables, components, or factors) leaving most of the measurement noise behind as residuals. In other words, each object is regarded as a 'mixture' of a few underlying phenomena. The aim is to identify and quantify these phenomena with minimal effects of measurement noise.[40]

4. PLS to handle multi-collinearity problem and large p small n situation

4.1 Characteristics of PLS in comparison with other methods

The PLS method is based on the singular value decomposition of $\mathbf{X}^T\mathbf{Y}$: $\mathbf{X}^T\mathbf{Y} = \sum a_i \mathbf{f}_i \mathbf{g}_i$ where ($\mathbf{f}_i$) and ($\mathbf{g}_i$) are orthonormal vectors of appropriate dimension and (a_i) are the singular values arranged in decreasing order. Hőskuldsson (1988) have shown that upon convergence the weight vectors $\boldsymbol{w}_1$ and $\boldsymbol{q}_1$ correspond to the first pair of left and right singular vectors obtained from a singular vector decomposition (SVD) of the matrix of cross products $\mathbf{X}_0{}^T\mathbf{Y}_0$. Since the

dominant singular value equals $\boldsymbol{w}_1\mathbf{X}_0^{\mathrm{T}}\mathbf{Y}_0\boldsymbol{q}_1 = \boldsymbol{t}_1^{\mathrm{T}}\boldsymbol{u}_1 = (n-1)\cdot\mathrm{cov}(\boldsymbol{t}_1,\boldsymbol{u}_1)$, the score vectors $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$ have maximum covariance among all score vectors obtainable by applying normalized weights to $\mathbf{X}_0$ and $\mathbf{Y}_0$, respectively.[41] For detailed interpretations, please refer to Höskuldsson (1988) and Helland (1990).

On what basis can one expect PLS regression methods to perform better than OLS and other well-known regression techniques? The answer is the stability of predictors derived from PLS methods.[41] The essential criteria in evaluating the predictability of models is the number of variables included in the models. The uncertainty of the estimated parameters quickly becomes the dominating factor in the variability of predictors. Therefore, it is important to keep the number of variables as low as possible. In PLS, components are selected to give 'maximal' reduction in the covariance $\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ of the data.

According to Wold et al. (1984), the PLS method is equivalent to the conjugate gradient method used in numerical analysis for related problems. The original algorithm by Wold is essentially another description for the conjugate gradient algorithm for solving the least squares problem with singular $\mathbf{X}$. PLS utilizes the principle of dimension reduction by obtaining a small number of latent components that are linear combinations of the original variables to avoid multi-collinearity. [28] The procedures of OLS and PCR occupy the opposite ends of a continuous spectrum, with partial least squares lying in between. There are two adjustable 'parameters' controlling the procedure: α, in the continuum [0, 1], and ω, the number of regressors finally accepted. Where α is a real number in the interval [0, 1], with the values 0, ½, and 1 corresponding to OLS, PLS and PCR respectively. The role of α suggests the obvious title for the procedure – 'continuum regression'. These control parameters are chosen by cross-validation.[42]

Compared with other approaches, the PLS method has following advantages: the PLS solution is similar to PCR except that the projection **T** is computed *both* to model **X** *and* to correlate with **Y,** while the PCR, **T** is computed only to model **X**. This is accomplished by introducing a weight matrix **W** and a set of latent variables for **Y** denoted by **U** with the corresponding loading matrix $\mathbf{B}_y$ (**Figure1**). This makes the PLS solution have equal or better predictive properties for **Y**; better in the case when the information in **X** about **Y** appears among the later singular vectors of **X** (corresponding to small singular values).[43] PLS is considered especially useful for constructing prediction equations when there are many explanatory variables and comparatively little sample data.[41]

4.2 Application of PLS method in soft tissue prediction after surgery- Previous studies

Mandibular setback surgeries

Recently, Suh (2012) reported the application of multivariate PLS method to predict soft tissue profile after mandibular setback surgeries.[9] In the study, prediction model was constructed by PLS method with data from patients underwent mandibular setback surgeries that correct Class III skeletal discrepancy. PLS method demonstrated more accurate predictions than the conventional ordinary least squares (OLS) method [9] for Class III subjects who had mandibular setback surgeries.

Class II surgeries

To explore whether PLS method provide an accurate prediction model for surgeries to correct skeletal Class II relationship, data from patients who were in skeletal Class II category before orthognathic surgery were collected. The investigation suggested that PLS method provides an accurate profile prediction after surgery for Class II patients as well.[10]

Class III surgeries

PLS method is known to be accurate in predicting two-jaw surgeries and additional surgeries correcting Class III skeletal discrepancies.[29] Number of components was introduced in the study. Root mean squared error of prediction (RMSEP) curve was used to select the best prediction model. During building a prediction equation in the training set, the more components were included, the smaller predicton error were obtained. The full prediction model with the entire components had zero error in the training set. However, when validating the equation to each individual subject, there was an optimum number of components to minimize prediction error in the test set. Since it is typical to choose the smallest model that minimizes the expected prediction error[39], the PLS prediction equation with 30 number of components was selected as the final prediction model. The validation result demonstrated that the multivariate PLS prediction model with 30 orthogonal components showed the best prediction quality among others.

A prospective study

To test the validity of a new prediction method, its prediction errors should be expected to be in reasonable agreement again for future patients. Therefore, a prospective study with a new set of data was performed. Yoon and his colleagues tested the validity of the soft tissue prediction method developed for Class II surgery patients.[44] The multivariate partial least squares (PLS) regression again returned more accurate prediction results than the conventional method did.[44]

Prediction performance of PLS method - cross-validation

The leave-one-out method is one of the classical cross-validation procedures. Recent research suggested that the leave-one-out method might be the best validation strategy in a clinical research framework. [44] In the leave-one-out cross-validation (LOOCV) method, each subject is

part of the validation data set, and can play the role of a new prospective data set. This makes the best use of a study's sample size and maximizes the potential use of each subject. In addition, it is possible to measure the individual error and variability after the validation test.[44] Therefore the LOOCV method may be the most optimal and safe validation method for improving the prediction error in a given sample.

Based on aforementioned research, LOOCV was selected for cross-validation to test validity and model selection to choose smallest model that minimizes the prediction error in the present study.

5. PLS and Sparse PLS

Partial Least Squares (PLS) Regression

Partial least squares (PLS) regression is a regression technique suggested by Wold (1972) in order to solve the problem when the data are seriously correlated. This approach regresses response variables on latent variables. Latent variables are linear combinations of the predictor variables that maximize the variance between response and predictor. Thus, our aim is to find latent variables satisfying

$$max_{b,c}\, Cov(\boldsymbol{X}b, \boldsymbol{Y}c),$$

where b and c depend on both directions and norm as bellow.

$$b^t b = 1 \text{ and } c^t c = 1.$$

The objective function, $h(t)$, can be defined as

$$h(b, c, \lambda_1, \lambda_2) = b^t X^t Y c - \lambda_1(b^t b - 1) - \lambda_2(c^t c - 1).$$

By Lagrangian multiplier, we can solve the equation and obtain latent variables. The number of

latent variables is usually determined by a *K*-fold cross-validation technique.

Sparse Partial Least Squares (SPLS) Regression

Sparse partial least squares (sparse PLS) regression is a penalized regression method proposed by Chun and Keles (2010) in order to conduct dimension reduction and variable selection, *simultaneously*. Since the estimation procedure of the PLS regression is ultimately affected by the large number of predictors, it causes inconsistent coefficient estimates.[45]

To solve the problem, Chun and Keles (2010) proposed to use a penalty, such as lasso or elastic net, in a pre-processing step before PLS regression fit.[28] As a result, it provides easy interpretation of correlations among predictors under the framework of the PLS regression.

By obtaining sparse linear combination of the original predictors, sparse PLS can achieve good predictive performance and variable selection *simultaneously*. Under this procedure, small loadings of the sparse PLS regression may be set to zero without significantly decreasing the total variance, imposing sparsity when constructing the direction vectors. Using a prediction matrix **X** and a response matrix **Y**, the sparse PLS regression estimates the k^{th} column of the weight matrix **W** to identify the principal components (or, latent variables) that have the maximum covariance with the response variables. Moreover, it selects the optimal set of predictor variables by tuning L_1 and L_2 penalty terms as follows.

$$\min_{w,c} -\kappa w^T X^T Y Y^T X w + (1-\kappa)(c-w)^T X^T Y Y^T X (c-w) + \lambda_1 \left|\left|c\right|\right|_1 + \lambda_2 \left|\left|c\right|\right|_2,$$

$$\text{s.t.} \, w^T \mathrm{w} = 1 \,,$$

where *c* is estimated a priori as

$$\hat{c} = \left(|Z| - \eta \max_j |Z_j|\right)_+ \text{sign}(Z), \quad \text{with } Z = \frac{X^T Y}{||X^T Y||}, \quad (x)_+ = \max(0, x),$$

and κ and η are pre-specified tuning parameters.

With the regularization, the latent variables depend not on a whole set, but on a subset of the predictors. In this framework, the first term of the equation can be fitted by an ordinary PLS algorithms, such as SIMPLS, kernel PLS, and so on. Moreover, Leave-one-out cross-validation technique can be used to select the optimal value of the penalty term.

6. Cross-validation for model assessment and model selection

6.1 Purposes

The purposes of cross-validation (CV) are to estimate the risk of an estimator or to perform model selection.[46] In addition, CV is probably the simplest and most widely used method for estimating prediction error.[39]

One of the major purposes of statistical analysis is to use past events to predict future events. In orthodontics, predicted outcomes of treatment are fundamentally needed in order for orthodontists to properly diagnose and treatment plan their patients. Given the huge inter-individual variation among human beings, the accuracy of these predictions is a wide open question. A prediction model may produce promising results when developed from its original data, but often cannot reproduce the same promising results when applied to new subjects.[39,47] Therefore, a validation report is required to better scrutinize a prediction prior to its clinical application.

The ability to make accurate predictions is the major goal of a prediction study. In general, two steps are necessary: a model building procedure and a validation step. After the prediction

model has been developed from the original data, assessing a model's prediction error using an independent dataset is defined as *validation,* and that dataset is called the *validation set*.

There are several related terms that need to be defined. *Learning data* (or set) refers to the data from which a prediction model is derived or built. The learning dataset is also called the *training* or *study* dataset. *Validation data* is a set or sets of data used to estimate the future accuracy of the prediction formula derived from the learning data. The validation data set can also be called the *test*, *real*, or *prediction* dataset.

The model building procedure is performed using the learning data set. This is the process by which various prediction equations or models are developed. After the model fitting, the *learning error*, which assesses the *goodness-of-fit* of the prediction model, is calculated. This is the method by which the best prediction model is selected based on its goodness-of-fit and resulting calculated learning error. These procedures are all calculated using the original dataset. After selecting the final prediction model, the validation step assesses how well the chosen prediction method performs on new subjects. Therefore, the true test of the prediction formula is how well the method predicts the outcomes for a different data set. To restate, after building a prediction model with the learning data, the prediction performance, or the *validity* of the prediction model, is assessed according to how well the prediction model fits with the *validation set* and the calculated *validation error* (**Figure 4**).[9,39,44,47]

Using the validation data set to assess prediction accuracy is a crucial step in the development of a new prediction equation. The validation step measures not only the accuracy but also the generalizability of the new prediction model.[48] Therefore, it is always recommended to test the validity of a prediction model with a validation data set, rather than just reporting the goodness-of-fit measure or the learning error in the learning data. Unfortunately, validation reports are

often missing from orthodontic studies. Omitting a validation report is one of the most common errors in prediction studies.

6.2 Definition

CV is a method for model selection according to the predictive ability of the models.[49] Suppose that n data points are available for model selection. The data set can be split into two parts. The first part contains n_c data points used for fitting a model (model construction), whereas the second part contains $n_v = n - n_c$ data points reserved for assessing the predictive ability of the model (model validation). Strictly speaking, model validation is carried out using not just n_v, but all the $n = n_v + n_c$ data. There are $\binom{n}{n_v}$ different ways to split the data set. Cross validation, as its name indicated, selects the model with the best average predictive ability calculated based on all (or some) different ways of data splitting.

6.3 Various cross-validation methods

The simple validation method

The simple validation method has been a popular method since the early 1930's.[46] This technique is also known as the *simple split*, *learning-test* split, or *holdout* method.[48]

Cross-validation methods

Various splitting strategies lead to several related cross-validation techniques. In the K-fold cross-validation method, the original data set is randomly partitioned into K subsamples. Of the K subsamples, a single subsample is then retained for use as the validation data set, and the remaining $K - 1$ subsamples are used as the learning data set. The cross-validation process is

then repeated *K* times (the *folds*), with each of the *K* subsamples used exactly once as the validation data. The validation results from each fold can then be averaged (or otherwise combined) to produce a single estimation. The advantages of this method over completely separate subsets or over repeated random sub-sampling are that all of the observations are used for both learning and validation, and each observation is used for validation exactly once.[49] The *K* can be equal to any value. The optimal *K* is usually limited to between 5 to 10.[39,46] The case where *K* = *N*, that is where *K* is equal to the number of subjects, is known as the *leave-one-out* (LOOCV) method; it is the most extreme version of the *K*-fold cross-validation method (**Figure 4**).

6.4 Measure of prediction performance and model selection in the PLS method

The goodness-of-fit for an OLS model is expressed as the coefficient of determination, also called R^2. However, in case of PLS prediction no proper overall goodness-of-fit measure exists.[50] For this reason, the cross-validated R^2, also called Q^2 was proposed:

$$Q^2 = 1 - \sum \frac{(\mathrm{Y_{predicted}} - \mathrm{Y_{observed}})^2}{(\mathrm{Y_{predicted}} - \mathrm{Y_{mean}})^2}$$

A model with good internal predictive performance will have Q^2 value close to 1. A Q^2 with a negative value indicates chance correlations.[51]

Predicted residual sum of squares (PRESS) is the total sum of squares of predictions minus observations. PRESS has been most frequently used as a measure of the predictive power of the model among other statistics.[52] The square root of PRESS is the root mean squared error of prediction (RMSEP). This is used to measure the prediction performance and model selection in

the “pls” package written in statistic language R.[53,54] Using the “pls” package, a researcher can fit the model, draw cross-validated RMSEP curves and find the optimal number of components of the model. RMSEP have simple advantage of keeping the original unit intact and behaves like a standard deviation.[55]

Ⅲ. MATERIALS AND METHODS

1. Subjects

The subjects consist of 318 patients (191 women, average age 24.1 years; 127 men, average age 23.6 years) who had undergone the surgical correction of severe Class II or Class III malocclusion. All patients were treated at the Department of Orthodontics, and surgery was performed at the Department of Oral and Maxillofacial surgery, Seoul National University Dental Hospital. All subjects were of Korean ethnicity. Subjects had a mandibular surgery and/or maxillary surgery. Patients who had a cleft lip and palate, severe asymmetry, facial deformity due to syndrome were excluded.

All patients were treated with fixed orthodontic appliances before and after surgery. During the preoperative orthodontic treatment, dental decompensation and proper arch coordination was achieved. Postoperative orthodontic treatment was limited to completing the adjustment of the occlusion, and minimal incisor movement was required. The institutional review board for the protection of human subjects reviewed and approved the research protocol (institutional review board no. S-D 20140019).

2. Cephalometric analysis

Every patient took lateral cephalograms before and after orthognathic surgery. Patients were instructed to hold their teeth in occlusion with their lips relaxed when they take X-ray images. Preoperative lateral cephalograms were taken close to the time of surgical intervention.

Postoperative cephalograms were taken at least 3.7 months after the surgery to allow the postoperative swelling to be resolved.

To orientate a subject's pre- and post-operative tracings to the same head position, the two tracings were superimposed on the anterior cranial base to confirm whether the Sella-Nasion planes were coincident. The anatomical tracing, cephalometric landmarks, soft tissue outline, and their abbreviations used in the study are illustrated in **Figure 3.** Forty-six skeletal landmarks and 32 soft tissue landmarks from glabella to the terminal point were identified. Upper-case letters were used to demarcate hard tissue landmarks. Lower-case letters were used to indicate the soft tissue landmarks.

With its origin at Sella, the vertical reference was established perpendicular to Sella-Nasion +7 degrees. Sella-Nasion is considered to be relatively stable beyond 7 years of age.[31] The *x* coordinates represented the horizontal distance from the vertical axis, and the *y* coordinates represent the vertical distance from the horizontal axis measured in millimeters. Using a custom digitizing program via Microsoft Visual C# 2010 (Microsoft, Redmond, Washington), the coordinates of every landmark on each tracing were sequentially computed in relation to the *x* and *y* reference system.

3. Variables in predictor and response matrices

A total of 232 input variables, also called predictor variables, were entered into the prediction equation. The *predictor variables* included: the patient's age, sex, time after surgery, existence of bimaxillary surgery, existence of genioplasty, 78 pre-surgical skeletal measurements, 64 pre-surgical soft tissue measurements; and 78 variables with regard to the surgical skeletal repositioning in both anteroposterior and vertical directions. The output variables, also called

response variables, were the soft tissue responses at the 32 soft tissue landmarks both in *x*- and *y*-axes summing up 64 output variables. The prediction algorithm used in this study was based on the PLS and SPLS methods. **X** matrix was a 318 $(N)_{\text{subjects}}$ × 232 $(K)_{\text{variables}}$ matrix of predictor variables and **Y** matrix was a 318 $(N)_{\text{subjects}}$ × 64 $(M)_{\text{responses}}$ matrix of response variables (**Figure 2**). The more detailed algorithm is available in previous publications.[9,10,29]

4. Two multivariate methods to make prediction equations

Two multivariate methods, the PLS method and the SPLS method were used to construct prediction equations.

The PLS prediction equation may be written as $\mathbf{Y} = \mathbf{X}\mathbf{B}_{\text{PLS}} + \mathbf{E}$, where **E** is an $n \times k$ matrix of residual for **Y**, and $\mathbf{B}_{\text{PLS}}$ is a $p \times k$ matrix of PLS prediction coefficients (**Figures 1** and **2**). In the equation itself, the PLS method resembles the stepwise OLS method. But in contrast to the OLS method, the PLS method is applicable even if the variables are strongly intercorrelated (multi-collinear), contain significant noise, and even if the number of variables is greater than the number of subjects. (i.e., "small *n* large *p* situations").[40]

Sparse partial least squares (SPLS) regression is a penalized regression method proposed by Chun and Keles (2010). Chun and Keles (2010) proposed to impose a penalty in a pre-processing step before PLS regression fit to obtain sparse linear combination of the original predictors.

Under this procedure, small loadings of the sparse PLS may be set to zero without significantly decreasing the total variance, imposing sparsity when constructing the direction vectors. Using a prediction matrix **X** and a response matrix **Y**, the Sparse PLS estimates the k^{th} column of the

weight matrix **W** to identify the principal components (or, latent variables) that have the maximum covariance with the response variables. Moreover, it selects the optimal set of predictor variables by tuning L_1 and L_2 penalty terms. Leave-one-out technique was utilized to select the optimal value of the penalty terms.

5. Comparing the prediction performance between PLS method and SPLS method

When developing a prediction method, the model is fitted from part of the data (the training dataset), and the quality of the fit is judged by how well it predicts the other part of the data (the test set, also called the validation dataset). To evaluate the predictive performance of the prediction equations, the leave-one-out cross-validation method was used. Mean training errors and mean test errors were calculated for the prediction model based on PLS method and models based on SPLS method. Mean training errors indicate the accuracy of the model on the training dataset which was used in model construction. Mean test errors mean the accuracy of the model on the new data, test dataset, which was not used in model fitting. The mean test errors simulate errors in a real situation more closely than mean training errors.

After fitting the equation, the mean absolute error, $|\mathbf{Y}_{\mathbf{actual}} - \mathbf{Y}_{\mathbf{predicted}}|$ is used as the *criterion of goodness-of-fit* (**Table 2**).

6. Model selection for SPLS method

The aim of the model selection is to choose a model incorporates minimum number of variables yet still accurate, i.e., to build a parsimonious prediction model. To minimize variables used in building prediction model for simple and more interpretable prediction algorithm, I increased

sparsity by raising eta value. Eta value was raised until there is a significant difference between PLS regression and SPLS regression.

Leave-one-out cross-validation method was used to evaluate the prediction performance and to determine the optimal SPLS model, i.e., a model selection.[56]

Scattergrams and 95% confidence ellipses were constructed to compare the performance of selected model (**Figures 5** and **6**). Among the 32 soft-tissue landmarks, only six landmarks were selected: pronasale, stomion superius, lower lip, pogonion and menton. They are most prominent and clear landmarks that can show the change of soft-tissue of the selected areas.

Language R (Vienna, Austria)[54], which is a free software environment for statistical computing, was used. Detailed codes of the PLS, SPLS and validation algorithm for use with language R is available upon request to the author.

Ⅳ. RESULTS

1. Detailed description of study subjects

Table 1-1 provides further details of study subjects. Sixty-two percent of the subjects were females. The average time after surgery was 9.5 months. **Table 1-2** shows the vast variety of surgical movements that subjects of present study underwent. Two thirds of the subjects, 204 patients underwent surgeries to correct Class III skeletal discrepancy. The remaining 114 patients had Class II skeletal discrepancy and the surgical repositioning was to correct mandibular insufficiency.

Among Class III patients, 71 patients had the mandibular surgery only, 133 patients had bimaxillary surgery, and 81 patients had additional genioplasty. When the patient had maxillary surgery, point A was slightly advanced (1.4 mm) and moved upward (1.1 mm). Point B moved backward and upward. The average amount of surgical repositioning at point B was 7 mm posteriorly and 3 mm superiorly.

In Class II patients, 20 patients received a mandibular advancement surgery. 94 patients had bimaxillary surgery including Le Fort I surgery and/or anterior segmental osteotomy in the maxilla. The average amount of surgical repositioning at point B was 6 mm anteriorly.

2. Building a universally applicable prediction model

Prediction models are built based on the data set of surgery patients. The surgeries the patients had include almost all surgeries that are commonly performed to correct skeletal discrepancies: mandibular setback surgery, mandibular advancement surgery, bimaxillary surgery including Le

Fort I surgery and/or anterior segmental osteotomy in the maxilla and additional genioplasty

The result of the prediction errors after applying the prediction equations in the test dataset from the two methods are summarized in **Table 2**. The soft tissue landmarks that we presented in the table were chosen to concisely describe the validity and accuracy of the suggested soft tissue prediction method.

The derived prediction method was successfully cross validated. A comparison test based on the means between the predicted and the actual soft tissue profile may not be appropriate since underestimates and overestimates will cancel each other out, showing no significant difference between the means.[9,57,58] Instead, as previously suggested,[9,58] scattergrams and 95% confidence ellipses for several soft tissue landmarks were constructed to assess the prediction error of PLS and SPLS methods (**Figure 5**). The ellipsoid satisfies $(\mathbf{z} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{z} - \mu) \leq \chi^2(\alpha)_2$, where $\mathbf{z}$ is the two dimensional (*x*- and *y* coordinates) vector for the error, μ is the mean vector for $\mathbf{z}$, $\mathbf{\Sigma}^{-1}$ is the inverse matrix of the covariance matrix, and $\chi^2(\alpha)_2$ is the upper 95th percentile of a chi-square distribution with two degrees-of-freedom.[35] The contour of an ellipse indicates the 95% confidence boundary. A negative value indicated the prediction was more posterior in the *x*-axis or more superior in the *y*-axis compared to the actual result. If any points are outside the ellipse, they can be called *outliers*.[57] The size of the 95% confidence ellipses for the PLS method and the SPLS method showed acceptable predictive performance (**Figure 5**). PLS regression was proven to be an appropriate method to predict soft tissue profile after orthognathic surgery.[9,10,29] Prediction models by SPLS regression showed comparable prediction performance (**Table 4**).

Soft tissue prediction accuracy was compared between Class II and Class III surgery patients, which is given in **Figure 7**. Among the 32 soft-tissue landmarks, only six landmarks were selected to concisely report the results: pronasale, stomion superius, lower lip, pogonion and

menton. They are most prominent and clear landmarks that can show the change of soft-tissue of the selected areas. There was no significant difference in the test error between Class II surgery and Class III surgery.

Figures 8 through **10** present comparison of prediction accuracy between different groups of patients: female and male; patients had additional genioplasty and no additional genioplasty; patients underwent maxillary surgery and no maxillary surgery. There were no significant differences in the sizes of ellipses depending on surgical movement or gender of the subjects. Thus, the more accurate PLS and SPLS models built in present study are universally applicable.

3. Development of an optimal SPLS model for an accurate soft tissue prediction

To minimize variables used in building prediction model for simple and more interpretable prediction algorithm, I promoted sparsity. Eta value was raised until there was a significant difference between PLS regression and SPLS regression.

SPLS showed no significant difference in prediction performance until eta value was raised to 0.9 which includes independent 104 variables (**Tables 2**, **3** and **4**). When the eta value was raised to 0.99, there were some landmarks that are predicted by PLS regression better, however, the excellence of prediction performance of PLS regression was not significant clinically even in eta=0.99 situation (**Figure 6**). When eta was raised to 0.99, the number of independent variables needed in building prediction model became 34. **Figure 11** illustrates these landmarks on the patient's photographs.

Ⅴ. DISCUSSION

1. Building a universally applicable prediction model

In spite of numerous attempts to predict soft tissue change after orthognathic surgery, there is no satisfying method that is accurate enough to use in clinical situation. Although unexplained individual variations are inevitably present, statistical and mathematical approaches used previously were not suitable to make a soft tissue prediction model.

Orthodontics has problems with which conventional statistical inferences cannot properly deal. Statistics is an utmost essential tool in finding clinically significant evidences for proper decision making procedures for both clinicians and researchers. Most familiar statistical tests in orthodontics, such as the *t* test, analysis of variance, and regression analyses are all based on the conventional ordinary least squares (OLS) method. The OLS models basically require the prerequisite condition of independence between the predictor variables as well as normality and equality of variance among them.

Since anteroposterior skeletal anatomy is significantly related with vertical repositioning and vice versa, these highly correlated variables should be considered before predicting soft tissue results. In this respect, multivariate methods, like the PLS method, enjoy large popularity in a wide range of fields in natural sciences. Partial least squares (PLS) is the preferred method for constructing a prediction equation when the factors are many and highly collinear or correlated. Applying the PLS method is even possible when the sample size is less than the number of variables.[52,53]

Furthermore, previous studies were not able to analyze and interpret more than one type of surgery or more than one vector of movement once at a time of investigation. Since

conventional methods could not properly handle the complex data structures, only identical surgical procedures could be analyzed. Isolated mandibular prognathism occurs in a relatively small portion of Class III patients.[3,32] Therefore, the combination of a Le Fort I osteotomy of the maxilla and a mandibular setback surgery seems to be the current trend for skeletal Class III treatment.[31,59] With previous prediction methods, even an additional genioplasty was considered as a confounding variable when predicting soft tissue responses.[1] For those patients who undergo maxillary surgery and/or genioplasty, the vectors of movement are not uniform. Thus patients undergoing an additional jaw surgery would have less predictable results than those undergoing a relatively simple mandibular setback surgery.[5,7,30] Compared to Class III surgery studies, the soft tissue response to Class II surgeries seems to behave differently from that of the Class III patients. Several other papers also reported this difference. Due to stretching of the soft tissue, mandibular advancement surgery has a positive "lifting-effect" on the soft tissue profile. In mandibular advancements, the lower lip followed the hard tissue less so than in mandibular setbacks, 50% in advancements and 100% in setbacks.[1-3,60]

One of the aims of the present study was to develop an accurate soft tissue prediction model that can be applied to various modes of surgical correction: Class II correction, Class III correction, the mandibular surgery and/or the maxillary surgery, and additional genioplasty. There were no statistically significant differences in the prediction errors depending on directions of surgical movements, types of surgery and gender. (**Figures7**, **8**, **9** and **10**) The confidence ellipses obtained from the comparison between the groups showed apparently that the types of surgery did not influence the accuracy of prediction model constructed in present study.

Previous studies have reported that the predictability of the upper lip position after mandibular setback or advancement is poor and highly variable in both directions.[1,2] On the other hand,

other papers have reported that the main area of inaccuracy was the lower lip.[7,23,61] In this study, however, both the upper lip and lower lip areas demonstrated accurate prediction results. Errors may come from the changes in the defined points on the cephalometric tracings and superimpositions. The defined preoperative landmark would not always be at the same location on the soft tissue line after surgery. For example, the preoperatively defined soft tissue pogonion would not always be transferred to be the same 'point' after surgery. When locating the most anterior point of the chin, pogonion, it is likely easier to determine the position on the *x*-axis coordinate than on the *y*-axis, thus producing the large vertical variation in the distributions for pogonion. By the same token, since soft tissue menton is the most inferior point of the chin, there is greater horizontal variation in the *x*-axis than in the *y*-axis. This has been a common problem in reporting cephalometric reliability.[58] Therefore, the landmark definition for several soft tissue landmarks might have contributed to the resultant prediction error.[9]

From the clinical point of view, the validity of a prediction equation is the single most important factor influencing the usefulness of a prediction equation. It is necessary to identify the extent to which a prediction estimates soft tissue change in groups of subjects other than that from which it was derived. Accuracy of test error in the test set is far more important than the training error or the goodness-of-fit of the prediction method in the training set from which the equation was developed.

Introduced in early 1930's, the simple validation method was the first type of validation procedure used. It was also referred to as the hold-out validation method. In order to check their true significance, prediction models should ideally be tested on independent data. Training an algorithm and evaluating its predictive performance on the same data, yields overly optimistic results. Unfortunately, in most real applications, only limited datasets are available, thereby the

simple validation method should not be used. Starting from the idea that testing a prediction algorithm on new data, and not the same data from which it was developed, yields a good estimate of its performance, the idea of splitting the data and cross-validation methods were developed.[42,46]

The leave-one-out method is one of the most classical cross-validation procedures. In this method, *K* equals to the number of total subjects. During the leave-one-out cross-validation, each subject serves as a validation dataset. Each individual can play a role as a "new dataset" without arbitrarily splitting the whole dataset. After validation, therefore, the results of the leave-one-out validation method can preserve each subject's information with regard to the prediction error or the individual pattern. This may be one of the most advantageous features of the leave-one-out method. In this respect, the leave-one-out method might be the best validation strategy in a clinical research framework.

2. Development of an optimal SPLS model for an accurate soft tissue prediction

After finding the universally applicable model based on the PLS method and SPLS method, the next step was choosing the best prediction model. The best prediction model can be defined as the simplest model that minimizes the test error. For the model selection criteria, mean absolute error of prediction (MAEP) was used. The MAEP has been frequently used to assess the prediction performance and to choose the optimal model.

To build a parsimonious model, I increased sparsity. Eta value was raised until there was significant difference between PLS regression and SPLS regression. SPLS showed no significant difference in prediction performance until eta value was raised to 0.9, which includes 104 independent variables. (**Tables 2**, **3** and **4**) When the eta value was raised to 0.99, there

were some landmarks that were predicted by PLS regression better, however, the excellence of prediction performance of PLS regression was not significant clinically even in eta=0.99 situation (**Figure 6**). When eta was raised to 0.99, the number of independent variables needed in building prediction model became 34. Reminding the number of independent variables to build a PLS regression model for prediction algorithm was 232, it is expected that the number of variables reduced to 34 will profoundly simplify the prediction model. This prediction model is more interpretable and pragmatic to implement as an algorithm for imaging programs.

I explored the complex relationship between predictor variables and soft tissue responses by SPLS method. To identify landmarks that influence the soft tissue response significantly, I illustrated selected landmarks on the patient's photographs (**Figure 11**). These landmark coordinates reveal the points on face that significantly influence soft tissue prediction.

One of the aims of present study was to reduce the complexity of prediction model, and SPLS was found to be a powerful and objective tool for simplification. The major advantages of SPLS are as follows. First, SPLS facilitates the interpretation of data. Traditional PLS yields an extremely large number of non-zero loadings, which makes it difficult to interpret. Second, SPLS is capable of identifying important variables. The "small n, large p" problem is increasingly prevalent in research today, partly due to the difficulties involved in obtaining a large number of subjects. In addition, the number of variables has grown due to the increasing sophistication of electronic data acquisition devices. Therefore, it is necessary to develop an efficient and effective method of reducing the dimensionality of information while ensuring that the information remains as intact as possible. We suggest that SPLS provides a relatively easy means of reducing complexity of prediction model.

Computer assisted predictions have become an integral part of surgical-orthodontic treatment planning.[62] However, existing software predictions still result in considerable errors. One of the reasons for the errors might be caused by over-simplistic OLS algorithms that were integrated into most commercially available computer programs. These programs have never been clearly published or opened to the public. We hope that the soft tissue prediction method presented in this study will provide a practical algorithm to improve surgical treatment simulation programs. Detailed algorithms written in language R will be opened to public through General Public Licensure or by request to the author.

The methodology in this study may facilitate the further development of soft tissue prediction algorithms for various surgical and orthodontic treatment objectives. Although this study lacks frontal morphology changes, when additional information is provided, the method could be implemented to incorporate all the other meaningful variables. The use of three-dimensional (3D) images is becoming increasingly popular. [63-65] At the same time, use of computer algorithms became inevitable modality for human face recognition techniques. Two-dimensional face landmarks have not exceeded 100 at best. [66-68] But in case of 3D face images, the number of face landmarks is not comparable to that of the 2D landmarks and well reaches to hundreds of variables. In addition to the inclusion of 3D curvatures identification, this is because a single landmark includes coordinate information of all three planes of space, thus tripling the number of variables. I hope that by using SPLS method, additional datasets can be incorporated with the goal of developing a comprehensive clinical predictive model for anticipating the effect of changes upon an individual's face.

VI. CONCLUSIONS

1. It was our observation that the PLS and SPLS methods resulted in accurate prediction of soft tissue change associated with various types of surgical intervantion.

2. It was possible to build a parsimonious and interpretable prediction model by SPLS method. SPLS method was found to be a powerful and objective tool for simplification.

3. Based on our findings, we propose that the SPLS method might provide an improved algorithm in predicting surgical outcomes after orthognathic surgery.

REFERENCES

1. Joss CU, Joss-Vassalli IM, Berge SJ, Kuijpers-Jagtman AM. Soft tissue profile changes after bilateral sagittal split osteotomy for mandibular setback: a systematic review. J Oral Maxillofac Surg 2010;68:2792-2801.

2. Joss CU, Joss-Vassalli IM, Kiliaridis S, Kuijpers-Jagtman AM. Soft tissue profile changes after bilateral sagittal split osteotomy for mandibular advancement: a systematic review. J Oral Maxillofac Surg 2010;68:1260-1269.

3. Joss CU, Vassalli IM, Thuer UW. Stability of soft tissue profile after mandibular setback in sagittal split osteotomies: a longitudinal and long-term follow-up study. J Oral Maxillofac Surg 2008;66:1610-1616.

4. McCollum AGH, Dancaster JT, Evans WG, Becker PJ. Sagittal soft-tissue changes related to the surgical correction of maxillary-deficient Class III malocclusions. Semin Orthod 2009;15:172-184.

5. Eckhardt CE, Cunningham SJ. How predictable is orthognathic surgery? Eur J Orthod 2004;26:303-309.

6. Smith JD, Thomas PM, Proffit WR. A comparison of current prediction imaging programs. Am J Orthod Dentofacial Orthop 2004;125:527-536.

7. Kaipatur NR, Flores-Mir C. Accuracy of computer programs in predicting orthognathic surgery soft tissue response. J Oral Maxillofac Surg 2009;67:751-759.

8. Chew MT, Sandham A, Wong HB. Evaluation of the linearity of soft- to hard-tissue movement after orthognathic surgery. Am J Orthod Dentofacial Orthop 2008;134:665-670.

9. Suh HY, Lee SJ, Lee YS, Donatelli RE, Wheeler TT, Kim SH et al. A more accurate method of predicting soft tissue changes after mandibular setback surgery. J Oral Maxillofac Surg

2012;70:e553-562.

10. Lee HJ, Suh HY, Lee YS, Lee SJ, Donatelli RE, Dolce C et al. A better statistical method of predicting postsurgery soft tissue response in Class II patients. Angle Orthod 2014;84:322-328.

11. Boulesteix AL, Strimmer K. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief Bioinform 2007;8:32-44.

12. Lee D, Lee Y, Pawitan Y, Lee W. Sparse partial least-squares regression for high-throughput survival data analysis. Stat Med 2013;32:5340-5352.

13. Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage 2011;56:455-475.

14. Lee SJ. Modified partial least squares method incorporating mixed effect model. Department of Statistics, Graduate School. Seoul, Korea: Korea University; 2012: p. 1-50.

15. Alves PV, Mazucheli J, Vogel CJ, Bolognese AM. How the lower face soft tissue changes after mandibular advancement or setback. J Craniofac Surg 2008;19:593-598.

16. Burden D, Johnston C, Kennedy D, Harradine N, Stevenson M. A cephalometric study of Class II malocclusions treated with mandibular surgery. Am J Orthod Dentofacial Orthop 2007;131:7 e1-8.

17. Naoumova J, Soderfeldt B, Lindman R. Soft tissue profile changes after vertical ramus osteotomy. Eur J Orthod 2008;30:359-365.

18. Mobarak KA, Krogstad O, Espeland L, Lyberg T. Factors influencing the predictability of soft tissue profile changes following mandibular setback surgery. Angle Orthod 2001;71:216-227.

19. Chou JI, Fong HJ, Kuang SH, Gi LY, Hwang FY, Lai YC et al. A retrospective analysis of the stability and relapse of soft and hard tissue change after bilateral sagittal split osteotomy for

mandibular setback of 64 Taiwanese patients. J Oral Maxillofac Surg 2005;63:355-361.

20. Kolokitha OE. Validity of a manual soft tissue profile prediction method following mandibular setback osteotomy. Eur J Dent 2007;1:202-211.

21. Dolce C, Hatch JP, Van Sickels JE, Rugh JD. Five-year outcome and predictability of soft tissue profiles when wire or rigid fixation is used in mandibular advancement surgery. Am J Orthod Dentofacial Orthop 2003;124:249-256.

22. Kau CH, Cronin A, Durning P, Zhurov AI, Sandham A, Richmond S. A new method for the 3D measurement of postoperative swelling following orthognathic surgery. Orthod Craniofac Res 2006;9:31-37.

23. Ksiezycki-Ostoya BK, McCollum AGH, Becker PJ. Sagittal soft-tissue changes of the lower lip and chin associated with surgical maxillary impaction and consequent mandibular autorotation. Semin Orthod 2009;15:185-195.

24. McCollum AGH, Gardener GJM, Evans WG, Becker PJ. Soft-tissue changes related tomandibular advancement surgery. Semin Orthod 2009;15:161-171.

25. Gjorup H, Athanasiou AE. Soft-tissue and dentoskeletal profile changes associated with mandibular setback osteotomy. Am J Orthod Dentofacial Orthop 1991;100:312-323.

26. Stella JP, Streater MR, Epker BN, Sinn DP. Predictability of upper lip soft tissue changes with maxillary advancement. J Oral Maxillofac Surg 1989;47:697-703.

27. Kasai K. Soft tissue adaptability to hard tissues in facial profiles. Am J Orthod Dentofacial Orthop 1998;113:674-684.

28. Chun H, Keles S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. J Roy Stat Soc B 2010;72:3-25.

29. Lee YS, Suh HY, Lee SJ, R.E. D. More accurate soft-tissue prediction model for Class III 2-

jaw surgeries. Am J Orthod Dentofacial Orthop 2014:724-733.

30. Jones RM, Khambay BS, McHugh S, Ayoub AF. The validity of a computer-assisted simulation system for orthognathic surgery (CASSOS) for planning the surgical correction of class III skeletal deformities: single-jaw versus bimaxillary surgery. Int J Oral Maxillofac Surg 2007;36:900-908.

31. Johnston C, Burden D, Kennedy D, Harradine N, Stevenson M. Class III surgical-orthodontic treatment: a cephalometric study. Am J Orthod Dentofacial Orthop 2006;130:300-309.

32. Bailey LJ, Cevidanes LH, Proffit WR. Stability and predictability of orthognathic surgery. Am J Orthod Dentofacial Orthop 2004;126:273-277.

33. Enacar A, Taner T, Toroglu S. Analysis of soft tissue profile changes associated with mandibular setback and double-jaw surgeries. Int J Adult Orthodon Orthognath Surg 1999;14:27-35.

34. Chun H, Keles S. Expression quantitative trait loci mapping with multivariate sparse partial least squares regression. Genetics 2009;182:79-90.

35. Johnson RA, Wichern DW. Applied multivariate statistical analysis. Pearson Prentice Hall; 2007.

36. Geladi P, Kowalski BR. Partial least-squares regression - a tutorial. Anal Chim Acta 1986;185:1-17.

37. Lee YS. Developing a prediction model for soft tissue changes after orthognathic surgeries in skeletal Class III patients based on the partial least squares method Department of Orthodontics, Graduate School. Seoul, Korea: Seoul National University; 2014.

38. Wold S, Ruhe A, Wold H, Dunn WJ. The collinearity problem in linear-regression - the

partial least-squares (pls) approach to generalized inverses. Siam J Sci Stat Comp 1984;5:735-743.

39. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Data mining, inference, and prediction. New York: Springer Verlag; 2009.

40. Martens M, Martens H, Wold S. Preference of cauliflower related to sensory descriptive variables by partial least-squares (pls) regression. J Sci Food Agr 1983;34:715-724.

41. Höskuldsson A. PLS regression methods. J Chemometr 1988;2:211-228.

42. Stone M, Brooks RJ. Continuum regression - cross-validated sequentially constructed prediction embracing ordinary least-squares, partial least-squares and principal components regression. J Roy Stat Soc B Met 1990;52:237-269.

43. Lindberg W, Persson JA, Wold S. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic-acid and ligninsulfonate. Anal Chem 1983;55:643-648.

44. Yoon KS, Lee HJ, Lee SJ, Donatelli RE. Testing a better method of predicting postsurgery soft tissue response in Class II patients: A prospective study and validity assessment. Angle Orthod 2014.

45. Bastien P, Bertrand F, Meyer N, Maumy-Bertrand M. Deviance residuals-based sparse PLS and sparse kernel PLS regression for censored data. Bioinformatics 2014.

46. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surveys 2010;4:40-79.

47. Faraway JJ. Practical regression and ANOVA using R. 2002.

48. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. Bioinformatics 2005;21:3301-3307.

49. Shao J. Linear-model selection by cross-validation. J Am Stat Assoc 1993;88:486-494.

50. Hulland J. Use of partial least squares (PLS) in strategic management research: a review of four recent studies. Strategic Manage J 1999;20:195-204.

51. Zhou XF, Shao Q, Coburn RA, Morris ME. Quantitative structure-activity relationship and quantitative structure-pharmacokinetics relationship of 1,4-dihydropyridines and pyridines as multidrug resistance modulators. Pharm Res 2005;22:1989-1996.

52. Tobias RD. An introduction to partial least squares. TS-509. Cary, NC: SAS Institute Inc.; 2006.

53. Mevik BH, Wehrens R. The pls package: principal component and partial least squares regression in R. J Stat Softw 2007;18:1-24.

54. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

55. Geladi P, Kowalski BR. An example of 2-block predictive partial least-squares regression with simulated data. Anal Chim Acta 1986;185:19-32.

56. Wold S. Cross-validatory estimation of number of nomponents in factor and principal components models. Technometrics 1978;20:397-405.

57. Donatelli RE, Lee SJ. How to report reliability in orthodontic research: Part 1. Am J Orthod Dentofacial Orthop 2013;144:156-161.

58. Donatelli RE, Lee SJ. How to report reliability in orthodontic research: Part 2. Am J Orthod Dentofacial Orthop 2013;144:315-318.

59. Jung MH. Age, extraction rate and jaw surgery rate in Korean orthodontic clinics and small dental hospitals. Korean J Orthod 2012;42:80-86.

60. Mobarak KA, Espeland L, Krogstad O, Lyberg T. Soft tissue profile changes following mandibular advancement surgery: predictability and long-term outcome. Am J Orthod

Dentofacial Orthop 2001;119:353-367.

61. Donatsky O, Bjorn-Jorgensen J, Hermund NU, Nielsen H, Holmqvist-Larsen M, Nerder PH. Immediate postoperative outcome of orthognathic surgical planning, and prediction of positional changes in hard and soft tissue, independently of the extent and direction of the surgical corrections required. Br J Oral Maxillofac Surg 2011;49:386-391.

62. Pektas ZO, Kircelli BH, Cilasun U, Uckan S. The accuracy of computer-assisted surgical planning in soft tissue prediction following orthognathic surgery. Int J Med Robot 2007;3:64-71.

63. Jeon J, Kim Y, Kim J, Kang H, Ji H, Son W. New bimaxillary orthognathic surgery planning and model surgery based on the concept of six degrees of freedom. Korean J Orthod 2013;43:42-52.

64. Kim JH, Kim KB, Kim WC, Kim JH, Kim HY. Accuracy and precision of polyurethane dental arch models fabricated using a three-dimensional subtractive rapid prototyping method with an intraoral scanning technique. Korean J Orthod 2014;44:69-76.

65. Wu D, Li L, Zhang L. Feature constrained compressed sensing CT image reconstruction from incomplete data via robust principal component analysis of the database. Phys Med Biol 2013;58:4047-4070.

66. Hancock PJ, Burton AM, Bruce V. Face processing: human perception and principal components analysis. Mem Cognit 1996;24:21-40.

67. Kim JY, Lee SJ, Kim TW, Nahm DS, Chang YI. Classification of the skeletal variation in normal occlusion. Angle Orthod 2005;75:311-319.

68. Hwang HS, Youn IS, Lee KH, Lim HJ. Classification of facial asymmetry by cluster analysis. Am J Orthod Dentofacial Orthop 2007;132:279 e271-276.

Table 1-1. The subjects′ age, sex and other characteristics.

Variables		N	Mean	Standard deviation	Minimum	Maximum
Age (years)						
	Female	191	24.1	5.3	16.0	51.6
	Male	127	23.6	3.5	18.8	39.1
Time after surgery (months)						
			9.5	4.1	3.7	30.5
Maxillary surgery						
	No	91				
	Yes	227				
Genioplasty						
	No	139				
	Yes	179				

Table 1-2. The Features of Class II Subjects and Class III Subjects.

Variables		Class II Subjects			Class III Subjects		
		n	Mean	SD[a]	n	Mean	SD[a]
Maxillary surgery							
	No	20			71		
	Yes	94			133		
Genioplasty	No	16			123		
	Yes	98			81		
Overjet before surgery (mm)			7.7	2.4		-5.8	3.8
Overbite before surgery (mm)			2.7	3.1		-0.2	1.8
[b] Amount of surgical repositioning at point A (mm)							
	Anteroposterior repositioning		-0.2	2.1		1.4	1.9
	Vertical repositioning		-1.5	3.1		-1.1	2.3
[b] Amount of surgical repositioning at point B (mm)							
	Anteroposterior repositioning		6.0	3.9		-7.3	3.8
	Vertical repositioning		-1.0	4.7		-2.9	4.2

[a] SD indicates standard deviation; [b] a positive value indicates forward and downward in the anteroposterior- and vertical direction respectively, a negative value indicates either posterior direction or superior direction during surgical repositioning.

Table 2. The Mean Absolut Error of Prediction from partial least squares (PLS) and sparse partial least squares (SPLS) prediction models.

		Sparse PLSR													
	PLSR	Eta													
Response		0.1	0.2	0.3	0.4	0.5	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.99
	MAEP														
	(SE)														
Horizontal															
Pronasale	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.91	0.91	0.91	0.90	0.87	0.88	0.97
	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04
Upper lip	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.20	1.20	1.19	1.18	1.20
	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.07
Lower lip	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.08	1.08	1.20
	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
Pogonion	1.24	1.24	1.24	1.24	1.24	1.24	1.23	1.23	1.23	1.23	1.23	1.23	1.26	1.31	1.23
	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
Menton	2.09	2.09	2.09	2.09	2.09	2.09	2.08	2.09	2.08	2.09	2.09	2.12	2.10	2.07	2.05
	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.21	0.19	0.19
Vertical															
Pronasale	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.99	0.99	0.99	0.99	1.00	0.97	1.03
	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.04	0.05	0.04	0.05
Upper lip	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.14	1.16	1.19
	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Lower lip	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.25	1.28	1.27	1.30
	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
Pogonion	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.59	1.60	1.59	1.58	1.59
	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Menton	1.27	1.27	1.27	1.27	1.27	1.27	1.26	1.27	1.27	1.29	1.27	1.28	1.33	1.33	1.33
	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08

Table 3. The number of predictor variables in PLS method and the number of selected predictor variables in SPLS method as eta increases.

		Sparse PLSR													
	PLSR	Eta													
		0.1	0.2	0.3	0.4	0.5	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.99
Mean	232	232	232	232	232	232	229	227	220	207	181	147	104	76	34
(SE)	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.06	0.18	0.13	0.16	0.19	0.12	0.15	0.02

Table 4. Result of *t* test (p-value) to compare soft tissue prediction errors between partial least squares (PLS) and sparse partial least squares (SPLS) prediction methods.

		Sparse PLSR													
	PLSR	Eta													
		0.1	0.2	0.3	0.4	0.5	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.99
Horizontal															
Pronasale		0.50	0.50	0.50	0.50	0.50	0.50	0.47	0.48	0.44	0.47	0.48	0.49	0.52	0.44
Upper lip		0.50	0.50	0.50	0.50	0.50	0.50	0.51	0.48	0.53	0.49	0.53	0.54	0.52	0.52
Lower lip		0.50	0.50	0.50	0.50	0.50	0.49	0.51	0.49	0.53	0.47	0.49	0.55	0.41	0.51
Pogonion		0.50	0.50	0.50	0.50	0.50	0.49	0.49	0.51	0.53	0.47	0.50	0.56	0.40	0.47
Menton		0.50	0.50	0.50	0.50	0.50	0.49	0.48	0.49	0.48	0.44	0.42	0.45	0.45	0.45
Vertical															
Pronasale		0.50	0.50	0.50	0.50	0.50	0.50	0.47	0.53	0.52	0.51	0.53	0.55	0.48	0.46
Upper lip		0.50	0.50	0.50	0.50	0.50	0.50	0.51	0.54	0.50	0.48	0.46	0.65	0.56	0.50
Lower lip		0.50	0.50	0.50	0.50	0.50	0.50	0.51	0.49	0.49	0.49	0.46	0.60	0.47	0.51
Pogonion		0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.52	0.48	0.49	0.49	0.51	0.48	0.51
Menton		0.50	0.50	0.50	0.50	0.50	0.51	0.50	0.49	0.50	0.53	0.57	0.42	0.39	0.56

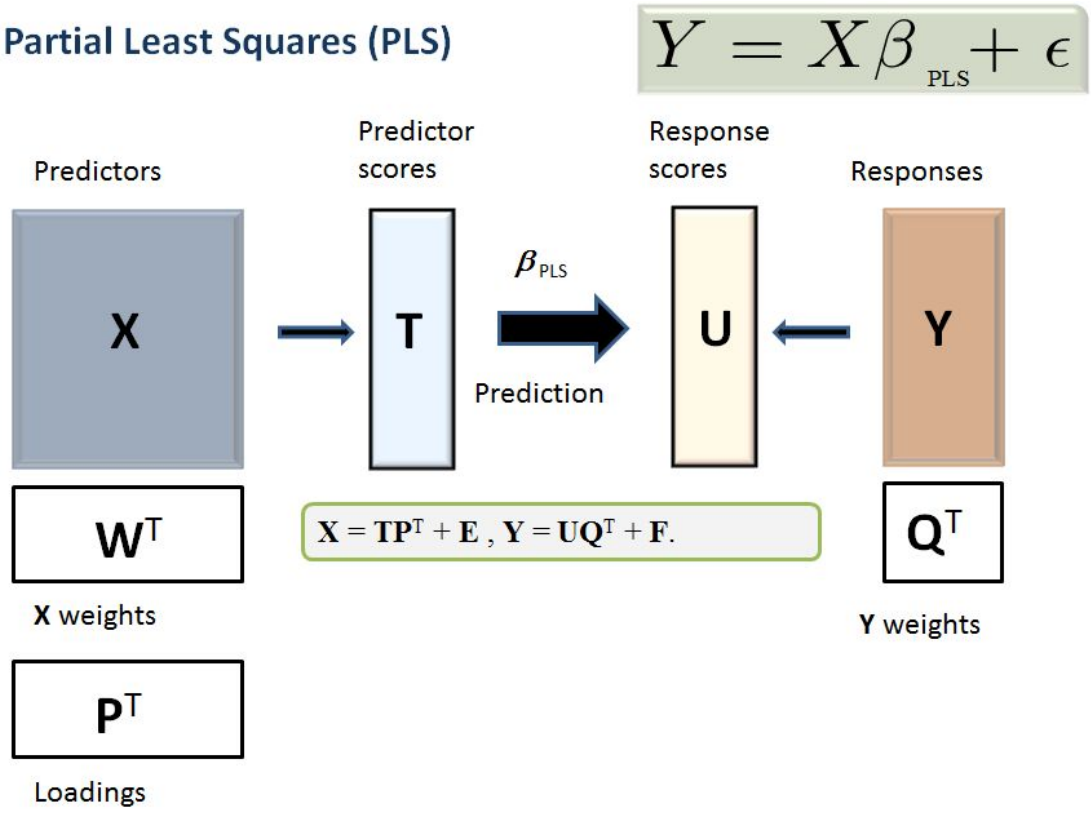

Figure 1. Figure illustrating partial least squares method. W^T, the matrix of weights for X; P^T, the matrix of loadings for X; T, the matrix of X scores; U, the matrix of Y scores; Q^T, the matrix of weights for Y.

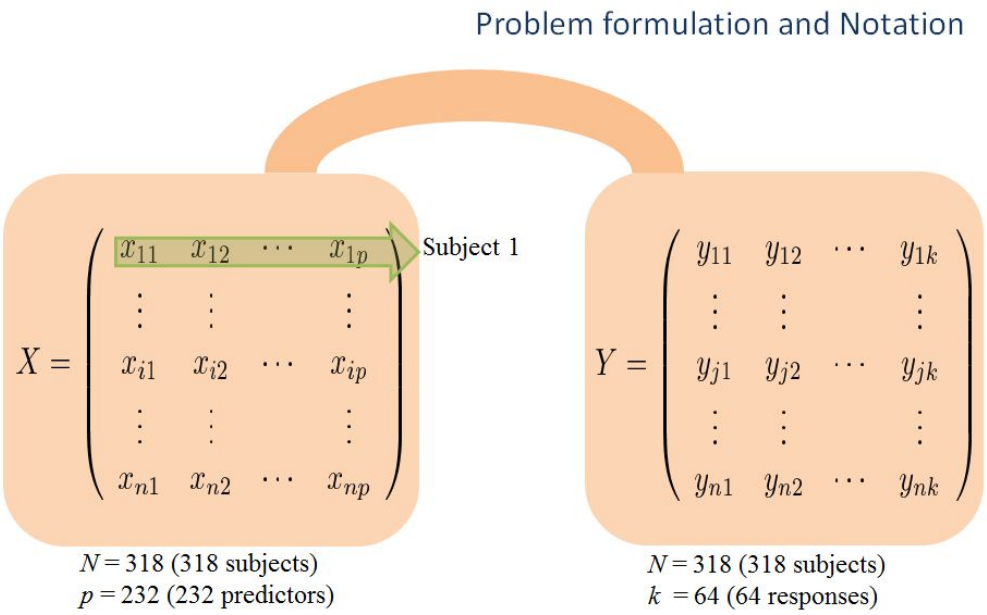

Figure 2. X and Y matrices formulated in the present study.

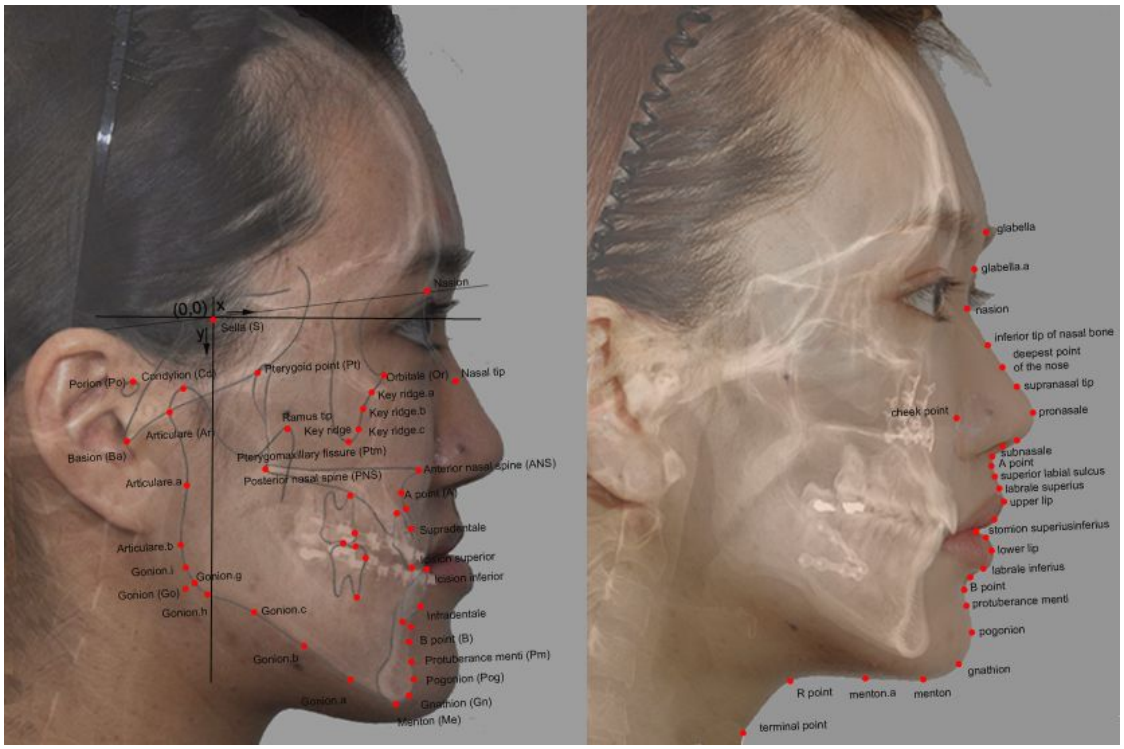

Figure 3. Diagram showing reference planes and cephalometric landmarks used in present study. ***Left,*** image composed from preoperative radiograph, with hard tissue landmarks in capital letters. ***Right,*** soft tissue landmarks in lowercase letters shown on the follow-up cephalogram.

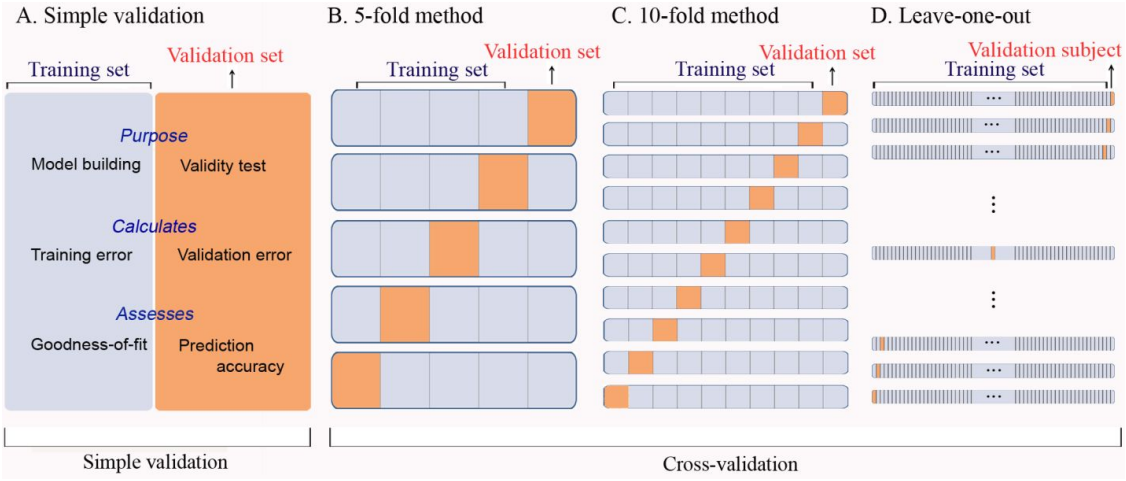

Figure 4. Schematic diagrams illustrating the validation methods applied in the present study. Simple validation (**A**) utilizes separate training and validation datasets. In 5-fold cross-validation **(B),** the dataset is divided into 5 portions. Each portion serves as a validation dataset in each round. In 10-fold cross-validation (**C**), the whole dataset has 10 portions for each training and validation trial. In leave-one-out cross-validation (**D**), each subject serves as a validation dataset.

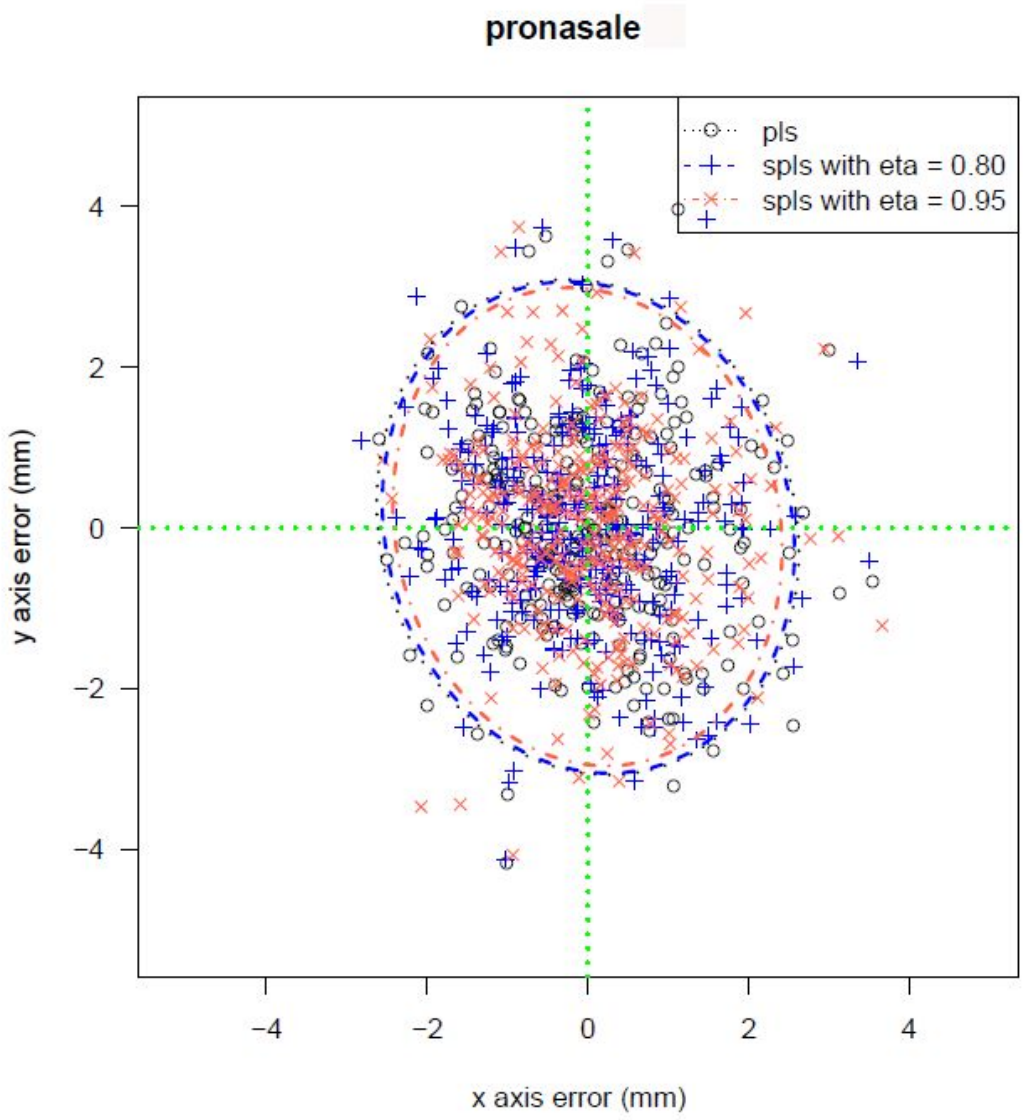

Figure 5, A. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (circle), SPLS with eta = 0.80 (plus) and SPLS with eta = 0.95 (cross) prediction methods. The plots clearly indicate that the bias in both *x*- and *y*-axes was not significantly different between PLS and SPLS methods. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

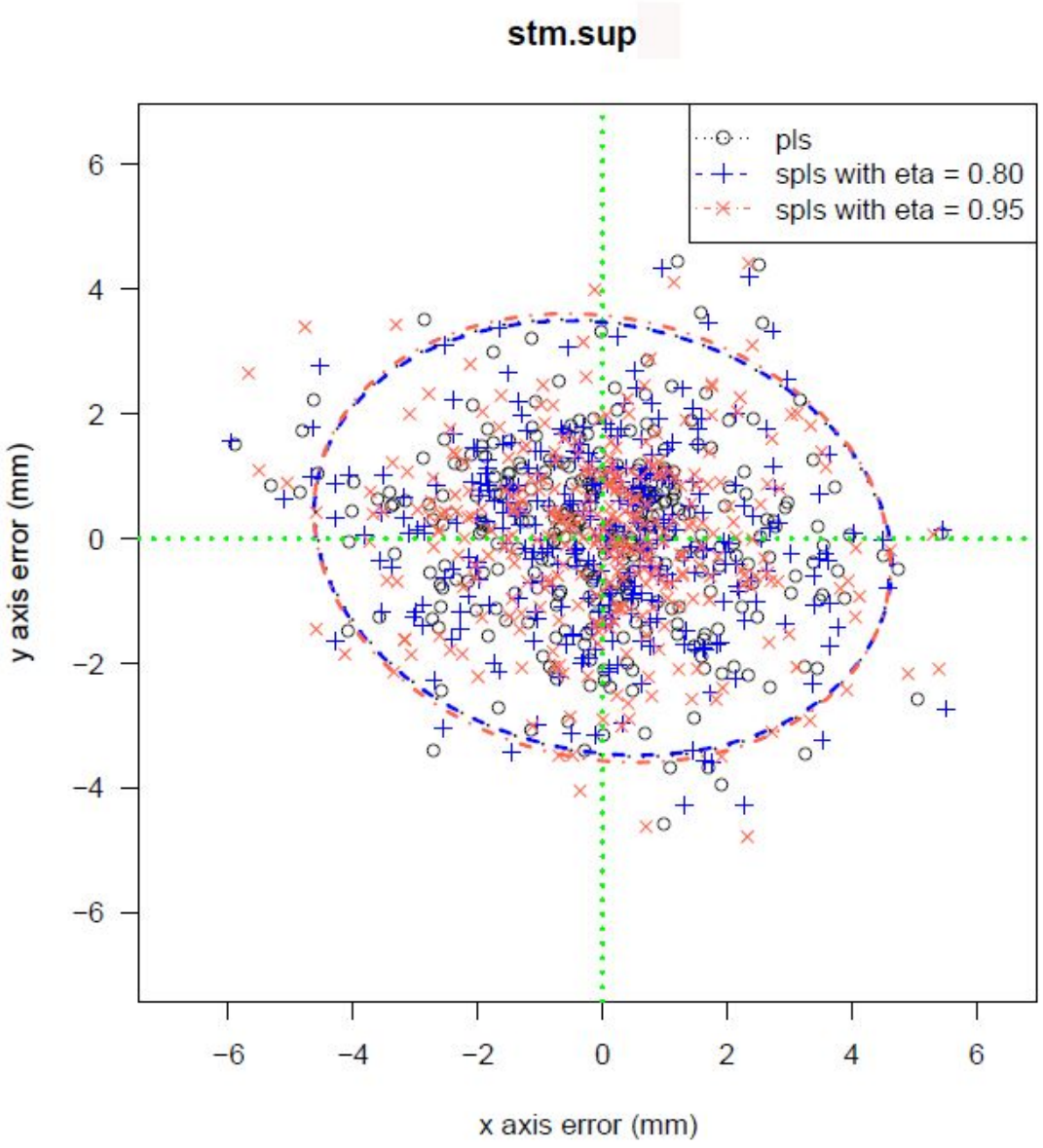

Figure 5, B. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (circle), SPLS with eta = 0.80 (plus) and SPLS with eta = 0.95 (cross) prediction methods. The plots clearly indicate that the bias in both *x*- and *y*-axes was not significantly different between PLS and SPLS methods. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

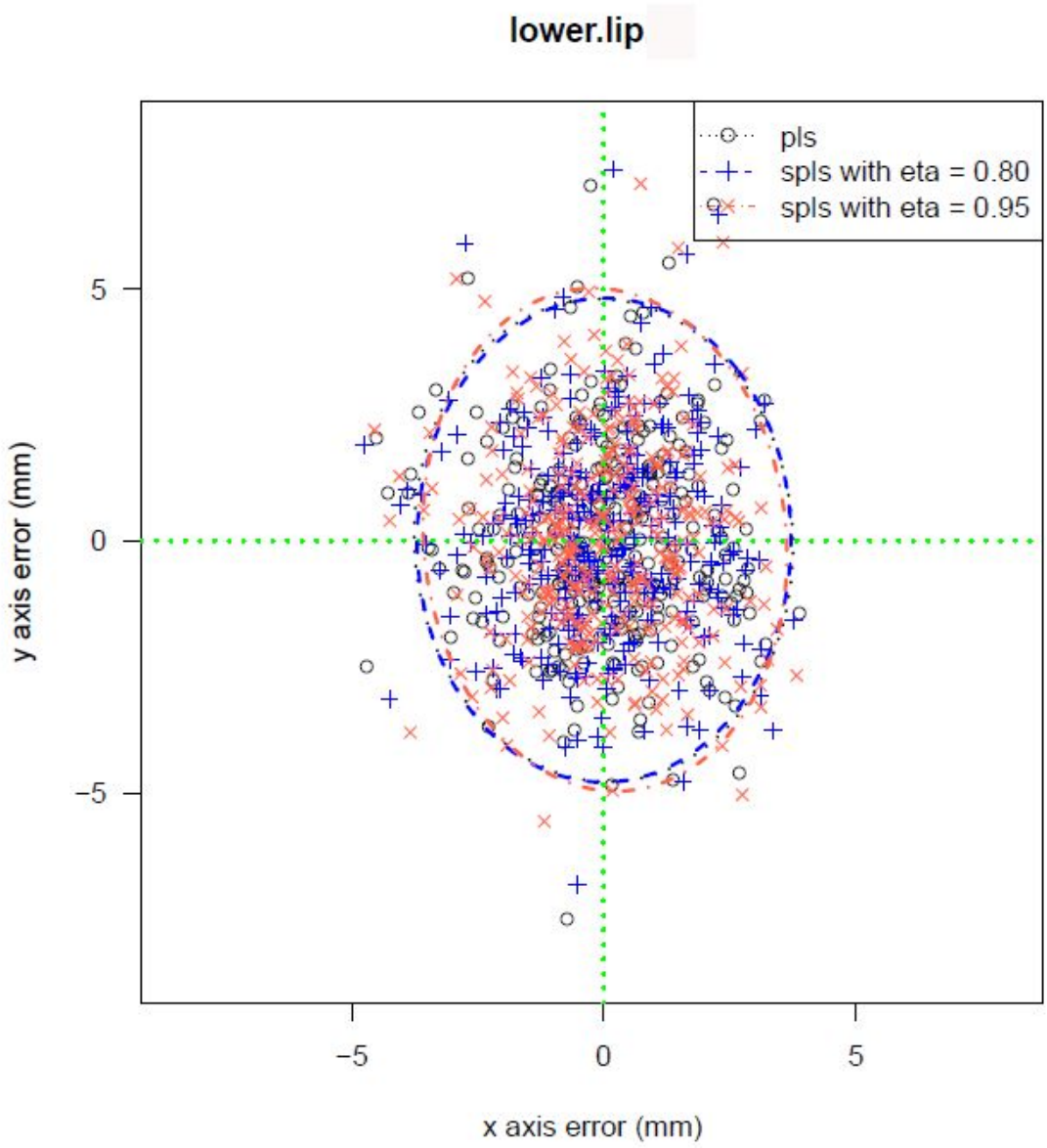

Figure 5, C. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (circle), SPLS with eta = 0.80 (plus) and SPLS with eta = 0.95 (cross) prediction methods. The plots clearly indicate that the bias in both *x*- and *y*-axes was not significantly different between PLS and SPLS methods. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

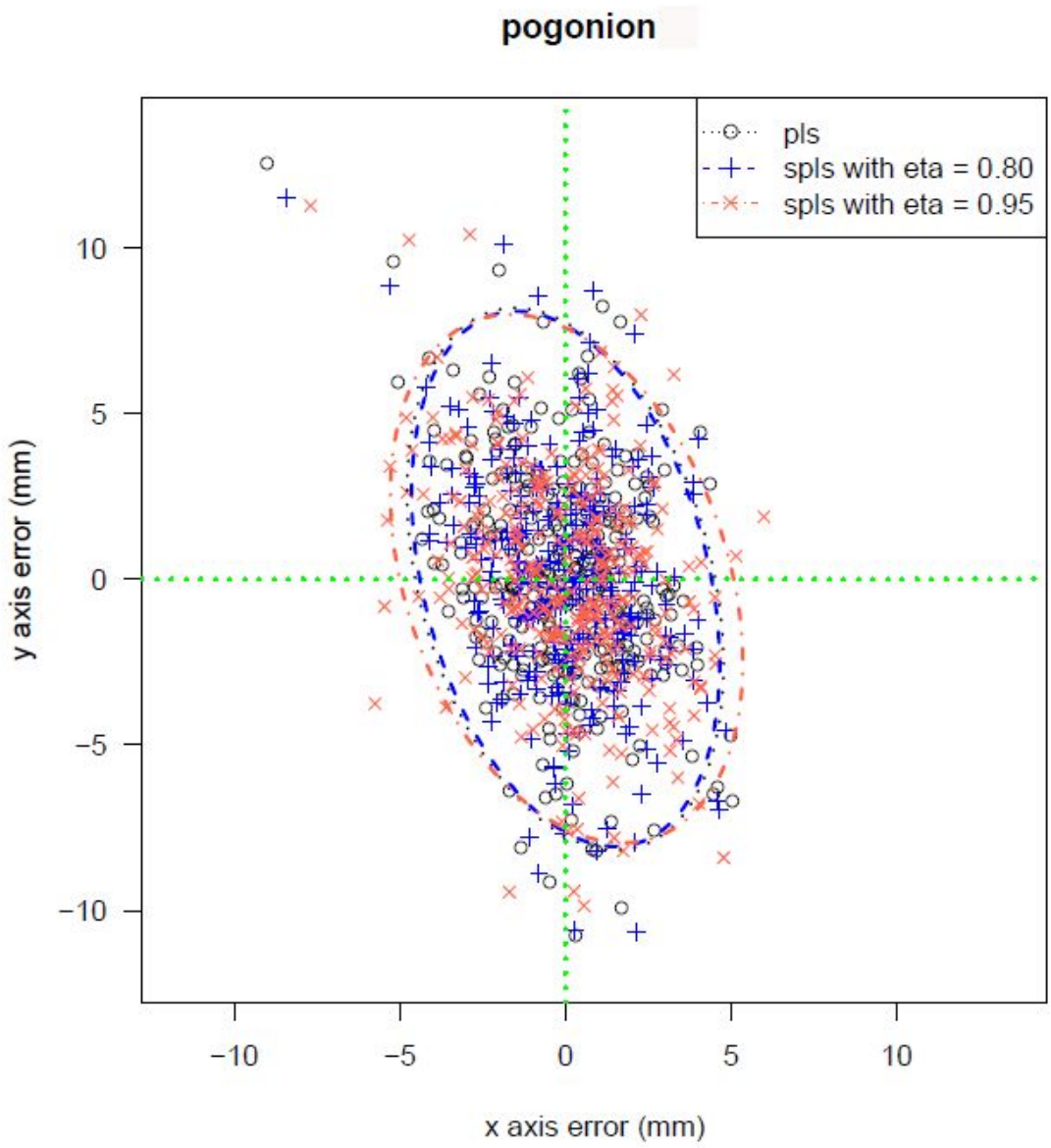

Figure 5, D. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (circle), SPLS with eta = 0.80 (plus) and SPLS with eta = 0.95 (cross) prediction methods. The plots clearly indicate that the bias in both *x*- and *y*-axes was not significantly different between PLS and SPLS methods. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

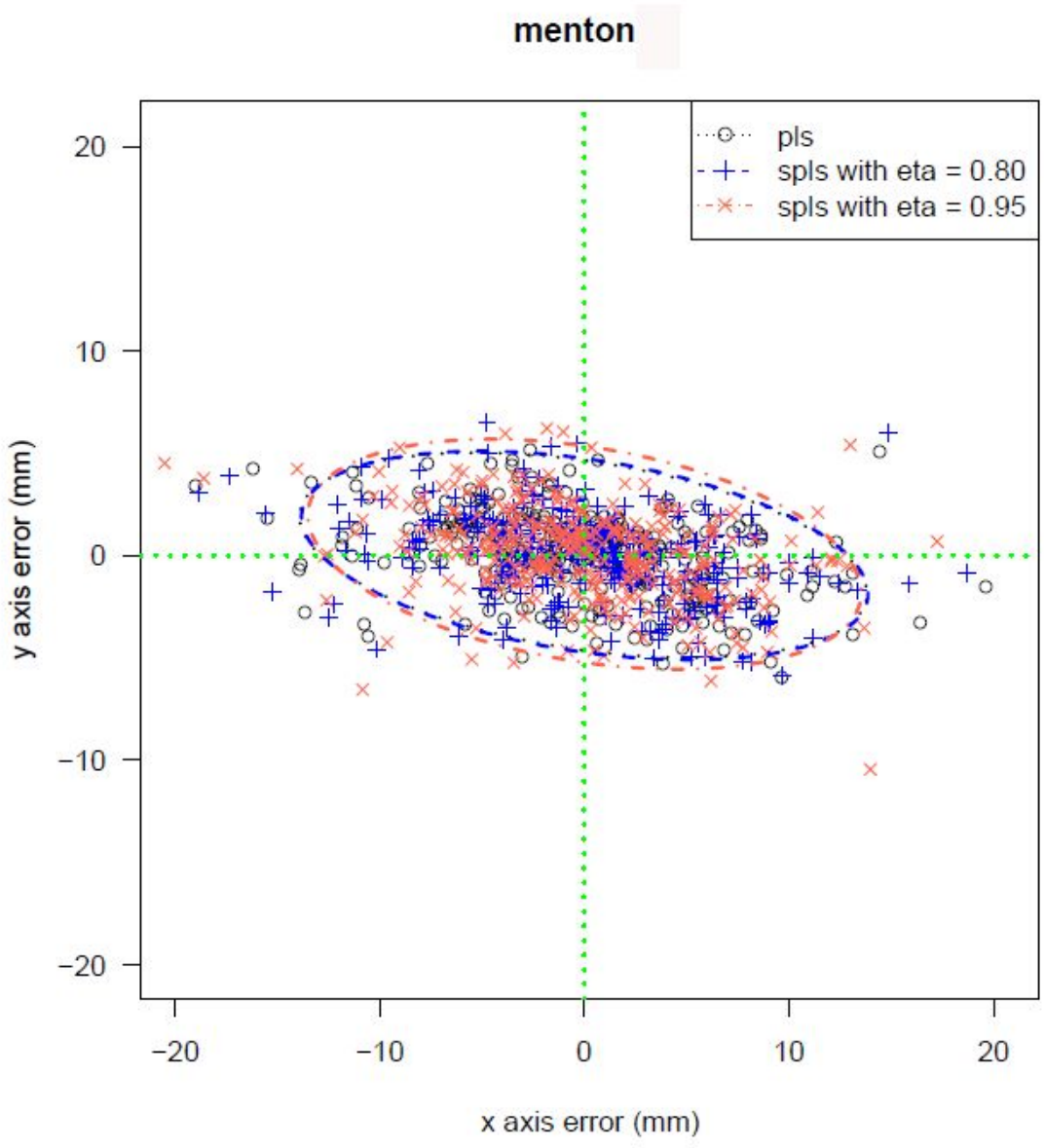

Figure 5, E. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (circle), SPLS with eta = 0.80 (plus) and SPLS with eta = 0.95 (cross) prediction methods. The plots clearly indicate that the bias in both *x*- and *y*-axes was not significantly different between PLS and SPLS methods. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

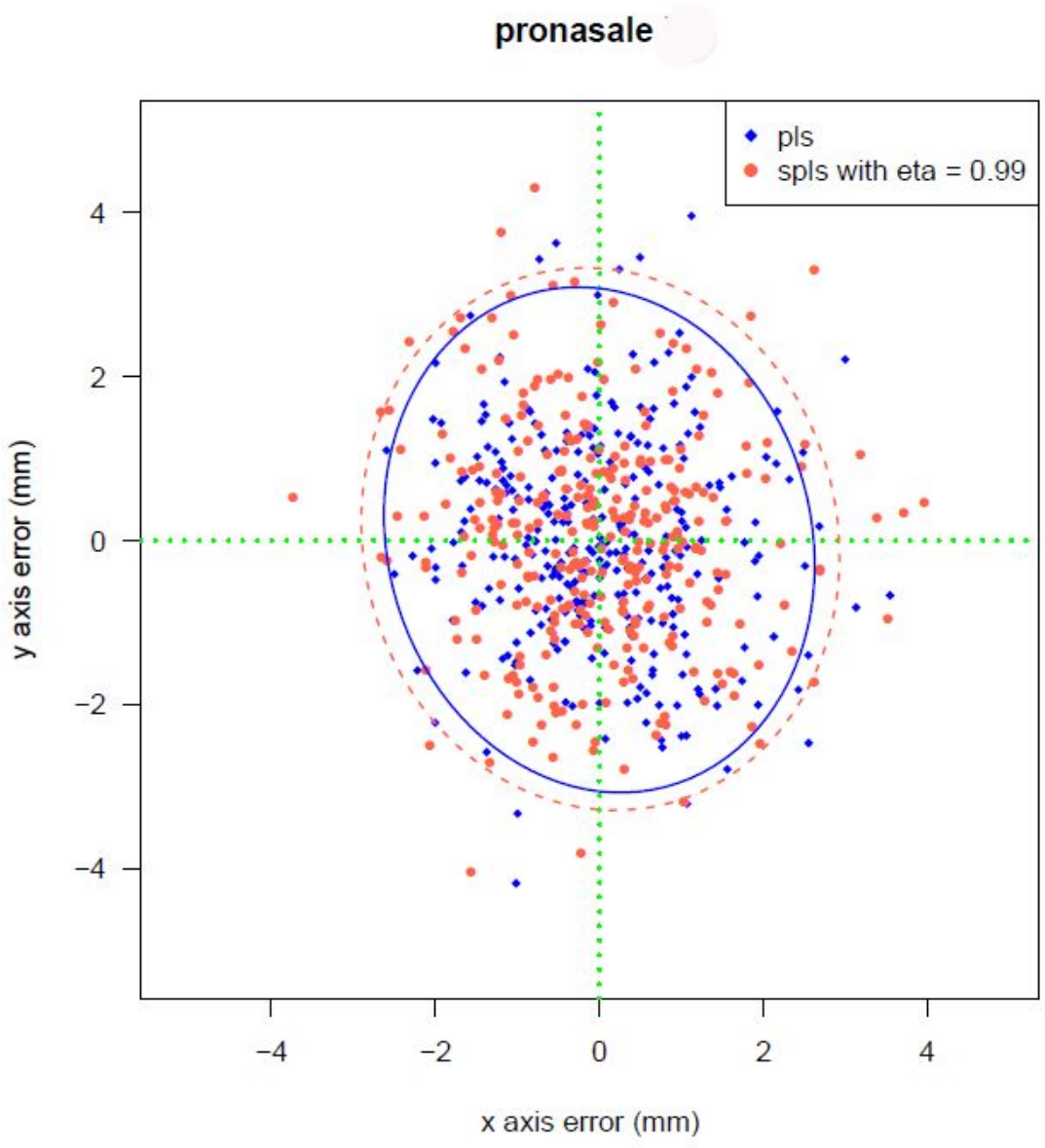

Figure 6, A. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. The plots clearly indicate that even at the eta=0.99 situation, the bias of SPLS method in both *x*- and *y*-axes was not significantly different from the bias of PLS method. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

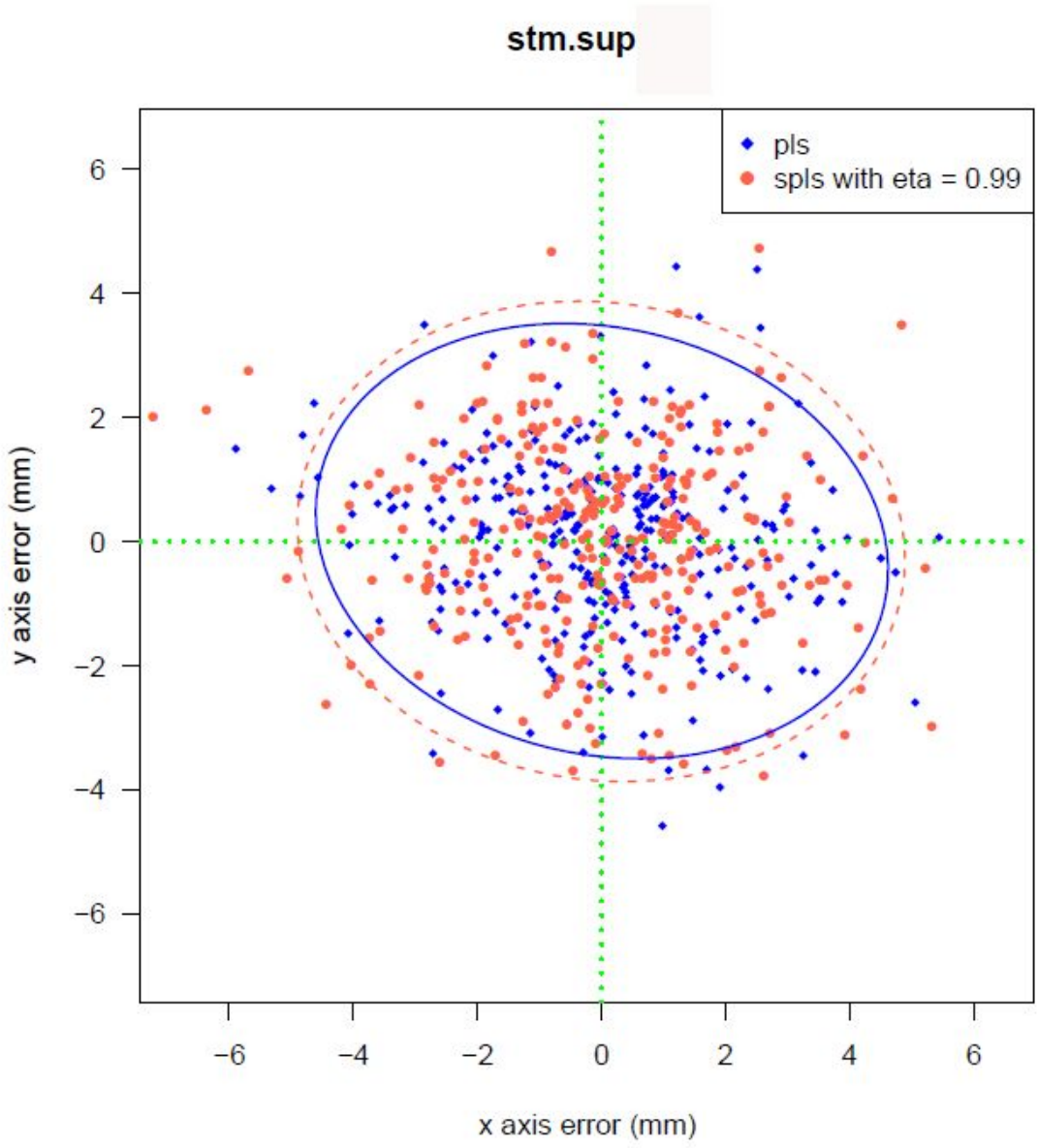

Figure 6, B. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. The plots clearly indicate that even at the eta=0.99 situation, the bias of SPLS method in both *x*- and *y*-axes was not significantly different from the bias of PLS method. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

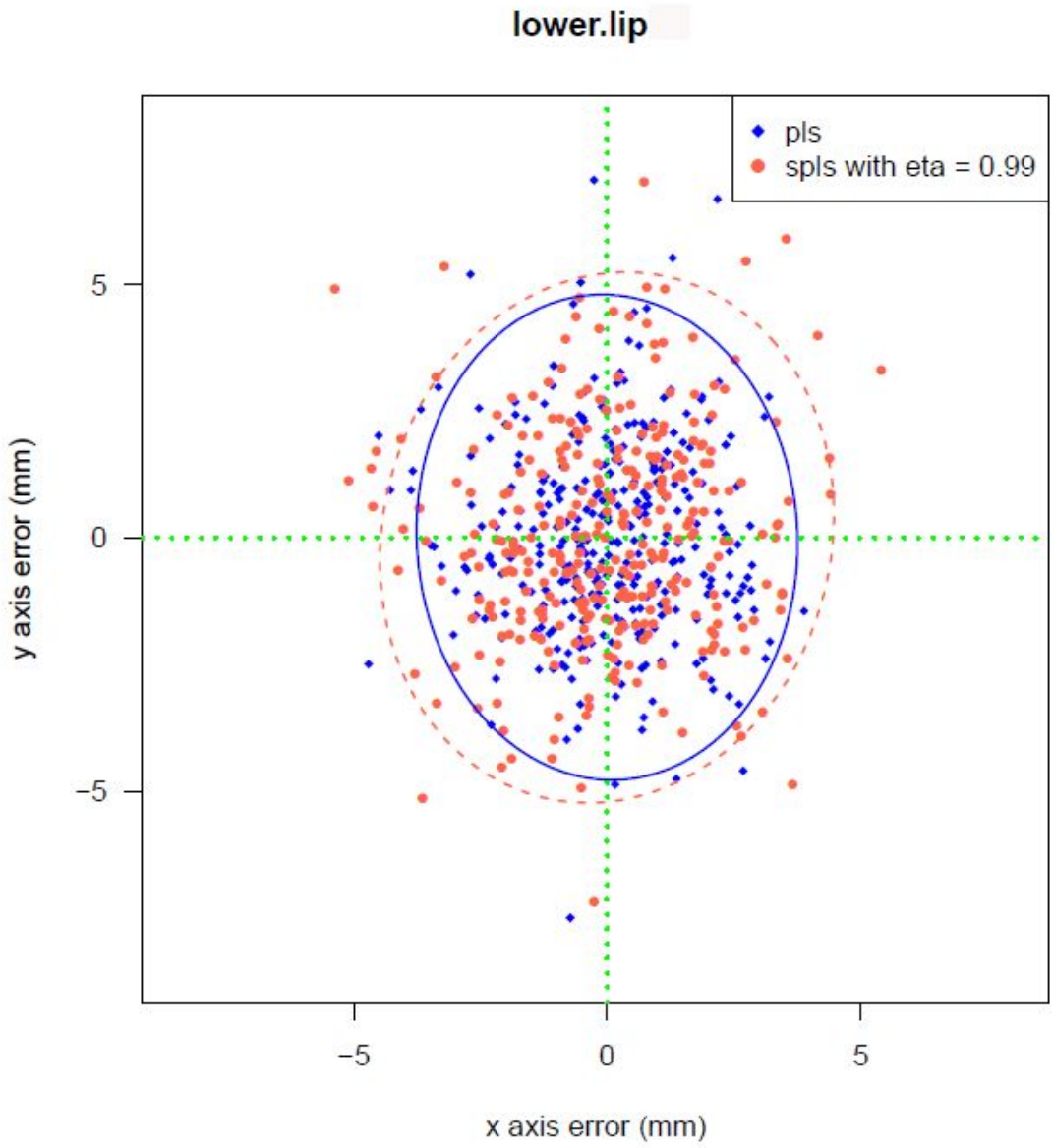

Figure 6, C. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. The plots clearly indicate that even at the eta=0.99 situation, the bias of SPLS method in both *x*- and *y*-axes was not significantly different from the bias of PLS method. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

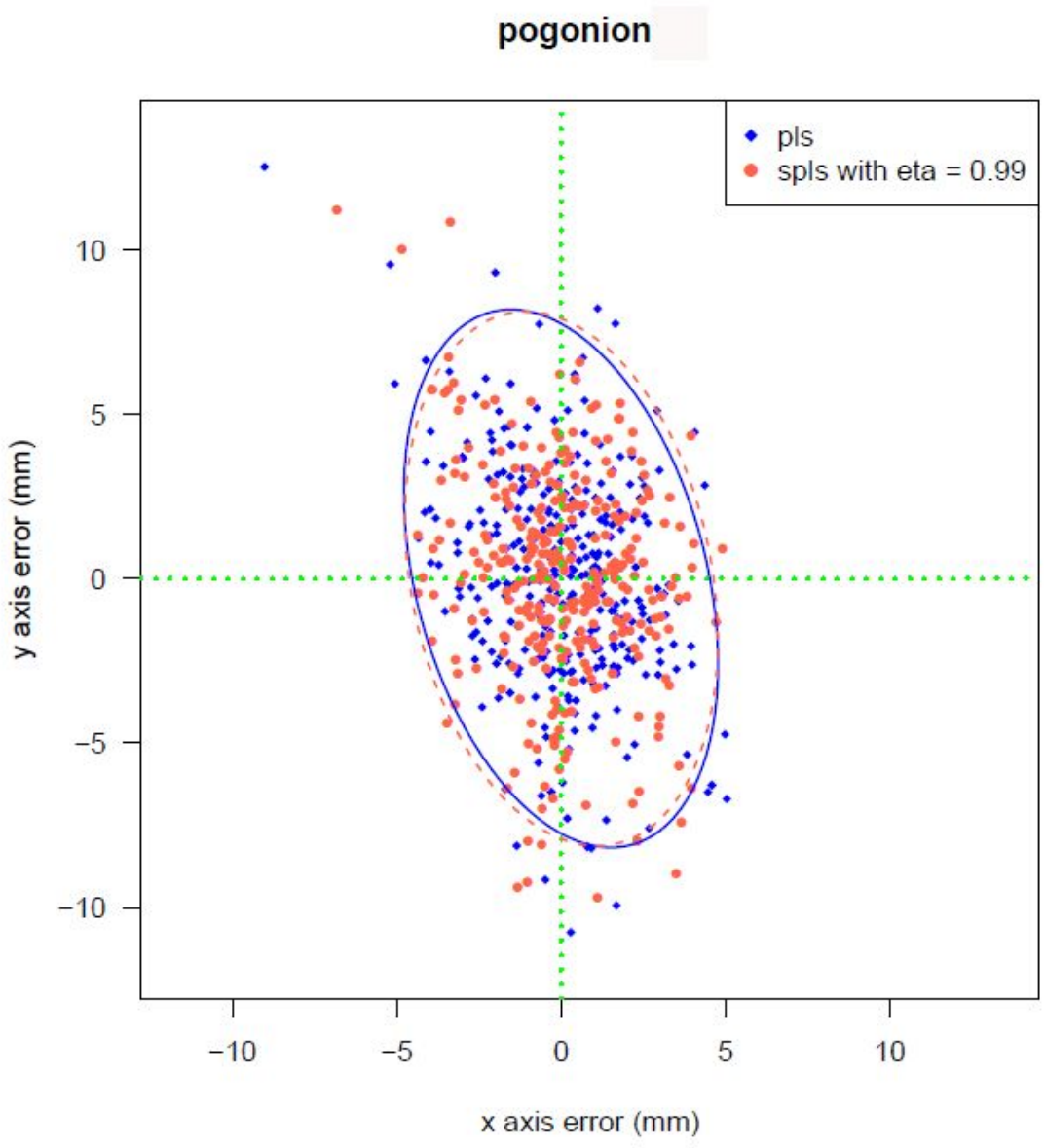

Figure 6, D. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. The plots clearly indicate that even at the eta=0.99 situation, the bias of SPLS method in both *x*- and *y*-axes was not significantly different from the bias of PLS method. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

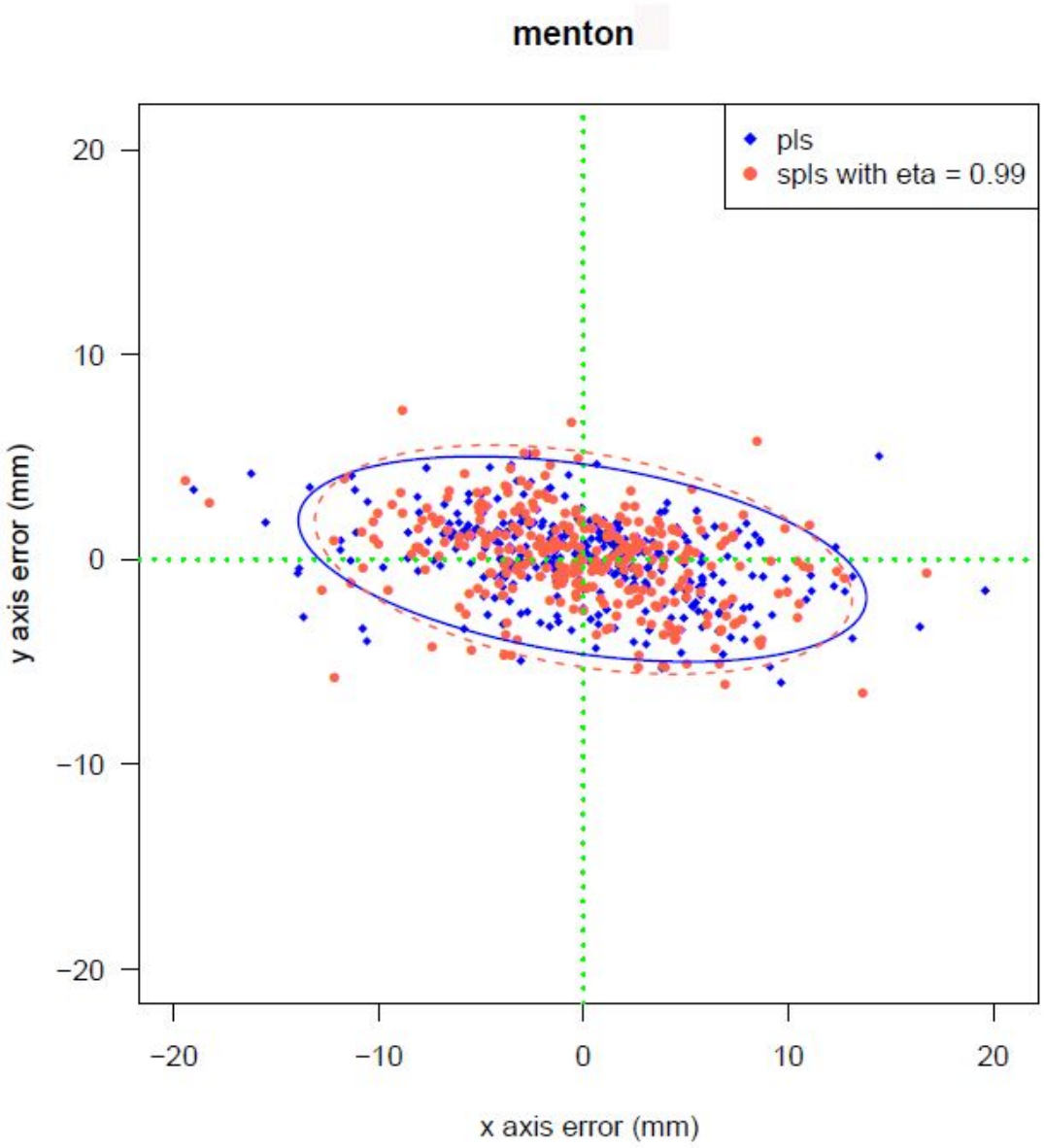

Figure 6, E. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. The plots clearly indicate that even at the eta=0.99 situation, the bias of SPLS method in both *x*- and *y*-axes was not significantly different from the bias of PLS method. Application of the equations to individuals may give rise to errors to this extent in 95% probability.

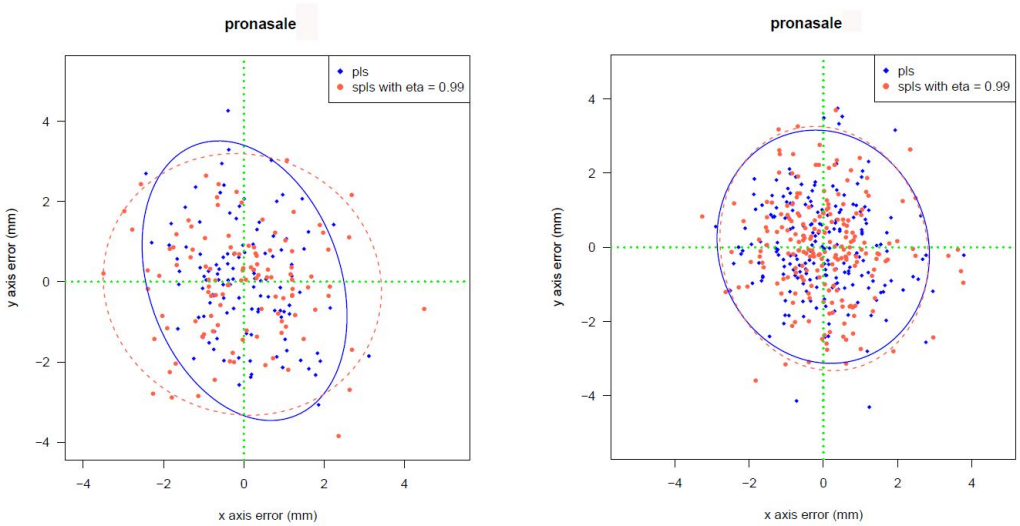

Figure 7, A. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of Class II patients. ***Right,*** plots drawn from dataset of Class III patients.

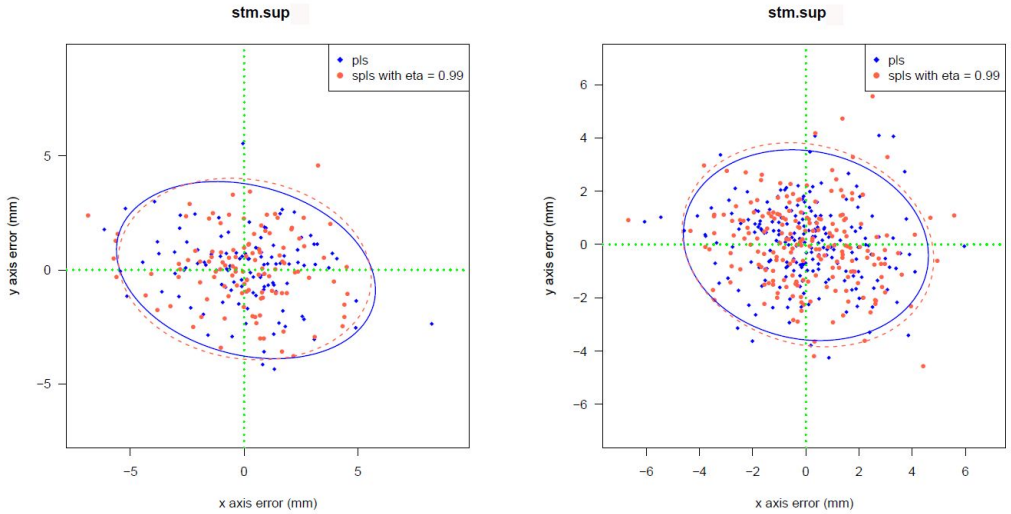

Figure 7, B. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of Class II patients. ***Right,*** plots drawn from dataset of Class III patients.

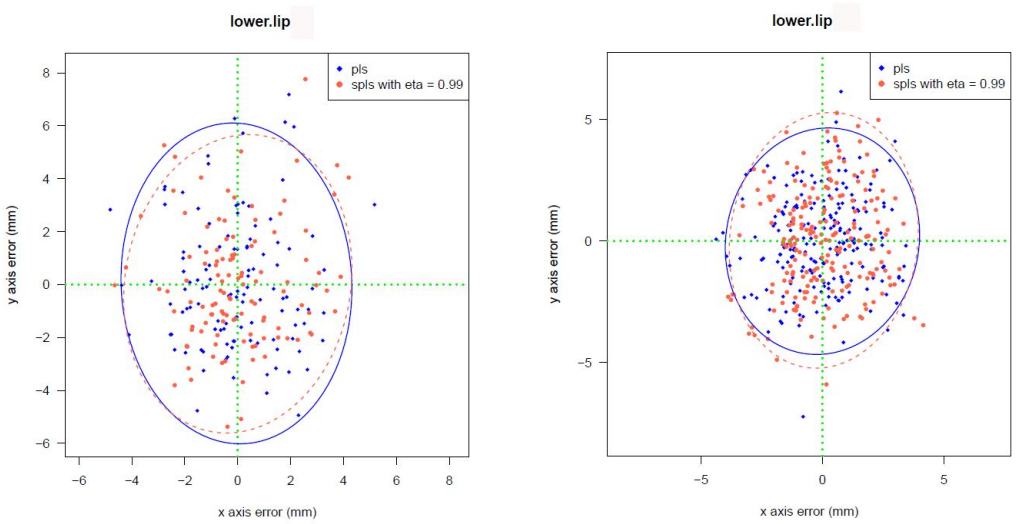

Figure 7, C. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of Class II patients. ***Right,*** plots drawn from dataset of Class III patients.

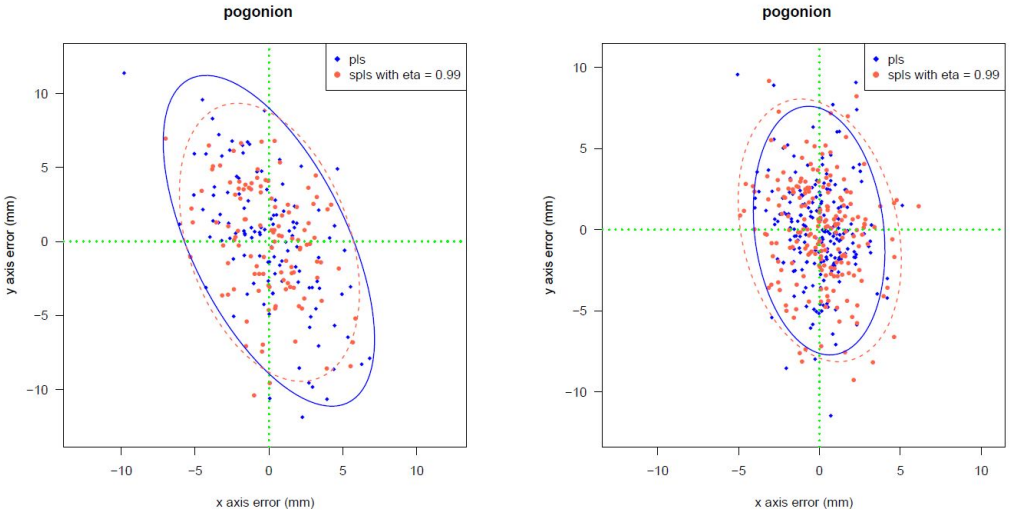

Figure 7, D. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of Class II patients. ***Right,*** plots drawn from dataset of Class III patients.

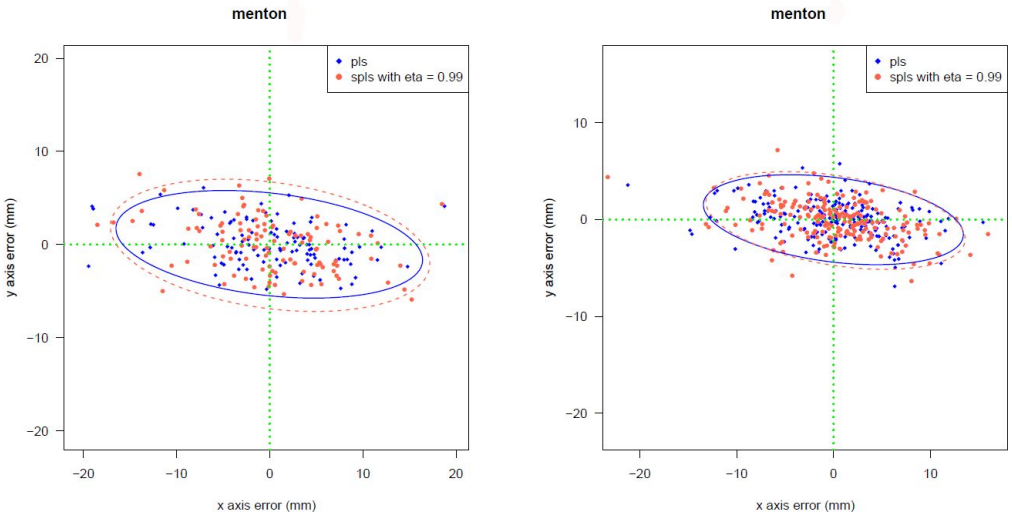

Figure 7, E. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of Class II patients. ***Right,*** plots drawn from dataset of Class III patients.

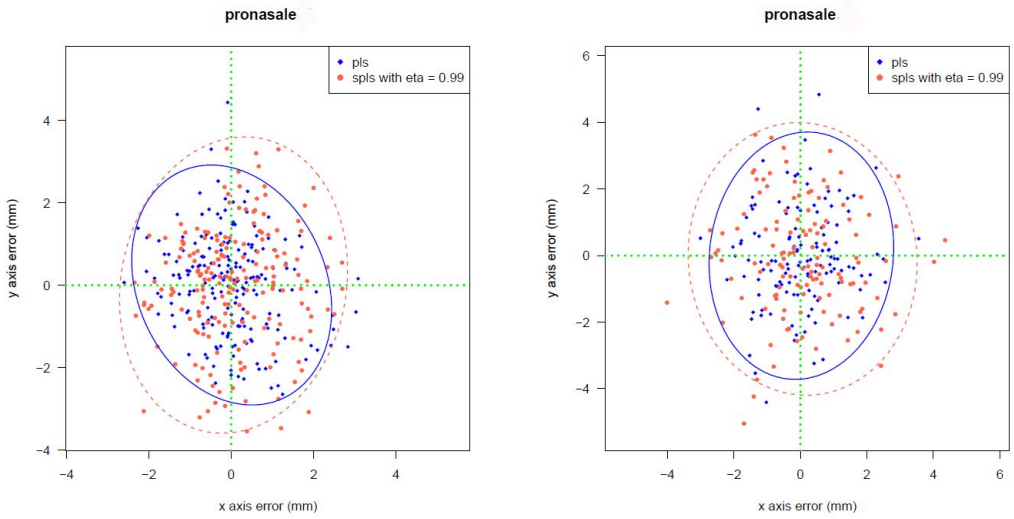

Figure 8, A. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of female subjects. ***Right,*** plots drawn from dataset of male subjects.

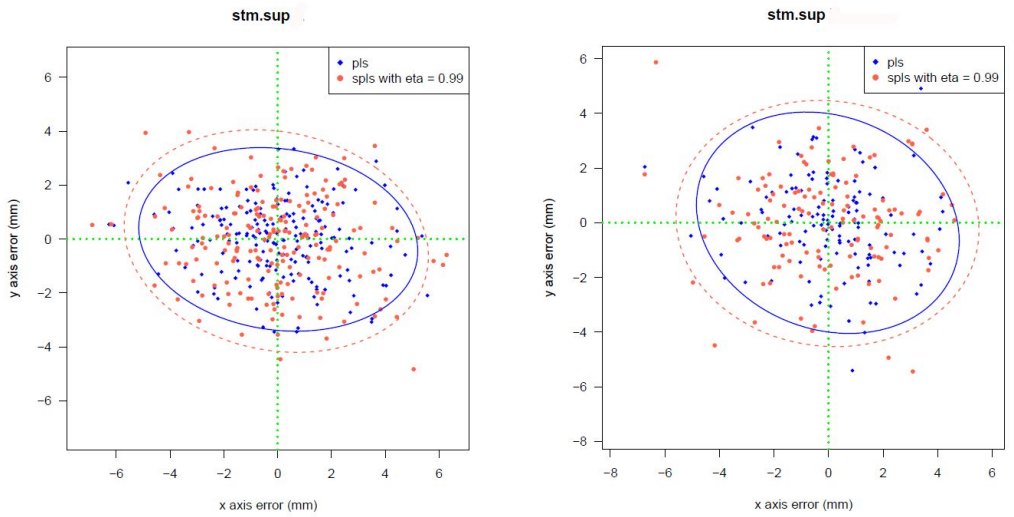

Figure 8, B. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of female subjects. ***Right,*** plots drawn from dataset of male subjects.

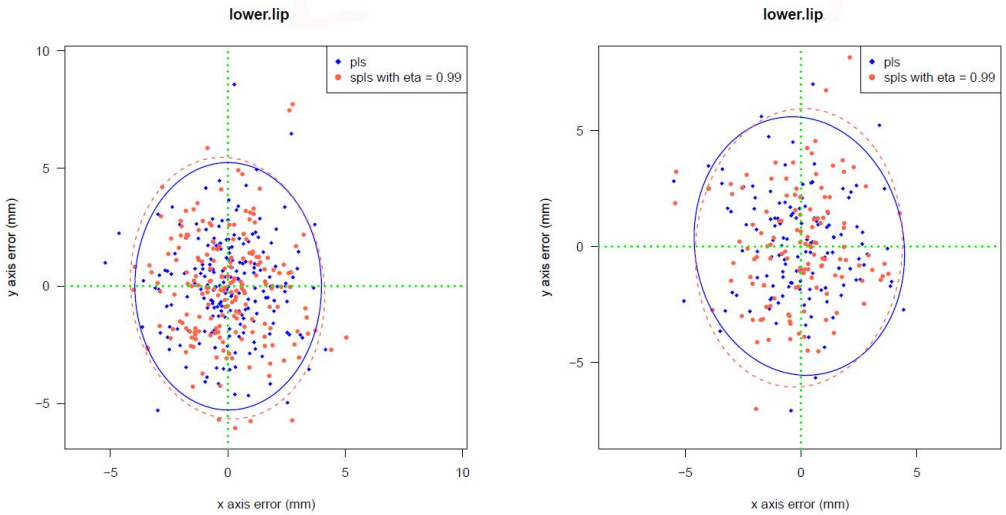

Figure 8, C. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of female subjects. ***Right,*** plots drawn from dataset of male subjects.

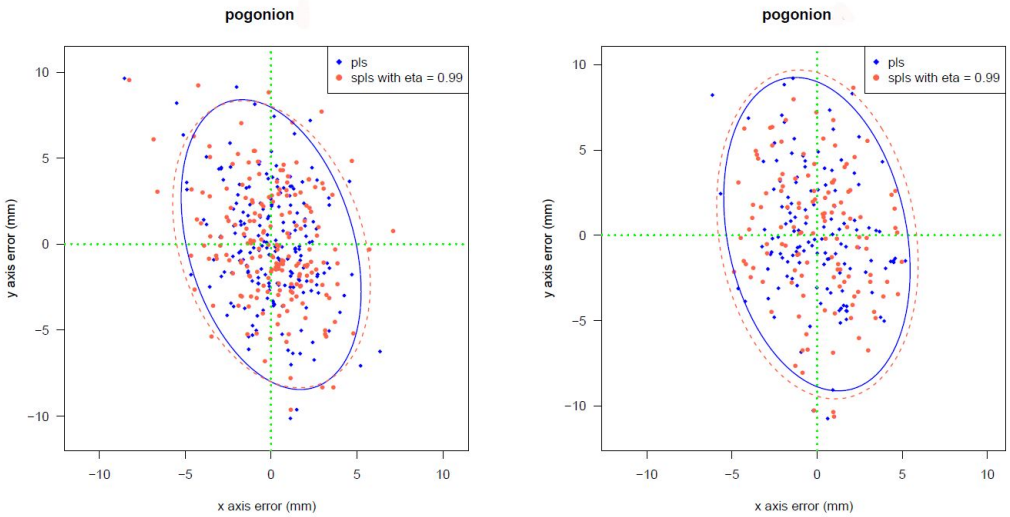

Figure 8, D. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of female subjects. ***Right,*** plots drawn from dataset of male subjects.

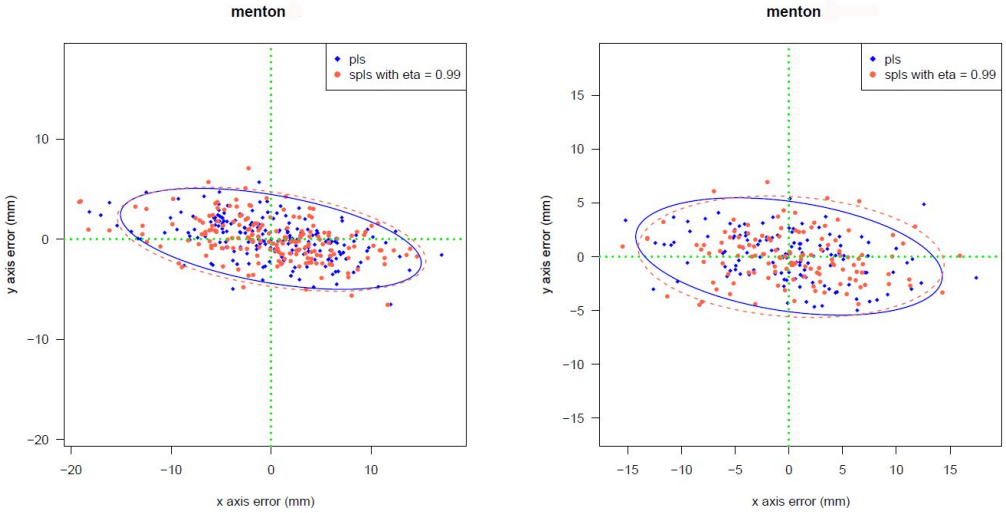

Figure 8, E. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of female subjects. ***Right,*** plots drawn from dataset of male subjects.

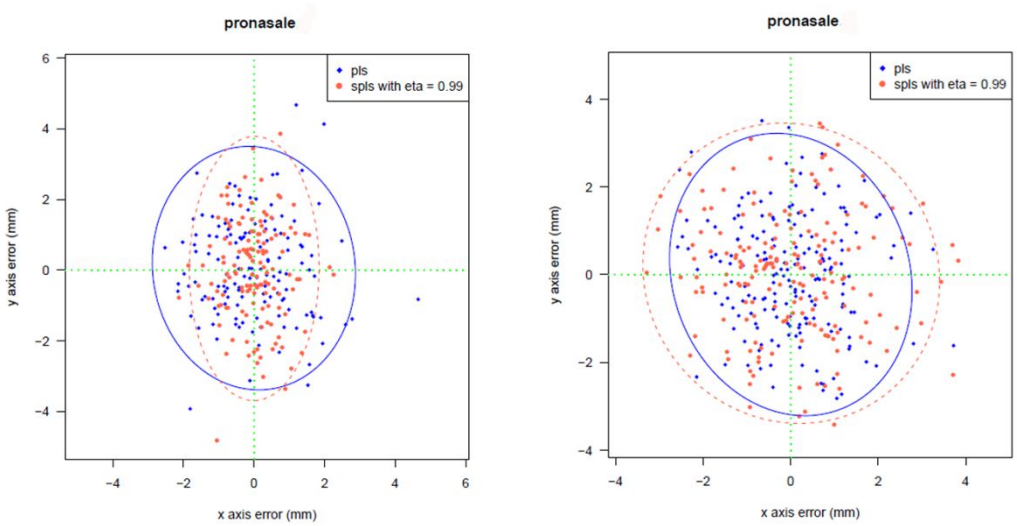

Figure 9, A. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients had no genioplasty. ***Right,*** plots drawn from dataset of patients had additional genioplasty.

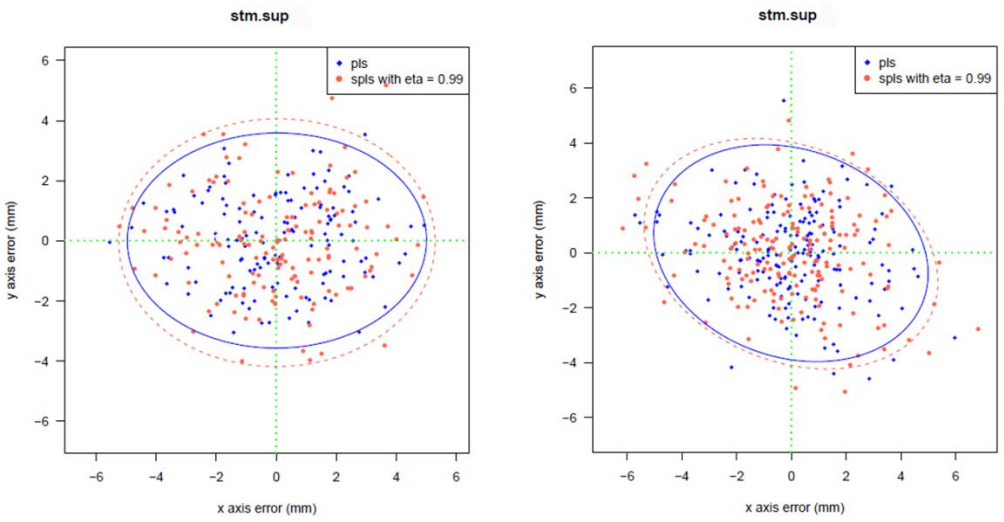

Figure 9, B. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients had no genioplasty. ***Right,*** plots drawn from dataset of patients had additional genioplasty.

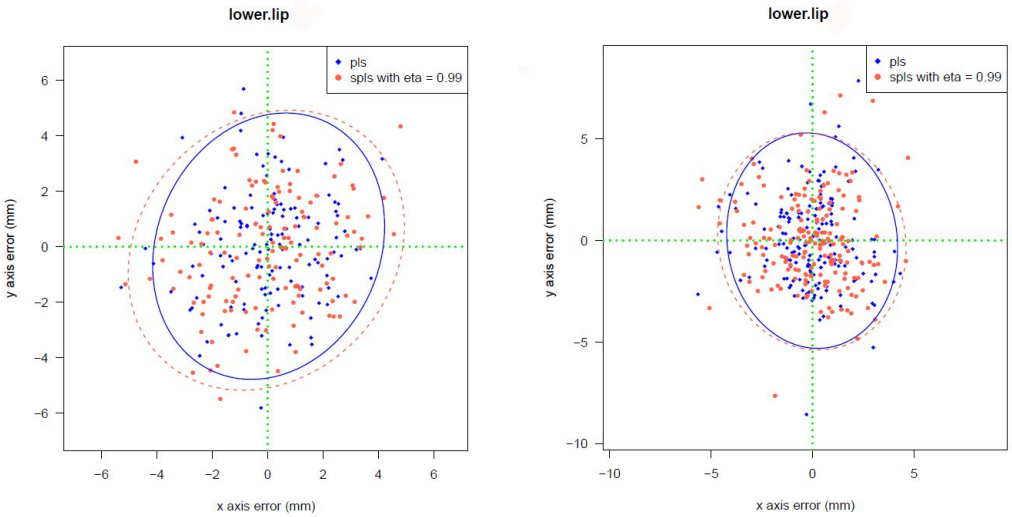

Figure 9, C. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients had no genioplasty. ***Right,*** plots drawn from dataset of patients had additional genioplasty.

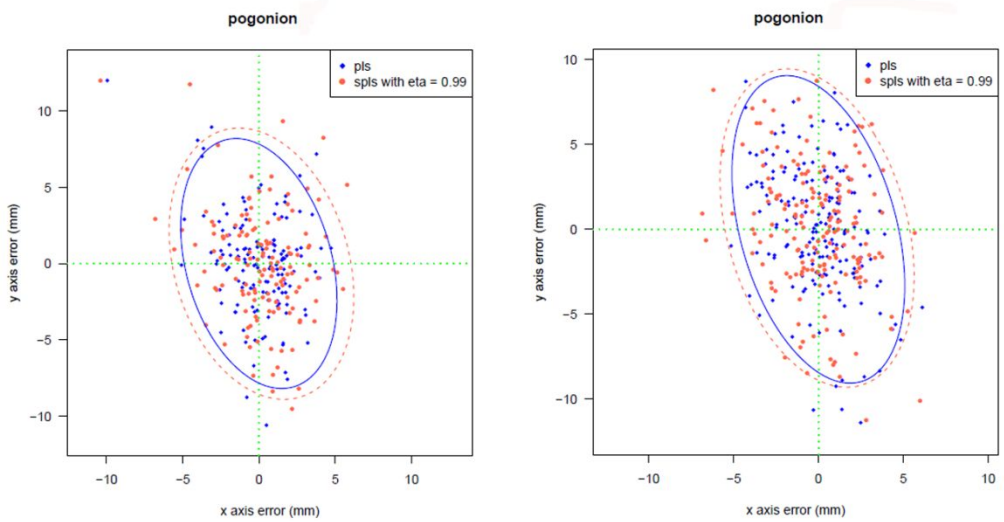

Figure 9, D. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients had no genioplasty. ***Right,*** plots drawn from dataset of patients had additional genioplasty.

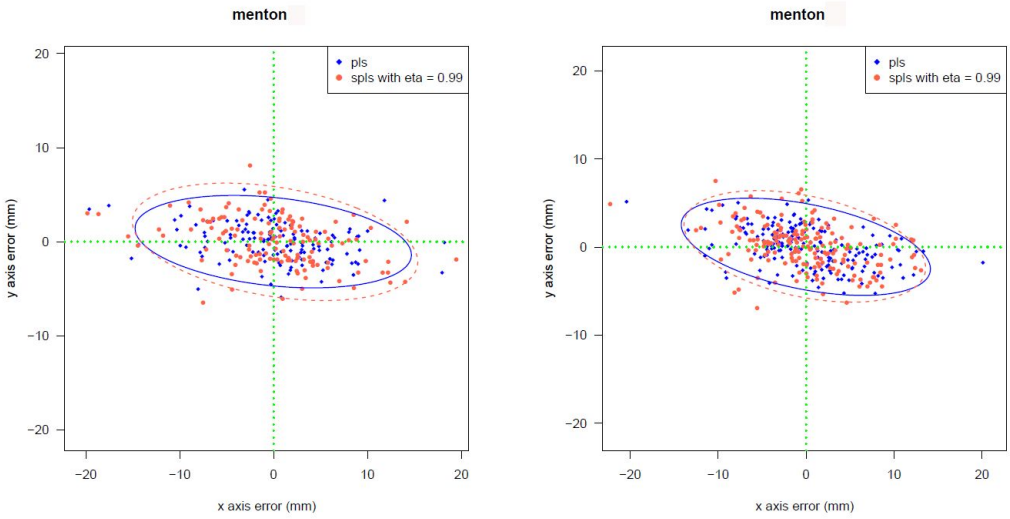

Figure 9, E. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients had no genioplasty. ***Right,*** plots drawn from dataset of patients had additional genioplasty.

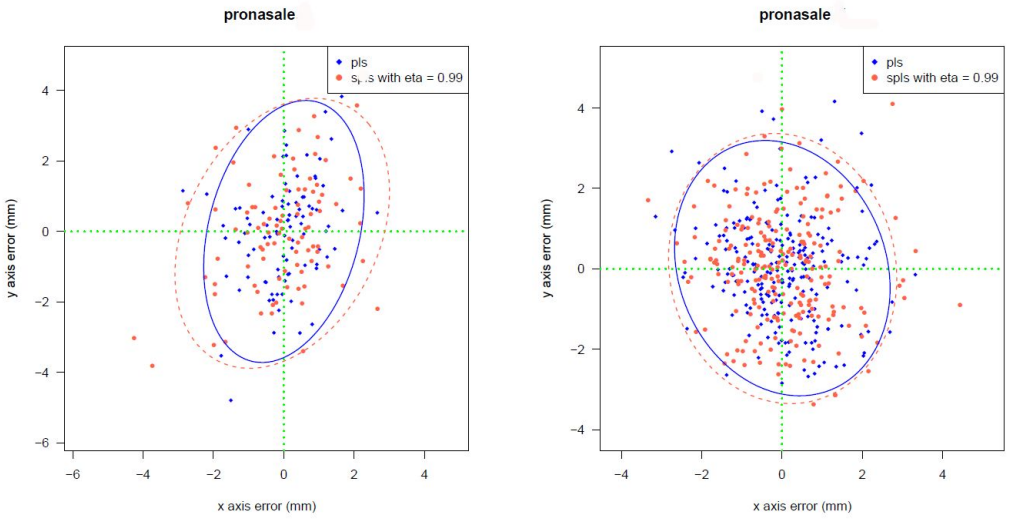

Figure 10, A. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients underwent mandibular surgery only. ***Right,*** plots drawn from dataset of patients underwent maxillary surgery.

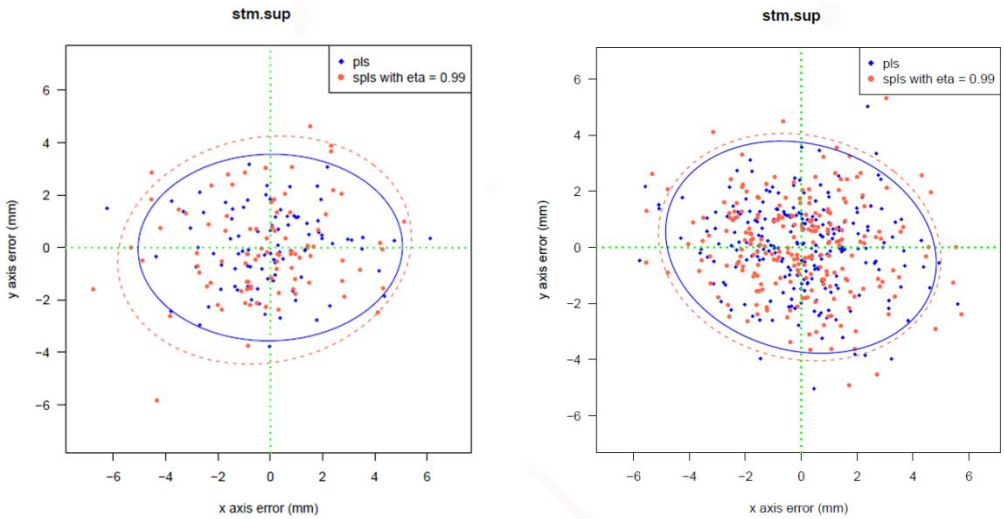

Figure 10, B. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients underwent mandibular surgery only. ***Right,*** plots drawn from dataset of patients underwent maxillary surgery.

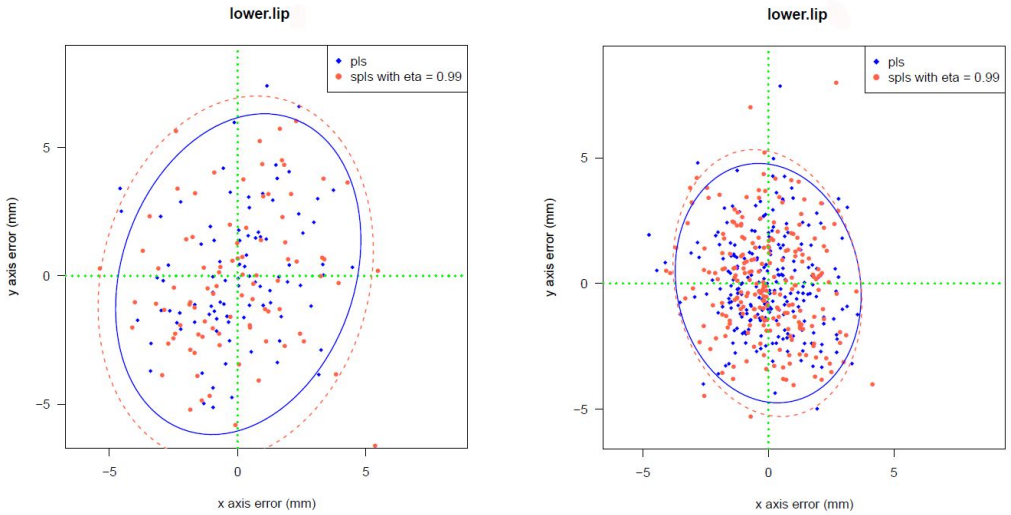

Figure 10, C. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients underwent mandibular surgery only. ***Right,*** plots drawn from dataset of patients underwent maxillary surgery.

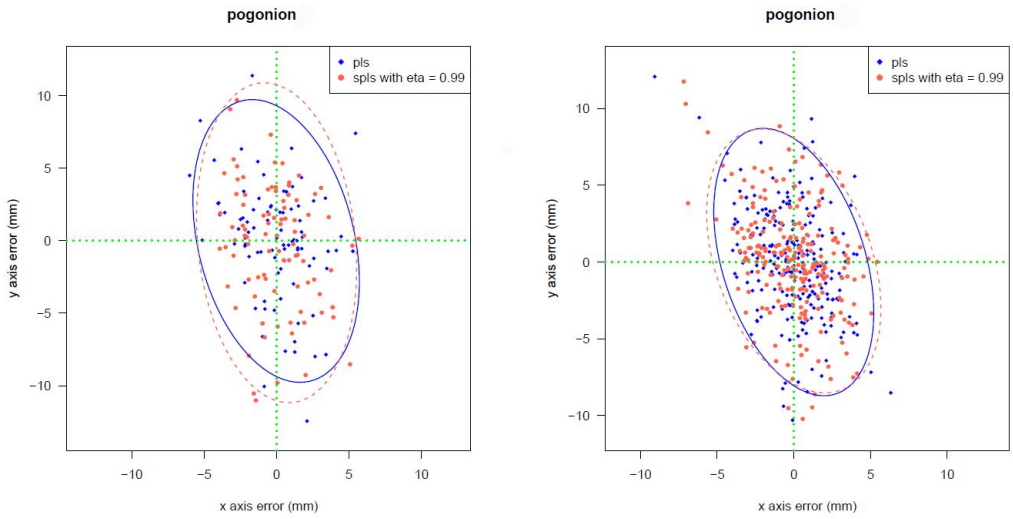

Figure 10, D. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients underwent mandibular surgery only. ***Right,*** plots drawn from dataset of patients underwent maxillary surgery.

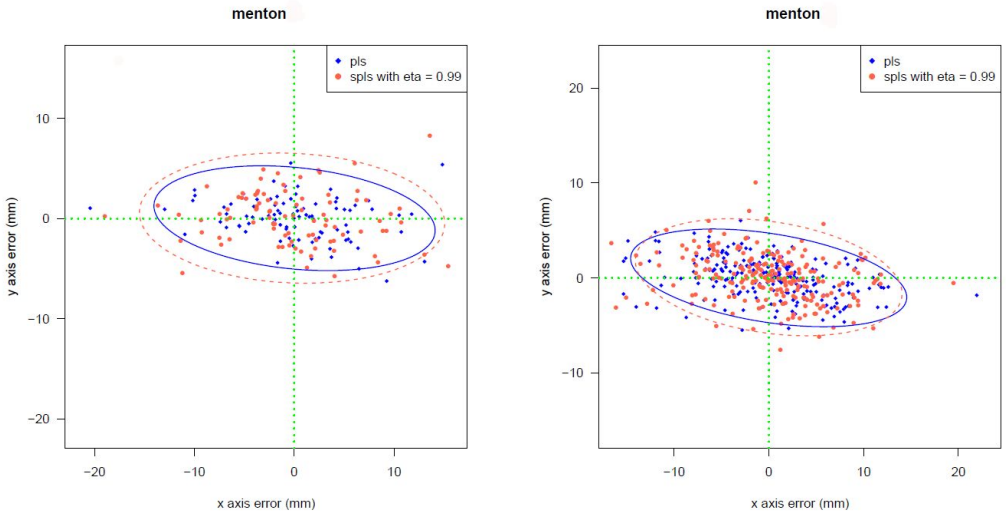

Figure 10, E. Scattergrams and 95% confidence ellipses for the bias that were obtained from the PLS (blue) and SPLS with eta = 0.99 (red) prediction methods. ***Left,*** plots drawn from dataset of patients underwent mandibular surgery only. ***Right,*** plots drawn from dataset of patients underwent maxillary surgery.

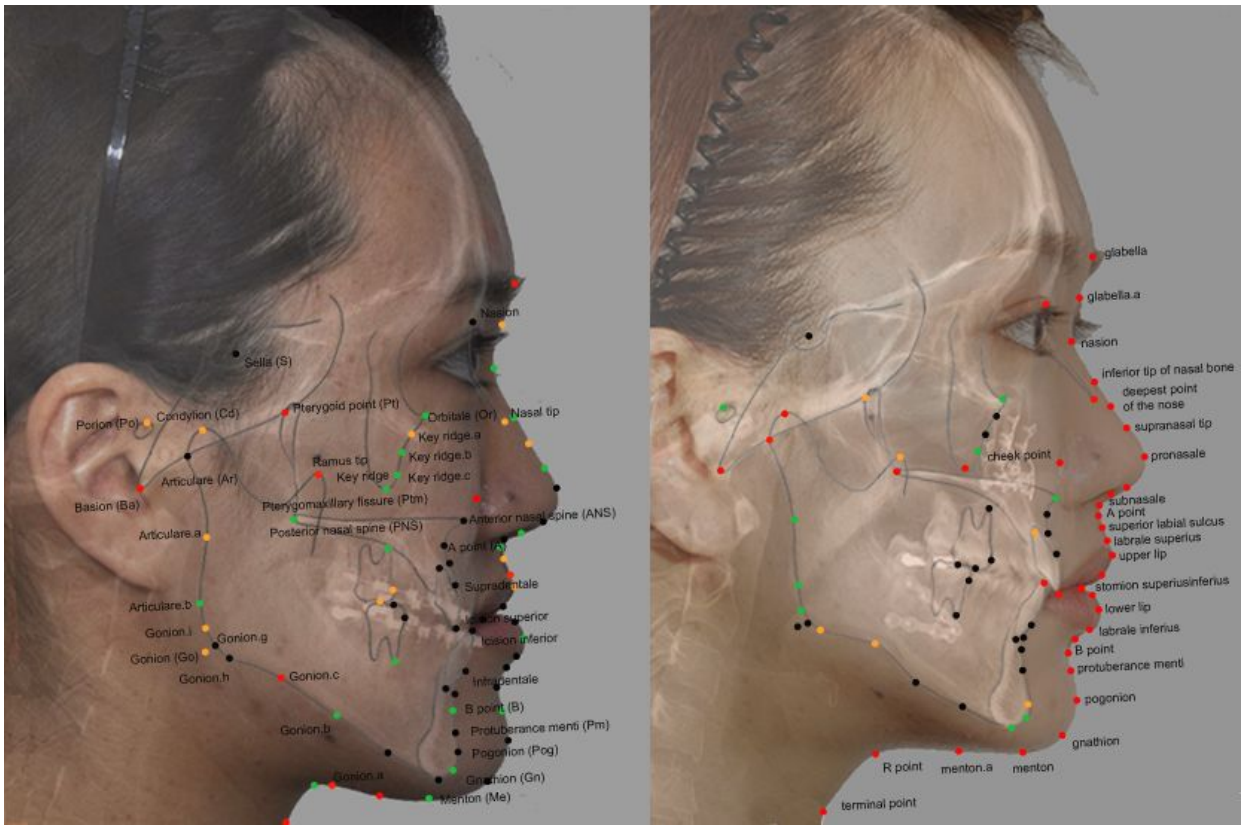

Figure 11. Diagram showing selected landmark coordinates in the present study. All cephalometric landmarks used in present study marked as black dots; selected x-coordinates indicated as orange dots; selected y-coordinates indicated as green dots; and selected landmarks (x- and y-coordinates) indicated as red dots. ***Left,*** image composed from preoperative radiograph, with hard tissue landmarks in capital letters. ***Right,*** soft tissue landmarks in lowercase letters shown on the follow-up cephalogram.

APPENDIX TABLE. List of notations, acronyms, terms used in this study

Notation	
N	the number of samples or observations
p	the number of predictor (x) variables
k	the number of response (y) variables
a	the number of latent factors used (≤ rank of **X**)
r	the number of levels in the random variable
$\boldsymbol{x}$	a column vector of features for the predictor variables (size $p \times 1$)
$\boldsymbol{y}$	a column vector of features for the response variables (size $k \times 1$)
X	a matrix of features for the predictor variables (size $n \times p$)
Y	a matrix of features for the response variables (size $n \times k$)
b	a column vector of coefficients
B	a matrix of coefficients for the multivariate methods (size $p \times k$)
$\mathbf{t}_h$	a column vector of scores for the **X** block, factor h (size $n \times 1$)
$\mathbf{p}_h^{\mathrm{T}}$	a row vector of loadings for the **X** block, factor h (size $1 \times p$)
$\mathbf{w}_h^{\mathrm{T}}$	a row vector of weights for the **X** block, factor h (size $1 \times p$)
T	the matrix of **X** scores (size $n \times a$)
$\mathbf{P}^{\mathrm{T}}$	the matrix of **X** loadings (size $a \times p$)
$\mathbf{u}_h$	a column vector of scores for the **Y** block, factor h (size $n \times 1$)
$\mathbf{q}_h^{\mathrm{T}}$	a row vector of loadings for the **Y** block , factor h (size $1 \times k$)
U	the matrix of scores (size $n \times a$)
$\mathbf{Q}^{\mathrm{T}}$	the matrix of **Y** loadings (size $a \times k$)
$\mathbf{M}_h$	a rank 1 matrix, outer product of $\mathbf{t}_h$ and $\mathbf{p}_h^{\mathrm{T}}$ (size $n \times p$)
$\mathbf{E}_h$	the residual of **X** after subtraction of h components (size $n \times p$)
$\mathbf{F}_h$	the residual of **Y** after subtraction of h components (size $n \times k$)
$\mathbf{b}_h$	the regression coefficient for one PLS component
$\mathbf{I}_n$	the identity matrix of size $n \times n$
Acronyms	
OLS	ordinary least squares regression, multiple multivariate linear regression
PCR	principal components regression
PLS	partial least squares regression
PLSM	modified PLS method implementing mixed effect model
CV	cross-validation
Terms	
Predictor variable	also called input-, explanatory-, descriptive-, regressor-, or independent variables
Response variable	also called output-, regressand-, or dependent variables
Training dataset	also called learning-, calibration-, or study dataset
Test dataset	also called real-, prediction-, or validation dataset
Component	also called latent variable or latent factor
Coefficient	also called loading or sensitivity
Score	also called model matrix