

UNIVERSITÉ DE MONTRÉAL

MODÉLISATION DE LA COURBURE SCOLIOTIQUE À PARTIR DE LA  
TOPOGRAPHIE DU TRONC

CHARLES BERGERON  
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(MATHÉMATIQUES APPLIQUÉES)

AVRIL 2005



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file    Votre référence*

*ISBN: 0-494-01282-X*

*Our file    Notre référence*

*ISBN: 0-494-01282-X*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.



UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

MODÉLISATION DE LA COURBURE SCOLIOTIQUE À PARTIR DE LA  
TOPOGRAPHIE DU TRONC

présenté par: BERGERON Charles

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de:

M. DESAULNIERS Guy, Ph.D., président

Mme CHERIET Farida, Ph.D., membre et directrice de recherche

M. LABELLE Hubert, M.D., membre et codirecteur de recherche

M. DANSEREAU Jean, Ph.D., membre

à tous ceux qui continuent de croire en moi.....

## REMERCIEMENTS

Tout d'abord, j'aimerais remercier l'équipe du Laboratoire d'imagerie et vision 4D à l'École Polytechnique de Montréal, dont les membres suivants : Michel Archambault, Jonathan Boisvert, Élodie Dusch, Jean-François Gauthier, Alexis Hétu, Thomas Hurtut, Quin Hu Liao, Mathias Mondamert, Charles-Auguste Saint-Pierre, Nancy Shawafaty, Ligen Wang, Luke Windisch, Rui Zhang, et surtout, l'associée de recherche Valérie Pazos. Des remerciements sont aussi transmis à l'équipe de support informatique de Luc Lalonde. Je désire aussi fortement souligner la contribution de ma directrice de recherche, le Prof. Farida Cheriet pendant ces 28 derniers mois inoubliables. Votre dévouement fut total, et je vous en remercie chaleureusement.

Ensuite, j'aimerais remercier l'équipe du Laboratoire d'informatique en scoliose 3D du Centre de recherche de l'Hôpital Sainte-Justine de Montréal qui apportait souvent d'excellentes suggestions dont : Kajsa Duke, Luc Duong, Anne-Marie Huynh, Julie Joncas, Philippe Labelle, Justin Novosad et Jean-Marc Mac-Thiong. Cette liste doit de plus inclure mon codirecteur, le Prof. Hubert Labelle de l'Université de Montréal, dont le soutien fut sans réserve. Malgré ses journées toujours très chargées, j'arrivais toujours à mettre ma main dessus lorsque nécessaire.

Des remerciements sont également transmis à l'équipe du *McCaig Centre for Joint Injury and Arthritis Research* de l'Université de Calgary. En particulier, je mentionne Ingrid Fjeld, Jacob Jaremko, PhD, Philippe Poncet, PhD, Daniela Robu et Hongfa Wu qui m'ont chaleureusement reçu lors de mon séjour à Calgary en juin 2003. L'implication de mes coauteurs d'articles de revue et d'actes de conférence, le Prof. Janet Ronsky (qui est l'origine de l'idée d'exploiter une base de données de la colonne vertébrale indépendante dans cette recherche, lors de son passage à

Montréal en avril 2004) et le Prof. Ronald Zernicke (qui retournait ses commentaires sur les multiples versions de nos écrits avec un délai ne dépassant jamais 48 heures), fut extraordinaire. Merci infiniment !

Que ce soit lors de colloques informels ou d'une journée thématique à l'automne 2004, mes échanges avec le groupe du Laboratoire d'imagerie, vision et intelligence artificielle de l'École de technologie supérieure furent très positifs. Je remercie en particulier le Prof. Mohammed Cheriet pour ses questions à la fois perçantes et éclairées.

Un appui dans mes études est venu également du Prof. Guy Desaulniers de l'École Polytechnique de Montréal. Celui-ci a coordonné mes charges des cours Calcul multivariable et vectoriel et Analyse numérique avec une confiance sans réserve. Il a de plus agi comme président-rapporteur du jury d'évaluation de ce mémoire. Je remercie par ailleurs l'ensemble du jury pour sa lecture attentive et critique de ce document, et spécifiquement le Prof. Jean Dansereau pour sa participation.

Au cours de l'été 2003, j'ai eu la possibilité de participer à trois activités d'enrichissement, soient : le *Graduate Industrial Mathematical Modeling Camp* au *Banff International Research Station* en Alberta, organisé par le *Pacific Institute for the Mathematical Sciences* ; le *Industrial Problem Solving Workshop* à l'Université de Calgary, organisé par le *Institute for Mathematics and its Applications* ; et le *Applied Mathematics Summer School* à la *Memorial University of Newfoundland* à Saint-John's, organisé par le *Atlantic Association for Research in the Mathematical Sciences*. Je remercie ces organismes subventionnaires pour leur contribution à ma formation.

Le Conseil de recherche en sciences naturelles et en génie du Canada et les Instituts de recherche en santé du Canada ont appuyé financièrement cette recherche. Ma

reconnaissance est également soumise au *Fraternal Order of Eagles (Alberta & Saskatchewan)*, au *Wood Professorship in Joint Injury Research* et au *SPIE Student Travel Contingency Grant Programme* pour leur support financier à ma participation à la conférence *SPIE Medical Imaging* qui eut lieu en février 2005 à San Diego.

## RÉSUMÉ

Ce mémoire de maîtrise traite de la question du suivi non invasif de patients scoliotiques. Celui-ci propose un développement innovatif en ce qui concerne la modélisation des structures anatomiques (soient la surface du tronc et la courbe de la colonne vertébrale) ainsi que la recherche d'une relation entre ceux-ci par des approches d'intelligence artificielle.

Ce document présente un nouveau paradigme pour l'apprentissage de systèmes pour la régression lorsque (i) le nombre d'échantillons est petit et contient des interdépendances, et (ii) chaque échantillon est formé de vaste quantités de données discrètes qui représentent des structures fonctionnelles. L'application motivant ce travail est le problème de suivi scoliotique non invasif. En effet, les déformations de la colonne vertébrale scoliotique sont généralement évaluées à partir de radiographies à des intervalles réguliers, un protocole qui augmente l'incidence de développement de plusieurs cancers. Ce mémoire accepte l'hypothèse classique selon laquelle l'asymétrie dans la topographie du tronc est indicative de la déformation des structures internes. Donc, ce travail cherche à exploiter la géométrie du tronc à partir d'un processus d'acquisition de données non invasives basé sur un système de caméras laser. L'objectif est l'identification d'une relation robuste entre les valeurs d'entrée représentant le tronc et celles de sortie représentant la colonne vertébrale.

Dans cette étude, les acquisitions laser de la topographie du tronc et des reconstructions de données de la colonne vertébrale à partir de rayons X chez des patients scoliotiques sont représentées par des fonctions décrivant des surfaces et des courbes en utilisant l'approche de l'analyse fonctionnelle de données. En particulier, la transformée des fractions de bruit minimales fonctionnelles (fMNF), une analyse par composantes principales dans un espace métrique non euclidien, est exploitée afin

d'extraire des composantes fonctionnelles qui sont globales, insensibles au bruit, et représentatives de la géométrie. Cette méthode permet également de réduire la dimensionnalité du problème.

Comme méthode d'apprentissage, la régression par vecteurs de support (SVR) a été sélectionnée pour sa robustesse. La pénalisation de la complexité du modèle évite le surapprentissage malgré le petit nombre de données d'apprentissage. De plus, la validation croisée fait en sorte que tous les échantillons sont utilisés pour tester l'apprentissage, augmentant la confiance dans la généralisation du modèle à des échantillons inconnus. Cette étude constitue une première prédiction 3D robuste de la colonne vertébrale à partir de données acquises non invasivement. Ce système conjoint fMNF-SVR explique 81% de la variabilité dans la projection latérale et 72% de la variabilité dans la projection frontale de la colonne vertébrale. Ceux-ci permettent une prédiction de cette courbe avec une erreur- $L_2$  de moins de 5 millimètres dans chaque plan, comparable aux erreurs dans les données de départ.

Ces résultats appuient l'incorporation de la méthode fMNF-SVR dans le développement d'un nouveau protocole de suivi scoliotique non invasif, et offre de vastes possibilités de modélisation plus complexe au fur et à mesure que de nouvelles données sont acquises, évitant ainsi le surapprentissage.

## ABSTRACT

This masters thesis addresses the problem of noninvasive scoliosis follow-up. In doing so, innovative strides are made with respect to the modeling of anatomical structures, that is, the trunk surface and spinal curve, and the search for a relation between these using artificial intelligence approaches.

This document proposes a framework for the training of learning systems for regression when (i) the number of examples is small and contains interdependencies, and (ii) each sample consists of large quantities of discrete data that are functional in nature. The application motivating this study is the problem of noninvasive scoliosis follow-up. Indeed, scoliotic spinal deformities are generally assessed from full spinal X-rays at regular intervals, a clinical protocol which increases the incidence of cancer development. This thesis accepts the classic hypothesis whereby asymmetry in the trunk topography is indicative of internal structure deformity, and exploits noninvasive data acquisition using laser scanning technology. The objective is the achievement of robust relations between inputs describing the trunk shape and outputs describing geometry of the spinal column.

In this study, laser scans of the trunk surface and reconstructions of spinal data from X-rays from scoliosis patients were functionally represented as surfaces and curves, using the framework proposed by functional data analysis. In particular, the functional minimum noise fractions (fMNF) transform, a principal components analysis in a non-Euclidean metric space, is exploited to extract leading functional components that become global, geometry-comprehensive and noise-insensitive features. Leading coefficients thereof achieved sufficient dimensionality reduction for the prediction of spine from trunk.

As a learning method, support vector regression (SVR) was chosen for its strong

generalizability capability that stems from penalizing model complexity. The small size of the training set resulted in some overfitting, but 5-fold cross-validation ensured confidence in the generalization of testing results. This study reports a first robust 3D prediction of the 3D spinal curve from a non-invasive scan of the trunk. The joint fMNF-SVR system explained 81% of lateral and 72% of frontal spinal curve projection variability. Those translated to a spinal curve prediction  $L_2$ -error under 5 millimetres, comparable to measurement error in data.

These results support the use of the fMNF-SVR method as part of the development of a new clinical protocol for noninvasive scoliosis monitoring, and offers vast potential for expansion towards more complex models as further data is obtained, thereby avoiding overfitting.

## TABLE DES MATIÈRES

|                                                                |      |
|----------------------------------------------------------------|------|
| DÉDICACE . . . . .                                             | v    |
| REMERCIEMENTS . . . . .                                        | vi   |
| RÉSUMÉ . . . . .                                               | ix   |
| ABSTRACT . . . . .                                             | xi   |
| TABLE DES MATIÈRES . . . . .                                   | xiii |
| LISTE DES TABLEAUX . . . . .                                   | xvi  |
| LISTE DES FIGURES . . . . .                                    | xvii |
| LISTE DES ABBRÉVIATIONS . . . . .                              | xix  |
| INTRODUCTION . . . . .                                         | 1    |
| CHAPITRE 1    ÉTAT DE L'ART . . . . .                          | 10   |
| 1.1    Angle de Cobb . . . . .                                 | 10   |
| 1.2    Indices d'évaluation de l'asymétrie externe . . . . .   | 15   |
| 1.3    Approximation de fonction par moindres carrés . . . . . | 17   |
| 1.4    Composantes principales fonctionnelles . . . . .        | 23   |
| 1.5    Fractions fonctionnelles de bruit minimal . . . . .     | 25   |
| 1.6    Régression linéaire . . . . .                           | 26   |
| 1.7    Réseaux de neurones . . . . .                           | 29   |
| 1.8    Régression par vecteurs de support . . . . .            | 30   |
| 1.9    Positionnement du mémoire . . . . .                     | 41   |
| CHAPITRE 2    PRÉDICTION DE LA COURBE SCOLIOTIQUE . . . . .    | 42   |

|            |                                                                                                               |    |
|------------|---------------------------------------------------------------------------------------------------------------|----|
| 2.1        | Synthèse . . . . .                                                                                            | 42 |
| 2.2        | Prediction of anterior scoliotic spinal curve from trunk surface using<br>support vector regression . . . . . | 44 |
| 2.2.1      | Introduction . . . . .                                                                                        | 44 |
| 2.2.2      | Principal components analysis . . . . .                                                                       | 47 |
| 2.2.3      | Support vector regression . . . . .                                                                           | 49 |
| 2.2.4      | Spine data acquisition and registration . . . . .                                                             | 53 |
| 2.2.5      | Trunk data acquisition and registration . . . . .                                                             | 54 |
| 2.2.6      | Spinal functional data analysis . . . . .                                                                     | 58 |
| 2.2.7      | Trunk functional data analysis . . . . .                                                                      | 60 |
| 2.2.8      | Results . . . . .                                                                                             | 62 |
| 2.2.9      | Discussion . . . . .                                                                                          | 64 |
| 2.2.10     | Conclusion . . . . .                                                                                          | 67 |
| CHAPITRE 3 | PRÉDICTION TRIDIMENSIONNELLE AMÉLIORÉE . .                                                                    | 69 |
| 3.1        | Synthèse . . . . .                                                                                            | 69 |
| 3.2        | Noninvasive assessment of scoliotic spinal curve using robust learning<br>techniques . . . . .                | 72 |
| 3.2.1      | Introduction . . . . .                                                                                        | 72 |
| 3.2.2      | Methods . . . . .                                                                                             | 75 |
| 3.2.2.1    | Minimum noise fractions . . . . .                                                                             | 75 |
| 3.2.2.2    | Support vector regression . . . . .                                                                           | 77 |
| 3.2.2.3    | Functional model assessment . . . . .                                                                         | 80 |
| 3.2.3      | Materials . . . . .                                                                                           | 81 |
| 3.2.3.1    | Data acquisition . . . . .                                                                                    | 83 |
| 3.2.3.2    | Data registration . . . . .                                                                                   | 86 |
| 3.2.3.3    | Functional data analysis: spinal curve . . . . .                                                              | 87 |
| 3.2.3.4    | Functional data analysis: trunk surface . . . . .                                                             | 89 |

|                                             |                                              |     |
|---------------------------------------------|----------------------------------------------|-----|
| 3.2.3.5                                     | Support vector regression learning . . . . . | 94  |
| 3.2.4                                       | Results . . . . .                            | 95  |
| 3.2.5                                       | Discussion . . . . .                         | 96  |
| 3.2.6                                       | Conclusion . . . . .                         | 101 |
| 3.3                                         | Addendum . . . . .                           | 102 |
| 3.3.1                                       | Implémentation SVR . . . . .                 | 103 |
| 3.3.2                                       | Résultats . . . . .                          | 104 |
| DISCUSSION GÉNÉRALE ET CONCLUSION . . . . . |                                              | 108 |
| RÉFÉRENCES . . . . .                        |                                              | 115 |

## LISTE DES TABLEAUX

|             |                                                                                                 |     |
|-------------|-------------------------------------------------------------------------------------------------|-----|
| Tableau 1.1 | Recueil de fonctions noyaux définis positifs . . . . .                                          | 37  |
| Table 2.1   | Statistics for selected clinical and spinal indices . . . . .                                   | 57  |
| Table 2.2   | Properties of spinal curve principal components . . . . .                                       | 60  |
| Table 2.3   | Properties of trunk surface principal components . . . . .                                      | 62  |
| Table 2.4   | SVR prediction of first spinal coefficient $\beta_1$ . . . . .                                  | 63  |
| Table 2.5   | SVR prediction of second spinal coefficient $\beta_2$ . . . . .                                 | 63  |
| Table 2.6   | SVR prediction of third spinal coefficient $\beta_3$ . . . . .                                  | 64  |
| Table 2.7   | Prediction of spinal curve using support vector regression . . . . .                            | 64  |
| Tableau 3.1 | Comparaison des deux articles . . . . .                                                         | 70  |
| Table 3.2   | Kernel functions . . . . .                                                                      | 79  |
| Table 3.3   | Statistics for selected clinical and spinal indices . . . . .                                   | 85  |
| Table 3.4   | Properties of spinal basis functions. . . . .                                                   | 89  |
| Table 3.5   | Support vector regression prediction of spinal minimum noise<br>fraction coefficients . . . . . | 96  |
| Tableau 3.6 | Prédiction SVR linéaire de coefficients fMNF rachidiens . . . . .                               | 104 |
| Tableau 3.7 | Prédiction SVR spline cubique de coefficients fMNF rachidiens                                   | 104 |
| Tableau 3.8 | Prédiction SVR de la courbe rachidienne (apprentissage et<br>test) . . . . .                    | 105 |

## LISTE DES FIGURES

|            |                                                                                     |    |
|------------|-------------------------------------------------------------------------------------|----|
| Figure 0.1 | Colonne vertébrale saine . . . . .                                                  | 2  |
| Figure 0.2 | Radiographie postéroantérieure d'un rachis scoliotique . . . .                      | 3  |
| Figure 0.3 | Problématique radiographique . . . . .                                              | 4  |
| Figure 0.4 | Forme asymétrique du dos d'un patient scoliotique . . . . .                         | 6  |
| Figure 0.5 | Approche de Jaremko . . . . .                                                       | 7  |
| Figure 0.6 | Approche de Wang . . . . .                                                          | 8  |
| Figure 1.1 | Définition clinique de l'angle de Cobb . . . . .                                    | 12 |
| Figure 1.2 | Mesure informatique de l'angle de Cobb . . . . .                                    | 14 |
| Figure 1.3 | Données de la courbe rachidienne latérale pour quatre patients                      | 17 |
| Figure 1.4 | Contour horizontal d'un tronc . . . . .                                             | 18 |
| Figure 1.5 | Erreurs empiriques d'une régression linéaire . . . . .                              | 27 |
| Figure 1.6 | Graphes de coefficients externes par rapport aux coefficients<br>internes . . . . . | 31 |
| Figure 1.7 | Fonctions de pénalité . . . . .                                                     | 33 |
| Figure 1.8 | Impact du paramètre d'insensibilité . . . . .                                       | 34 |
| Figure 1.9 | Quatre régressions non linéaires de données . . . . .                               | 38 |
| Figure 2.1 | Méthodologie pour le premier article . . . . .                                      | 43 |
| Figure 2.2 | Study objective flow chart . . . . .                                                | 48 |
| Figure 2.3 | Definition of SVR insensitivity parameter $\epsilon$ . . . . .                      | 51 |
| Figure 2.4 | Spine data acquisition and processing . . . . .                                     | 54 |
| Figure 2.5 | Data acquisition system . . . . .                                                   | 55 |
| Figure 2.6 | Trunk data acquisition and processing . . . . .                                     | 56 |
| Figure 2.7 | Spinal curve functional basis . . . . .                                             | 61 |
| Figure 2.8 | System flow chart . . . . .                                                         | 63 |
| Figure 2.9 | Comparison of spinal data with SVR-predicted spinal curves                          | 65 |
| Figure 3.1 | Méthodologie pour le second article . . . . .                                       | 71 |

|             |                                                                                                                 |     |
|-------------|-----------------------------------------------------------------------------------------------------------------|-----|
| Figure 3.2  | Learning system flow chart . . . . .                                                                            | 82  |
| Figure 3.3  | Data acquisition system . . . . .                                                                               | 84  |
| Figure 3.4  | Spinal functional bases . . . . .                                                                               | 90  |
| Figure 3.5  | Retained trunk sections . . . . .                                                                               | 91  |
| Figure 3.6  | Decreasing sequences of latent values . . . . .                                                                 | 93  |
| Figure 3.7  | SVR input dimensionality optimization . . . . .                                                                 | 95  |
| Figure 3.8  | Comparison of support vector regression (SVR) and maximum-likelihood (ML) spinal curve predictions . . . . .    | 97  |
| Figure 3.9  | Comparison of X-ray-reconstructed and support vector regression (SVR)-predicted spinal curves . . . . .         | 98  |
| Figure 3.10 | Distributions du coefficient de détermination fonctionnel . . . . .                                             | 106 |
| Figure 3.11 | Comparaison entre les reconstructions radiographiques et les prédictions SVR de la colonne vertébrale . . . . . | 107 |

**LISTE DES ABBRÉVIATIONS**

|       |                                                                     |
|-------|---------------------------------------------------------------------|
| 3D    | tridimensionnel, en trois dimensions                                |
| AI    | intelligence artificielle                                           |
| ANN   | réseau de neurones artificiel                                       |
| BSR   | rotation dorsale                                                    |
| CV    | validation croisée                                                  |
| DCT   | transformée cosinus discrète                                        |
| DLT   | direct linear transformation                                        |
| DWT   | transformée ondelettes discrète                                     |
| FDA   | analyse fonctionnelle de données                                    |
| fMNF  | fractions fonctionnelles de bruit minimal                           |
| FMS   | sinusoïde à fréquence modulée                                       |
| fPCA  | analyse par composantes principales fonctionnelles                  |
| FSR   | représentation par série de Fourier                                 |
| GA    | algorithme génétique                                                |
| LAT   | radiographie latérale                                               |
| ML    | vraisemblance maximale                                              |
| MLR   | multiple linear regression                                          |
| MNF   | fractions de bruit minimal                                          |
| MRI   | imagerie par résonance magnétique                                   |
| PA    | radiographie postéroantérieure                                      |
| PA-0  | radiographie postéroantérieure avec le tube horizontal au sol       |
| PA-20 | radiographie postéroantérieure avec le tube plavé à un angle de 20° |
| PCA   | analyse par composantes principales                                 |
| PCHIP | polynômes cubiques d'interpolation par morceaux de Hermite          |
| RBF   | fonction à base radiale                                             |

|     |                                        |
|-----|----------------------------------------|
| SD  | écart-type                             |
| SLR | simple linear regression               |
| SNR | ratio signal-à-bruit                   |
| SVC | classification par vecteurs de support |
| SVR | régression par vecteurs de support     |

## INTRODUCTION

La projection d'une colonne vertébrale saine est généralement droite dans le plan frontal et présente des courbures naturelles appelées cyphose et lordose dans le plan latéral (figure 0.1). La scoliose est une déformation complexe et tridimensionnelle (3D) de la colonne vertébrale, impliquant la rotation axiale des vertèbres et des courbures dans les plans frontal et latéral. Causant des problèmes de santé primaires et secondaires, cette condition affecte jusqu'à 4% de la population. La cause de la scoliose idiopatique est inconnue; elle se manifeste à jeune âge ( $< 3$  ans,  $< 1\%$  des cas), pendant l'enfance 3–10 ans, 12–21% des cas) ou pendant l'adolescence ( $> 10$  ans) (Reamy et Slakey, 2001). Un plus petit nombre de cas sont de source génétique, appelée scoliose congénitale.

Les facteurs de progression déterminants de cette maladie sont le sexe (les femmes présentent un risque de progression de la courbure vertébrale 10 fois plus fort que les hommes), le potentiel de croissance, l'apex de la courbe majeure, et l'ordre de grandeur de la courbure (Reamy et Slakey, 2001; Greiner, 2002). Ces deux derniers sont évalués à partir d'une radiographie complète de la colonne vertébrale du patient en position debout. La figure 0.2 illustre une radiographie postéroantérieure (PA, de l'arrière du patient) sur laquelle la déformation scoliotique est clairement visible.

Étant donné la variabilité de la croissance adolescente sur une base annuelle (Tuddenham et Snider, 1954) et même quotidienne (Thalange et al., 1996), un suivi clinique fréquent est nécessaire. Toutefois, le diagnostic et le suivi de la scoliose demeurent controversés. Par exemple, Reamy et Slakey (2001) recommandent l'acquisition radiographique aux 6 mois jusqu'à maturité osseuse. En contrepartie, Greiner (2002) considère seulement l'acquisition de rayons X pour les patients présentant

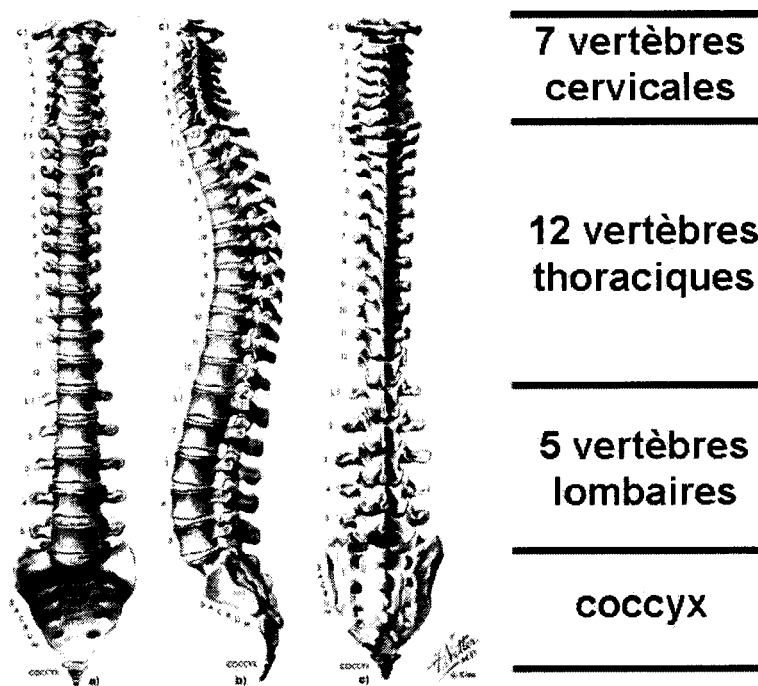


Figure 0.1 La projection d'une colonne vertébrale saine est généralement droite dans le plan frontal et présente des courbures naturelles appelées cyphose et lordose dans le plan latéral.

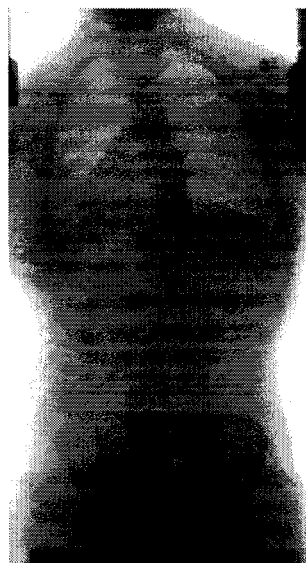


Figure 0.2 Sur cette radiographie postéroantérieure, une déformation scoliotique de la colonne vertébrale est visible.

une courbure suffisamment sévère pour nécessiter un traitement ou pour lesquels une telle progression est envisagée. De part et d'autre, le facteur limitant est le risque associé à l'exposition répétée aux rayons X nocifs dommageables à l'ADN (figure 0.3).

L'équipement moderne a permis de réduire la dose de radiations, et les vues PA ont réduit l'impact sur les tissus sensibles. L'incidence de cancer du sein, de la thyroïde, du poumon, de l'ovaire et de la moelle épinière est augmentée pour atteindre 2,4 pour mille patients—comparable à la probabilité de développement d'un cancer de la thyroïde dans la population générale. Étant donné que moins d'une scoliose sur dix exige un traitement au-delà du suivi périodique (Reamy et Slakey, 2001), il est désirable de minimiser la cadence de charge aux rayons X autant que possible. Un nouveau protocole pour le suivi non invasif de la scoliose est donc nécessaire.

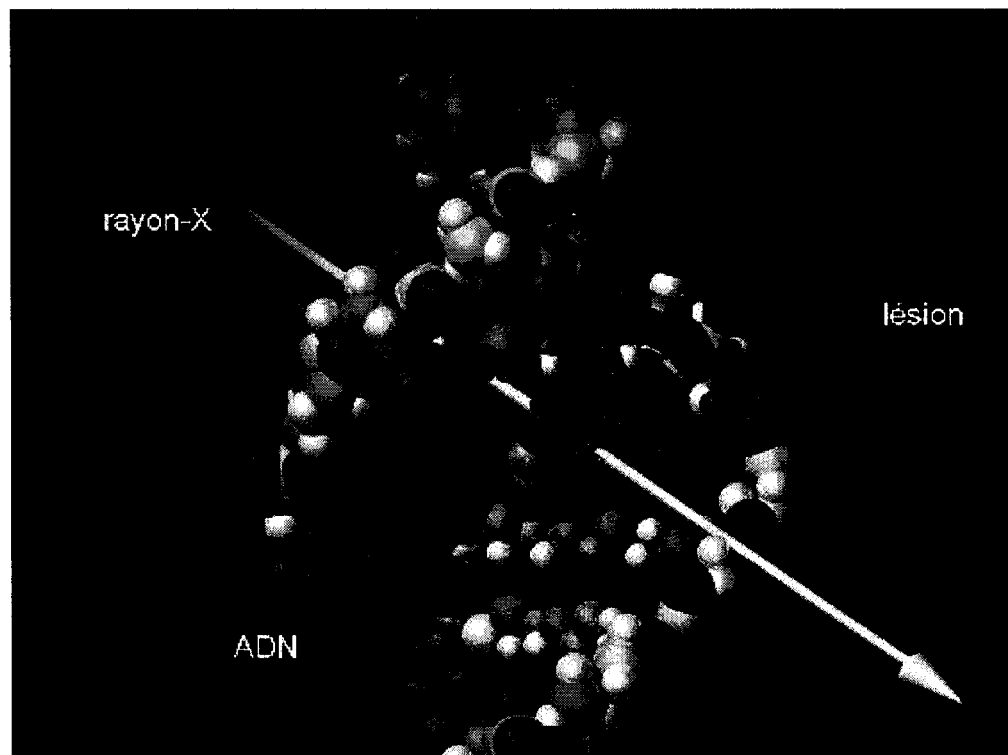


Figure 0.3 Les rayons X causent les lésions à l'ADN, doublant l'incidence de plusieurs cancers chez les patients scoliotiques. Figure modifiée de *www.nasa.gov*.

La technologie d'imagerie par résonance magnétique (en anglais, *magnetic resonance imaging*, MRI) constitue une possibilité de rechange. Toutefois, son utilisation est limitée aux cas les plus sévères en vertu de son coût élevé et sa faible disponibilité (Carro, 1999). De plus, MRI est moins efficace que les rayons X pour distinguer les os des autres tissus. Difficulté additionnelle, MRI est acquise en position couchée, dans laquelle les courbures scoliotiques sont réduites. Finalement, MRI ne peut être utilisée lorsque la colonne vertébrale est instrumentée suite à une intervention chirurgicale.

Une autre possibilité combine donc des technologies d'acquisition de la surface du tronc et un modèle mathématique adéquat pour relier l'asymétrie de la topographie du tronc externe (*l'externe*) à la déformation de la courbe de la colonne vertébrale (*l'interne*). Dans ce mémoire, c'est cette approche de solution qui est adoptée. Cette stratégie est motivée par trois faits :

- les changements dans la forme du tronc constituent souvent le premier indicateur de la scoliose, indiquant que cette forme est révélatrice de la dynamique scoliotique ;
- l'asymétrie externe demeure l'effet le plus visible de cette maladie (figure 0.4) ;
- et
- l'acquisition de données du tronc est un processus totalement non invasif.

Évidemment, cette stratégie présuppose que les liens physiologiques entre l'externe et l'interne sont suffisamment réguliers d'un patient à l'autre pour permettre qu'une relation entre ceux-ci puisse être identifiée, une hypothèse qui n'est pas certaine et qui peut justifier les erreurs commises dans la modélisation. Trois aspects fondamentaux concernant cette approche de suivi de patients scoliotiques sont :

- la précision des systèmes d'acquisition de données ;
- la représentation de la courbe de la colonne vertébrale (*la courbe rachidienne*) et de la surface du tronc ; et

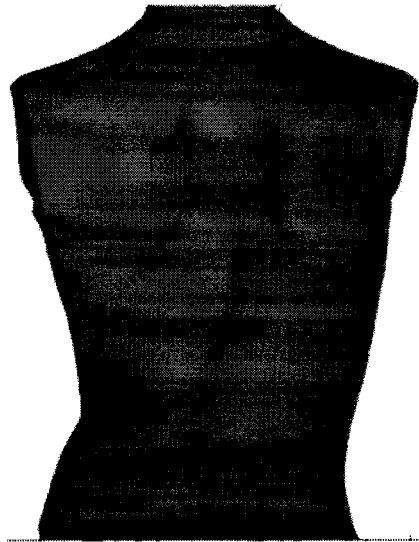


Figure 0.4 L'asymétrie dans la forme du dos demeure l'effet le plus visible de la scoliose. Figure générée à partir des données de l'Hôpital Sainte-Justine.

– la méthode d'identification de relation entre ces derniers.

Concernant ce premier aspect, les systèmes d'acquisition de la surface du tronc sont certes devenus plus précis au cours des trente dernières années, permettant en théorie une meilleure modélisation externe-interne, quoique les résultats demeurent trop faibles et trop peu robustes pour permettre l'innovation du suivi scoliotique dans un contexte clinique. Le présent mémoire n'est pas axé sur cet aspect, et aucun effort n'a été mis sur l'amélioration de la qualité des données obtenues utilisant la méthodologie développée à l'Université de Calgary (Poncet, 2001).

Par contre, les deux aspects restants, soient la représentation de la courbe rachidienne et de la topographie du tronc et la méthode d'identification de relation entre ces derniers, sont traités dans le présent mémoire. De façon générale, les approches proposées sont indépendantes des technologies d'acquisition de données.

L'approche classique de prédiction externe-interne a tenté, en premier lieu, de

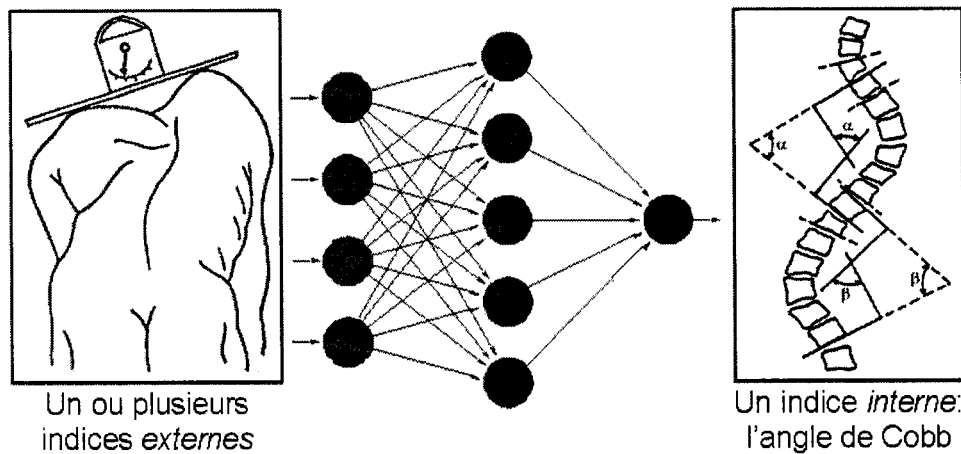


Figure 0.5 Jaremko et al. (2002a) ont tenté de prédire l'angle de Cobb à partir d'indices externes optimisés par l'algorithme génétique à l'aide d'un réseau de neurones artificiel.

définir des indices de déformation externes et internes, et en second lieu, de relier ceux-ci par la régression linéaire simple ou multiple (en anglais, *simple or multiple linear regression*, SLR ou MLR). Jaremko et al. (2002a) ont innové cette approche en proposant un grand nombre de nouveaux indices d'asymétrie externe, utilisant un algorithme génétique (en anglais, *genetic algorithm*, GA) pour sélectionner les meilleurs parmi ceux-ci, et recherchant la relation non linéaire avec l'angle de Cobb, une mesure clinique traditionnelle de la déformation de la colonne vertébrale scoliotique, à l'aide d'un réseau de neurones artificiel (en anglais, *artificial neural network*, ANN). La figure 0.5 illustre cette approche. La vaste gamme d'indices externes exploitées permet d'améliorer la capacité de prédiction, sans pour autant éviter le surapprentissage. En effet, leur coefficient de détermination pour la prédiction de l'angle de Cobb est de 0,94 en apprentissage et 0,87 en test.

Wang (2003) a, quant à lui, tenté d'associer les composantes principales des points de contrôle d'une représentation par B-splines de la surface du tronc à l'une

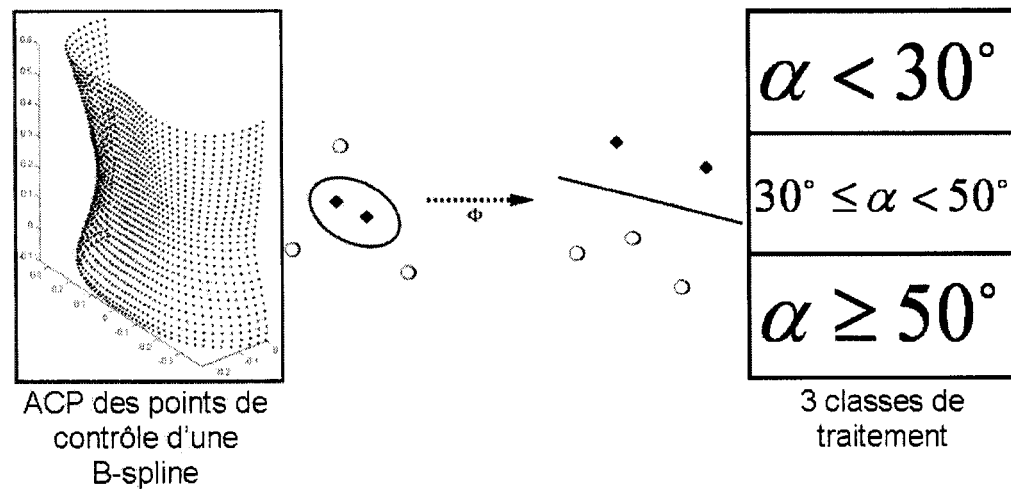


Figure 0.6 Wang (2003) a tenté de prédire l'une de trois classes de traitement définies par l'angle de Cobb à partir des composantes principales des points de contrôle d'une représentation B-spline de la surface du tronc à l'aide de la classification par vecteurs de support. Tirée en partie de (Schölkopf et Smola, 2002).

de trois classes de traitement définies par l'angle de Cobb. La figure 0.6 illustre cette approche. Malgré le fait que la classification par vecteurs de support (en anglais, *support vector classification*, SVC) ait été choisie comme méthode d'apprentissage, un algorithme conçu généralisable, la prédiction identifiée est très peu robuste, soit une capacité de classification de 89% en apprentissage et 66% en test. Cette faible performance est attribuée à la perte d'information critique sur la forme de l'externe lors de l'analyse par composantes principales (en anglais, *principal components analysis*, PCA) des points de contrôle.

Ce mémoire reprend des éléments de ces deux approches, soient la représentation de l'externe et de l'interne de façon plus détaillée et complète ainsi que l'utilisation de méthodes d'apprentissage. Ce faisant, le choix de techniques qui sont moins sensibles au bruit d'acquisition et robustes en ce qui concerne la prédiction sont recherchées. Ce faisant, ce mémoire adopte une méthodologie mathématique

très éloignée des définitions cliniques actuelles des déformations scoliotiques. C'est précisément cette façon différente de faire qui constitue l'originalité du travail ici présenté. Toutefois, il est entendu qu'un travail éventuel de transfert des connaissances à la discipline médicale sera nécessaire afin que celui-ci soit pleinement valorisé.

La structure de ce mémoire est comme suit. Le chapitre 1 présente une revue sur les méthodes de solution présentes et passées de ce problème de suivi scoliotique, permettant de faire le point sur l'état des connaissances dans ce domaine. Par la suite, le chapitre 2 contient un premier article présentant une méthodologie de prédiction de la vue frontale de la courbe de la colonne vertébrale à partir de la topographie du tronc. Cette approche est basée sur la représentation fonctionnelle de données et une technique d'apprentissage robuste. Le chapitre 3, quant à lui, développe et généralise cette façon de faire, permettant une prédiction tridimensionnelle de la courbe rachidienne scoliotique. Finalement, le dernier chapitre présente une discussion générale et une conclusion.

## CHAPITRE 1

### ÉTAT DE L'ART

Cet état de l'art présente une revue des connaissances et des méthodes sur deux grands thèmes. Tout d'abord, il y a la question de la représentation des données de la topographie du tronc et de la courbe de la colonne vertébrale. L'approche traditionnelle consiste à extraire directement des mesures qui sont des angles et des distances. Ces indices interne et externes sont brièvement présentés dans les deux premières sections. Toutefois, dans le contexte d'un calcul informatique d'un indice qui s'appelle l'angle de Cobb, la courbe scoliotique est représentée par une série trigonométrique de Fourier, et toute cette question d'analyse fonctionnelle de données est alors traitée dans les deux sections subséquentes.

Ensuite, la méthode utilisée pour découvrir la relation externe-interne doit être étudiée. L'approche classique consiste à relier les indices externes et interne par SLR ou MLR, mais d'autres techniques furent récemment tentées, telles que les ANN. Ces algorithmes, ainsi que la régression par vecteurs de support (en anglais, *support vector regression*, SVR), une méthode d'apprentissage relativement récente reconnue pour sa robustesse, font l'objet de rappel dans la deuxième moitié de ce chapitre.

#### 1.1 Angle de Cobb

La mesure clinique traditionnelle de la sévérité de la courbure de la colonne vertébrale est l'angle de Cobb (Cobb, 1948) (figure 1.1). Cet indice est obtenu directement à partir d'une radiographie PA utilisant un compas. Une courbe rachidienne présente

généralement plusieurs intervalles inflexionnels, et à chacun correspond un angle de Cobb. Sa mesure dépend d'un nombre de facteurs subjectifs, tels l'identification des vertèbres inflexionnelles et le traçage de droites tangentes et perpendiculaires directement sur la radiographie (Chockalingam et al., 2002), introduisant des variabilités de mesure intra- et interexamineur. Des valeurs pour l'écart-type de l'erreur intraexamineur sur l'angle de Cobb varient entre  $2,8^\circ$  et  $3,5^\circ$  (Carman et al., 1990; Goldberg et al., 1988; Morrissey et al., 1990; Ylikosky et Tallroth, 1990). L'erreur interexamineur est plus élevée, avec des valeurs de l'écart-type variant entre  $2,8^\circ$  et  $7,2^\circ$  (Goldberg et al., 1988; Morrissey et al., 1990; Ylikosky et Tallroth, 1990). Ces erreurs sont grandes dans le contexte où de nombreux orthopédistes considèrent une augmentation de l'angle de Cobb de  $5^\circ$  entre deux visites cliniques comme indicative d'une progression scoliotique significative, quoique Risser (1958) proposa qu'une différence de  $10^\circ$  soit exigée pour constituer une progression clinique significative. De part et d'autre, une augmentation de la fréquence de suivi exigerait une augmentation de la précision de mesure de l'angle de Cobb. Cette erreur constitue donc une première problématique d'utilisation de l'angle de Cobb ou d'autres indices internes semblables dans cette recherche.

Une seconde problématique de l'angle de Cobb se situe au niveau de sa capacité d'exprimer complètement la forme de la courbe rachidienne. Tout d'abord, cet angle ne décrit donc que la courbure d'une section de la colonne vertébrale, section qui est variable d'un patient à un autre. Encore faut-il que l'apex de la courbure soit spécifié! Donc, à une colonne vertébrale scoliotique correspond plusieurs angles de Cobb, et c'est l'angle maximal qui est généralement considéré. Ensuite, une aggravation de la courbe rachidienne 3D n'entraîne pas nécessairement une augmentation de l'angle de Cobb, puisque ce dernier ne considère que la projection de la colonne dans le plan frontal. Par conséquent, à un angle de Cobb maximal donné correspond une grande diversité de courbes rachidiennes. L'angle de Cobb ne peut

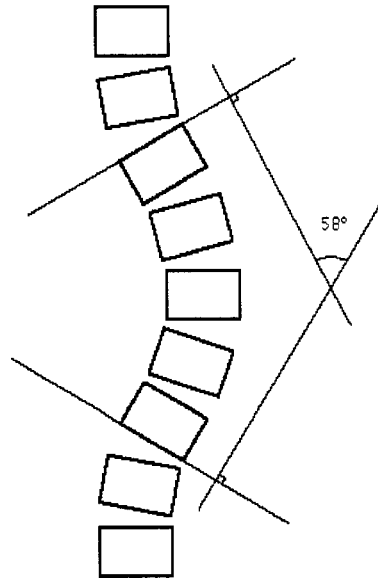


Figure 1.1 Définition clinique de l'angle de Cobb. L'angle de Cobb constitue une mesure clinique traditionnelle de la sévérité de la courbure rachidienne des patients scoliotiques. Figure tirée de (Jaremko, 2001)

caractériser complètement la forme de la colonne vertébrale 3D, tel qu'aucun scalaire ne le pourrait. Il est clair que cette approche d'évaluation de la déformation scoliotique interne par un indice, soit l'angle de Cobb, est à la fois incomplète et sensible à l'erreur de mesure.

Face à ses résultats, des méthodes informatiques pour rendre plus déterministe la mesure de l'angle de Cobb ont été tentées.

Une première approche consiste à identifier les vertèbres inflexionnelles sur des radiographies numérisées apparaissant sur un écran ; le reste du calcul de l'angle de Cobb est automatisé. Cette méthodologie permet de réduire l'erreur moyenne intraexamineur à  $0,739^\circ$  et interexamineur à  $1,22^\circ$  (Chockalingam et al., 2002).

Une seconde approche consiste à acquérir deux radiographies biplanaires, d'identifier un certain nombre de repères anatomiques de la colonne vertébrale, et d'exploiter une méthode de triangulation pour obtenir une reconstruction 3D : la transformation linéaire directe (en anglais, *direct linear transformation*, DLT). La méthode du DLT est une technique implicite largement utilisée (Marzan, 1976; Pearcy, 1985; Stokes et al., 1987; Dansereau et Stokes, 1987; Dansereau et al., 1990; Plamondon et Gagnon, 1990). La matrice de projection est obtenue sous forme de 11 paramètres, et l'objet est alors reconstruit par un algorithme linéaire basée sur les moindres carrés qui calcule les coordonnées 3D des points numérisés sur les deux radiographies (Poncet, 2001). Howell et Dickson (1989) ont utilisé quatre repères anatomiques par vertèbre, et Stokes et al. (1987) ont ajouté deux repères supplémentaires par vertèbre. Cette approche fut raffinée par Labelle et al. (1995) et Matsumoto et al. (1997). La courbe rachidienne est ensuite représentée par une série de Fourier trigonométrique de degré 3 (Dansereau et al., 1993; Labelle et al., 1995) et l'angle de Cobb est déterminé à partir des tangentes aux points d'inflexion (figure 1.2). Toutefois, le résultat dépend fortement de l'ordre de la série choisie, puisqu'une augmentation dans la précision du modèle entraîne également une augmentation du nombre de points d'inflexion, et, de ce fait, du nombre d'angles de Cobb. Cette problématique est largement indépendante de la classe de fonctions d'approximation choisie. Cette difficulté à précisément quantifier la déformation scoliotique de la colonne vertébrale par des indices internes suggère une approche alternative de représentation. Toutefois, l'idée de représenter la courbe rachidienne par une fonction est intéressante, puisque ses coefficients expriment complètement sa forme. Cette façon de faire est détaillée à la section 1.3.

Une dernière façon de représenter la courbe rachidienne doit être rappelée. En effet, Drerup et Hierholzer (1996) ont proposé l'utilisation d'une sinusoïde à fréquence modulée (en anglais, *frequency-modulated sinusoid*, FMS) comportant quatre pa-

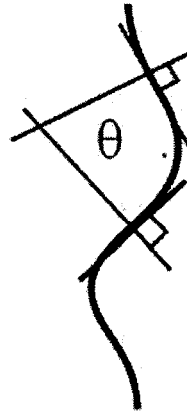


Figure 1.2 Mesure informatique de l'angle de Cobb. Une représentation par série de Fourier permet d'identifier les points d'inflexion de la courbe rachidienne. L'angle de Cobb est alors celui d'intersection des tangentes à ces points. Figure tirée de (Jaremko, 2001).

ramètres, à partir de laquelle des indices de courbure scoliotique tel l'angle de Cobb peuvent être calculés. Ce modèle, comparable en précision à une représentation par série trigonométrique de trois harmoniques, souffre de difficulté de calcul des paramètres : en effet, l'optimisation de ceux-ci exige l'exécution d'une procédure de Monte-Carlo. La fonction résultante présente des points d'inflexion différents de ceux obtenus d'une série trigonométrique, faisant ressortir la difficulté de précisément évaluer l'angle de Cobb.

Cette difficulté de choisir une méthodologie informatique de calcul de l'angle de Cobb, de précisément évaluer cet angle, ainsi que son manque de représentativité de la forme de la courbe de la colonne vertébrale font en sorte que des nouvelles mesures de la sévérité des déformations scoliotiques internes sont nécessaires.

## 1.2 Indices d'évaluation de l'asymétrie externe

Des indices de la forme du tronc peuvent être mesurés directement sur le patient, ou bien à partir d'imagerie 3D de la surface. Ces méthodes d'imagerie sont complètement non invasives et relativement peu dispendieuses, et les données permettent de décrire l'asymétrie de la surface avec beaucoup de détail (Poncet, 2001). Plusieurs technologies furent tentées depuis les trente dernières années, celles-ci sont énumérées ici :

- la topographie de Moiré (Adair et al., 2001; Moreland et al., 1981; Laulund et al., 1982; Suzuki et al., 1992), très sensible à de petits changements de posture du patient (Moreland et al., 1981) et exigeant un effort intensif pour la numérisation de la topographie de surface (Batouche et al., 1996) ;
- la rasterstéréographie (Frobin et Hierholzer, 1983; Stokes et al., 1987; Tredwell et Bannon, 1988; Turner-Smith et al., 1988; Upadhyay et al., 1988; Weisz et al., 1988; Vandegriend et al., 1995; Theologis et al., 1997; Thometz et al., 2000), toujours utilisée par certains groupes de recherche sous la forme du système *Quantec* (Goldberg et al., 2001; Liu et al., 2001) ;
- la vidéo numérique (Slupsky et al., 1992) ;
- les caméras laser, une solution retenue par l'équipe de l'Université de Calgary (Poncet et al., 2000).

L'évolution dans la performance de ces systèmes n'est pas discutée ici, puisque le présent mémoire n'améliore aucunement les méthodes d'acquisition de données. Par contre, des revues détaillées apparaissent dans Poncet (2001) et Jaremko (2001).

Une première source d'information sur la courbure rachidienne à partir de la surface du tronc se trouve dans le creux de la *vallée*, une courbe qui est généralement identifiable sur le dos. Plusieurs indices peuvent être dérivés de cette courbe, dont le pseudo-angle de Cobb (en anglais, *quasi-Cobb angle*) qui est mesuré de façon

analogue à l'angle de Cobb mais extraite de la surface du dos. Toutefois, le reste de la topographie du tronc est également importante; l'indice le plus connu est peut-être la rotation dorsale (en anglais, *back surface rotation*, BSR). Cet indice est évalué directement sur le patient à l'aide d'un instrument appelé scoliomètre, ou bien mesuré sur des données de la topographie du tronc.

Une source de variabilité entre ces études serait les différences dans la méthodologie utilisée par chaque équipe de chercheurs. Par exemple, la définition précise retenue pour l'extraction de la courbe correspondant à la vallée du dos aura un impact sur le calcul du pseudo-angle de Cobb. Ce mémoire n'entre pas dans tous ces détails de calcul d'indices externes, sauf pour remarquer qu'il s'agit d'une difficulté additionnelle de cette approche.

Ces indices du tronc souffrent des mêmes problèmes que l'angle de Cobb :

1. Ceux-ci sont assujettis à des erreurs de mesure relativement fortes. Par exemple, un intervalle de confiance de 95% pour l'erreur intraexamineur sur le BSR de  $4,8^\circ$  a été trouvé par Bunnell (1984), comparée à la valeur moyenne pour le BSR de  $8^\circ$ . L'erreur interexamineur était supérieure.
2. Il est difficile de complètement représenter la surface du tronc scoliotique avec un petit nombre d'indices externes.
3. La difficulté de développer des définitions et algorithmes de calcul précis et standardisés résultent en différentes approches de développement d'indices pour lesquels les comparaisons peuvent être difficiles.

Ce sont ces erreurs de mesure d'indices externes et interne et cette difficulté de complètement exprimer la géométrie de la surface du tronc et de la courbe rachidienne qui motivent la recherche d'autres techniques de représentation de données.

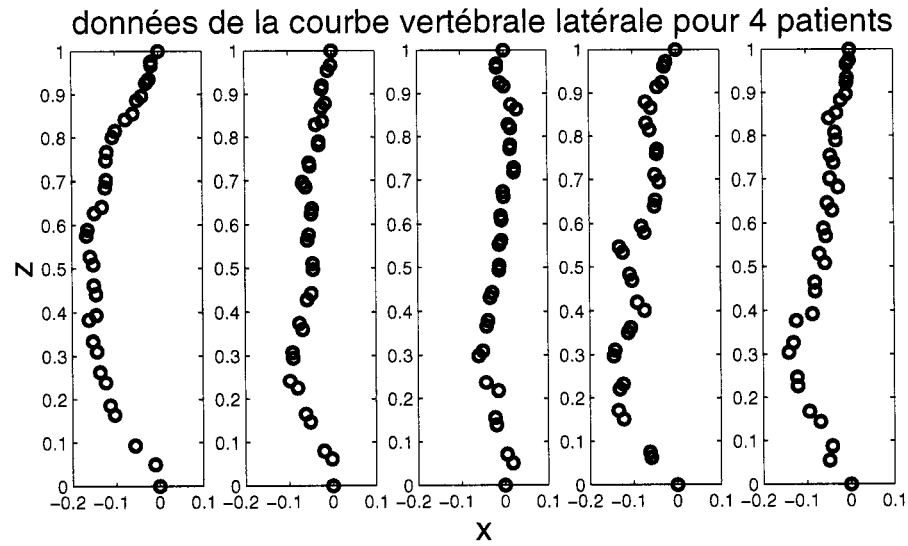


Figure 1.3 Données de la courbe rachidienne latérale pour quatre patients. Cette figure illustre que les données définissent des courbes rachidiennes latérales présentent un léger bruit. Figure générée à partir des données du *Alberta Children's Hospital*.

### 1.3 Approximation de fonction par moindres carrés

Cette section reprend l'idée de représenter les données scoliotiques externes et internes par l'approche d'analyse fonctionnelle de données (en anglais, *functional data analysis*, FDA). La figure 1.3 démontre des exemples de courbes rachidiennes, chacune définie par 34 repères anatomiques qui sont les centroïdes des plateaux supérieur et inférieur de chaque vertèbre thoracique et lombaire. La figure 1.4, quant à elle, présente un contour horizontal de la surface d'un tronc, formé de 360 points. Dans ces figures, les données semblent représenter des formes qui sont lisses, et apparaissent peu bruitées.

Les paragraphes qui suivent définissent et décrivent le problème fondamental en

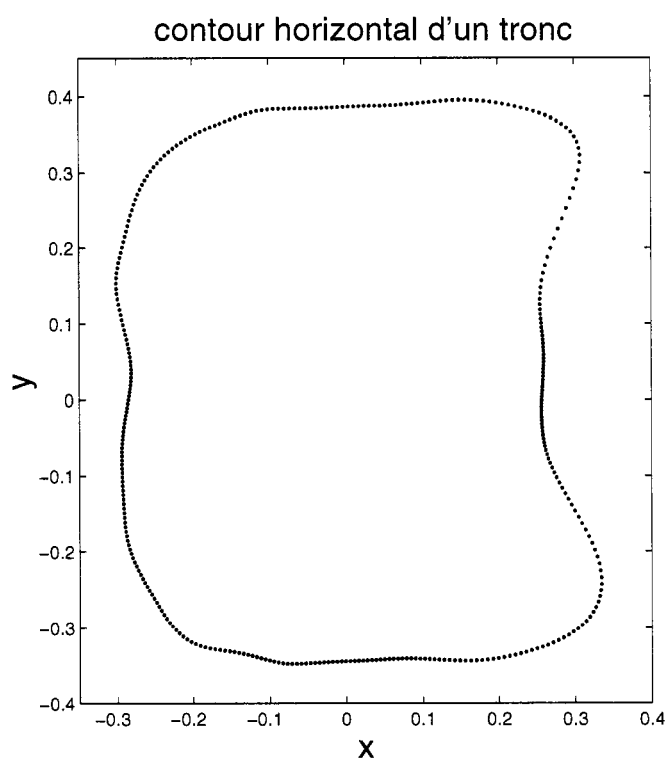


Figure 1.4 Contour horizontal d'un tronc. Cette figure affiche les données d'un contour horizontal de tronc, illustrant une forme très lisse et pas visiblement bruitée. Figure générée à partir des données du *Alberta Children's Hospital*.

FDA. L'approximation d'une fonction réelle  $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$  de  $n$  variables par une fonction d'approximation (ou approximant)  $F(\mathbf{x}, \mathbf{c})$  est définie par un nombre  $N$  fini de paramètres dénotés par  $\mathbf{c} = [c_1, \dots, c_N]^T$ . L'ensemble  $\mathcal{C} \subseteq \mathbb{R}^N$  constitue l'espace de ces paramètres. La validité d'un approximant  $F$  est évaluée selon la proximité de celle-ci à la fonction  $f$  ; cette proximité est mesurée à l'aide d'une fonction de distance  $\rho$ , qui peut ou non être une norme. Le problème d'approximation cherche à déterminer les paramètres  $\mathbf{c}^*$  dans  $\mathcal{C}$  telles que

$$\rho(f(\mathbf{x}), F(\mathbf{x}, \mathbf{c}^*)) \leq \rho(f(\mathbf{x}), F(\mathbf{x}, \mathbf{c})) \quad \forall \mathbf{c} \in \mathcal{C}$$

pour un choix donnée de  $\rho$ . Une solution  $F(\mathbf{x}, \mathbf{c}^*)$  de ce problème est dite la meilleure approximation.

Généralement, ce sont des normes qui sont utilisées comme fonction de distance (Rice, 1964). Deux normes fonctionnelles classiques sont les normes- $L_p$  de Hölder et la norme- $L_\infty$  de Tchebycheff. Lorsque  $p > 1$ , les grandes erreurs sont pénalisées plus fortement que les petites erreurs. Cette caractéristique se fait de plus en plus prononcée lorsque  $p \rightarrow \infty$ . Ces normes permettent l'introduction d'erreurs relativement grandes afin de réduire l'erreur maximale, puisque c'est cette dernière qui domine la norme. En contrepartie, pour  $p = 1$ , ce comportement disparaît, puisque les erreurs sont pénalisées également selon une relation linéaire. Le choix d'une norme dépend donc de l'objectif soutenant l'approximation. Dans le contexte de ce mémoire, la connaissance sur une fonction  $y = f(\mathbf{x})$ , soit une courbe rachidienne ou une surface du tronc sur un intervalle fermé  $\Omega$ , est limitée à un nombre fini de données  $(\mathbf{x}_i, y_i + \epsilon_i)$  où  $\epsilon$  représente un bruit dans les observations. L'objectif vise alors à identifier une meilleure approximation  $F(\mathbf{x}, \mathbf{c}^*)$  à partir de ces données, et le choix de la norme influence la solution  $\mathbf{c}^*$ . Plusieurs scénarios sont possibles, dont les trois suivants (Rice, 1964) :

- Lorsque  $\epsilon$  est une variable aléatoire indépendante présentant une distribution

gaussienne, il est connu que les paramètres  $\mathbf{c}^*$  par la norme- $L_2$  coïncident avec les estimés de vraisemblance maximale; ces derniers minimisent l'impact de  $\epsilon$  sur  $\mathbf{c}^*$ . Donc, la norme- $L_2$  constitue le meilleur choix.

- Lorsque  $\epsilon$  est une variable aléatoire indépendante présentant une distribution uniforme, c'est la norme de Tchebycheff qui est idéale.
- Toutefois, si les observations présentent des données aberrantes, ces dernières dominent les normes- $L_2$  et  $-L_\infty$ . Ce comportement n'est pas observé avec la norme- $L_1$ , qui pénalise toute erreur linéairement. Par conséquent, la norme- $L_1$  se voit alors être un choix supérieur.

En ce qui concerne la colonne vertébrale ou le tronc, les données grossièrement aberrantes sont rares, et les quelques cas d'exception sont corrigés manuellement ou bien exclus de cette étude. Les figures 1.3 et 1.4 témoignent de la faiblesse générale du bruit présent dans les données, faisant de la norme- $L_2$  un choix fort acceptable. Cette décision permet d'exploiter tout un paradigme d'approximation fonctionnelle existant appelée approximation par moindres carrés. Cette approche est idéale par sa facilité de calcul, l'existence de meilleure approximation unique, la capacité de déterminer explicitement les paramètres, et des solutions qui suivent de façon continue et uniforme la fonction à approcher (Rice, 1964). C'est cette approche qui fut exploitée par Staib et Duncan (1992) et Székely et al. (1996) pour l'identification des frontières de structures anatomiques dans le contexte de la segmentation MRI.

La solution du problème linéaire d'approximation- $L_2$

$$F(\mathbf{x}, \mathbf{c}) = \sum_{i=1}^N c_i \phi_i(\mathbf{x}),$$

telle que l'ensemble  $\Phi = \{\phi_i(\mathbf{x})\}$  contient des fonctions linéairement indépendantes (et habituellement imposées orthonormales pour faciliter les calculs) est donnée par

les coefficients  $c_i^*$  (Davis, 1963) :

$$c_i^* = \langle f, \phi_i \rangle = \int_{\Omega} f(\mathbf{x}) \phi_i(x) dx.$$

La somme s'appelle série généralisée de Fourier, et les paramètres s'appellent coefficients de Fourier. En théorie,  $N$  peut être infini, mais pour des raisons pratiques,  $N$  est habituellement imposé fini. Une discussion de la convergence s'avère toute-fois possible. Une condition nécessaire pour la convergence d'une telle série exige la convergence de la séquence de ses coefficients vers 0 (Swamy, 2002). En pratique, une convergence n'est pas toujours observée, car :

- l'évaluation des coefficients d'ordre élevé peut entraîner des instabilités numériques ; et
- les coefficients d'ordre élevé représentent possiblement le bruit dans les observations.

Aucun théorème formel ne définit la rapidité de la convergence ; toutefois, Davis (1963) remarque que la convergence d'une série généralisée de Fourier d'une fonction lisse est habituellement très rapide.

Il existe quelques sources d'inspiration concernant le choix de l'ensemble orthonormal  $\Phi$  :

1. Tout d'abord, l'adaptation de séries connues telles les séries de Taylor, de Laurent ou de Bernstein peut être considérée dans le choix d'un modèle (Rice, 1964).
2. Ensuite, les ensembles de fonctions explicites et transcendentaux qui sont linéairement indépendants sur un intervalle quelconque. L'exemple classique sont les sinusoides, périodiques avec période  $\omega$  : il peut s'agir d'une série sinus, cosinus, ou bien trigonométrique (la combinaison des deux). L'autre choix bien connu sont les polynômes : selon la nature de l'intervalle désiré, il

existe un ensemble orthonormal, tels les polynômes de Legendre sur  $[-1, 1]$  (Davis, 1963).

3. Une autre alternative consiste à diviser l'intervalle  $[a, b]$  en  $n$  sous-intervalles définis par les noeuds  $a = x_0 < x_1 < \dots < x_n = b$ , rendant chaque sous-intervalle orthogonal par rapport aux autres et permettant la définition d'un sous-ensemble orthonormal pour chaque sous-intervalle. Deux possibilités sont la droite brisée (de classe lisse par morceaux) interpolant les noeuds, ou bien la spline cubique (de classe lisse) interpolant les noeuds. En présence de noeuds bruités, ces méthodes d'interpolation peuvent être remplacées par des méthodes d'approximation en imposant une caractéristique lisse à l'approximant (Rice, 1964).
4. Finalement, les ensembles de fonctions implicites qui sont orthonormaux, soient les ondelettes. Il en existe plusieurs familles, dont celles de Daubechies ou de Meyer.

Cette classification est définitivement floue. Par exemple, les fonctions de Haar forment à la fois une famille d'ondelettes, des fonctions explicites d'expression sous la forme d'une série généralisée de Fourier, et une représentation par des fonctions constantes sur des sous-intervalles formant une droite par morceaux.

De façon générale, une base orthonormale est choisie selon les similarités qui existent entre cette base et les fonctions à approcher. En particulier, à considérer sont les attributs de la fonction  $f$  aux extrémités de l'intervalle d'intérêt lors du choix d'un approximant. Par exemple, les fonctions sinus de période  $\omega$  imposent la condition  $f(0) = f(\omega) = 0$  (Swamy, 2002), une contrainte acceptée pour la représentation de la courbe rachidienne dans l'article du chapitre 2.

## 1.4 Composantes principales fonctionnelles

Cette sous-section reprend l'idée de représenter un approximant sous une forme linéaire

$$F(\mathbf{x}, \mathbf{c}) = \sum_{i=1}^N c_i \phi_i(\mathbf{x}).$$

La convergence d'un approximant n'impose aucunement que les premiers coefficients soient grands ; la convergence n'explique que le comportement éventuel de la séquence. Par exemple, la représentation d'une fonction  $f$  dominée par de fortes oscillations de dixième mode par une série trigonométrique risque de comporter plusieurs coefficients de bas ordre presque nuls. Cette observation mène à la question *Est-il possible de redéfinir l'approximant de telle façon à accélérer la convergence ?* La réponse est affirmative. Le problème consiste à transformer l'ensemble orthonormal  $\Phi$  en un nouvel ensemble orthonormal  $\Psi = \{\psi(\mathbf{x})\}$  tel que la forme linéaire est reformulée comme suit :

$$F(\mathbf{x}, \mathbf{d}) = \sum_{i=1}^N d_i \psi_i(\mathbf{x})$$

pour laquelle les termes sont optimisés selon un critère spécifié.

La norme- $L_2$  est choisie comme critère, ce qui revient à maximiser la variance dans les premières composantes. Les composantes restantes se voient représenter de la variance résiduelle, et lorsque celles-ci sont jugées négligeables ou entraînées par du bruit, ces dernières peuvent être négligées, permettant une réduction de la dimensionnalité de  $\Psi$ . Cette méthode s'appelle l'analyse des composantes principales fonctionnelles (en anglais, *functional principal components analysis*, fPCA).

Les composantes principales sont calculées comme suit. Soit la matrice  $C$  formée des vecteurs de coefficients  $\mathbf{c}$  pour chaque échantillon. La matrice de covariance corres-

pondante est  $\Sigma = CC^T$ , pour laquelle la décomposition par valeurs singulières est donnée par  $V^T \Lambda V$ . La matrice  $\Lambda$  est diagonale, et contient les  $N$  valeurs singulières  $\lambda_i$  qui sont réelles, non-négatives, distinctes, et ordonnées de façon décroissante. La matrice  $V = [\omega_1 \omega_2 \cdots \omega_N]$  contient les vecteurs propres  $\omega_i$  associées à chaque valeur singulière. Cette matrice définit donc la transformée  $D = VC$ , où  $D$  contient les vecteurs de coefficients optimisés  $\mathbf{d}$  de chaque échantillon.

La variance représentée par chaque composante principale fonctionnelle est donnée par sa valeur singulière. La proportion de la variance totale représentée par la composante principale fonctionnelle  $k$  est donc donnée par

$$\frac{\lambda_k}{\sum_{i=1}^N \lambda_i}.$$

En définissant une tolérance, les composantes principales fonctionnelles  $\psi_i$  de grand ordre peuvent être négligées, réduisant la dimensionnalité de la base  $\Psi$  par rapport à  $\Phi$ .

En génie biomédical, PCA fut utilisé pour associer la forme de l'os fémoral à l'arthrite (Ramsay et Silverman, 2001), analyser la croissance adolescente (Ramsay et Silverman, 2001), détecter la frontière du disque optique de l'œil (Li et Chutatape, 2004), modéliser la forme de ligaments lors d'une chirurgie (Fleute et al., 1999), et automatiser la reconstruction vertébrale 3D à partir de radiographies biplanaires (Benamer et al., 2001). Dans la présente étude, les coefficients dérivés de la représentation fonctionnelle orthogonale de l'externe et de l'interne peuvent être calculés et ensuite réduits par cette approche fPCA.

### 1.5 Fractions fonctionnelles de bruit minimal

D'autres critères dans la définition de  $\Psi$  peuvent être choisis ; par exemple, il y a le ratio signal-à-bruit (en anglais, *signal-to-noise ratio*, SNR), un critère qui sépare alors les composantes de signal des composantes de bruit. Cette généralisation de fPCA dans un espace non euclidien s'appelle la transformée des fractions fonctionnelles de bruit minimal (en anglais, *functional minimum noise fractions*, fMNF).

En effet, lorsque les données utilisées sont bruitées, ce bruit se propage aux coefficients de la représentation fonctionnelle orthonormale. Sous ces conditions, il est intéressant de connaître le niveau de bruit présent dans chaque composante principale fonctionnelle afin d'effectuer un choix éclairé en ce qui concerne la réduction de dimensionnalité. En effet, si le bruit est blanc, additif et décorrélé, la séquence des valeurs singulières  $\lambda_i$  de la section précédente est telle que (Hyvärinen et al., 2001)

$$\lambda_1 > \lambda_2 > \dots > \lambda_{N'} > \lambda_{N'+1} \approx \dots \approx \lambda_N \approx \sigma_N^2$$

pour  $N' < N$  où  $\sigma_N^2$  est la variance de ce bruit sur les coefficients. Le SNR est le ratio du signal sur le bruit. Ayant estimé  $\Sigma_N$ , la matrice de covariance du bruit, et exploitant la propriété additive de la variance :

$$SNR = \frac{\omega^T \Sigma \omega}{\omega^T \Sigma_N \omega} - 1.$$

Par conséquent, le SNR est maximisé par les vecteurs propres  $\omega_i$  de  $\Sigma$  par rapport à  $\Sigma_N$ , et prend les valeurs

$$\kappa_k - 1$$

où  $\kappa_k$  sont les valeurs singulières associées.

Cette transformée fMNF fut proposée par Green et al. (1988) dans le contexte

de décomposition spectrale d'images satellite. Récemment, Larsen et Hilger (2003) ont appliqué cette transformée à l'annotation des poumons sur des radiographies et Hilger et al. (2003) l'ont appliquée à la modélisation de la croissance des mandibules chez l'humain. Dans la présente étude, des connaissances sur l'erreur dans les données externes ( $4,2 \pm 1,7$  millimètres) (Poncet et al., 2000; Tardif et al., 2000) et internes (2,2 millimètres dans la projection frontale et 5,6 millimètres dans la projection latérale) (André et al., 1992; Poncet et al., 2000) peuvent être exploitées pour estimer la matrice de covariance du bruit  $\Sigma_N$ . Il est supposé que ces erreurs présentent des distributions pouvant être modélisées par du bruit blanc, additif et décorréolé desquelles dépendent l'approche fMNF.

## 1.6 Régression linéaire

En rappel, une machine de régression linéaire optimise les paramètres qui sont le vecteur de poids  $\mathbf{w} \in \mathbb{R}^N$  et la constante de biais  $b \in \mathbb{R}$  dans la relation

$$F(\mathbf{x}, \mathbf{w}, b) = (\mathbf{w} \cdot \mathbf{x}) + b$$

avec l'objectif que ce modèle représente bien la fonction inconnue  $y = f(\mathbf{x})$  à partir des données empiriques  $\{(\mathbf{x}_m, y_m)\}$ , c'est-à-dire de minimiser l'erreur ou le risque empirique  $R^{emp}$  dans l'expression

$$f(\mathbf{x}) = F(\mathbf{x}, \mathbf{w}, b) + R^{emp}.$$

Deux aspects à noter sont :

- La notation  $F(\mathbf{x}, \mathbf{w}, b)$  devient conforme à celle utilisée à la section 1.3, pour laquelle l'approximant est dénoté  $F(\mathbf{x}, \mathbf{c})$ , lorsque les poids  $\mathbf{w}$  et le biais  $b$  sont rassemblés dans un même vecteur  $\mathbf{c} = [\mathbf{w}^T, b]^T$ .

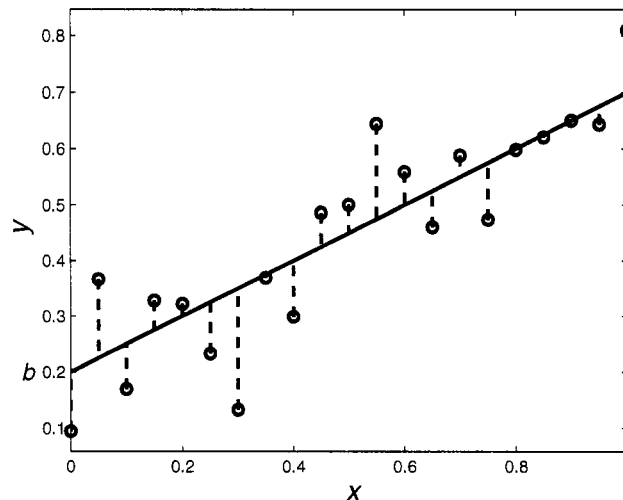


Figure 1.5 Erreurs empiriques d'une régression linéaire. Cette figure présente une régression linéaire  $y = F(x) = mx + b$  déterminée à partir d'un ensemble de données empiriques  $\{(x_i, y_i)\}$ . La pente de la droite est dénotée  $m$ . Les paramètres  $m$  et  $b$  sont déterminés de telle sorte à minimiser les erreurs, identifiées par des lignes tiretées, selon une règle quadratique. Figure inspirée de (Cristianini et Shawe-Taylor, 2000).

- Le terme  $R^{emp}$  n'est pas une fonction, mais représente plutôt une métrique de la différence entre les fonctions  $f$  et  $F$ .

Généralement, la norme-2 est choisie comme critère de minimisation de  $R^{emp}$ . La figure 1.5 illustre le scénario.

Une mesure d'efficacité d'une régression linéaire est le coefficient de détermination  $r^2$ , défini comme suit : soit un modèle  $y = F(x) = mx + b$  déterminé à partir d'un ensemble de données empiriques, les vecteurs  $\mathbf{x}$  et  $\mathbf{y}$ . Alors

$$r^2 = \frac{\|\mathbf{F}(\mathbf{x}) - \mathbf{y}\|}{\|\mathbf{y} - \bar{\mathbf{y}}\|},$$

la norme-2 est choisie et  $\bar{\mathbf{y}}$  dénote la moyenne des valeurs du vecteur  $\mathbf{y}$ .

Plusieurs études ont tenté d'identifier une relation entre des indices de l'asymétrie de la surface du tronc et l'angle de Cobb (Daruwalla et Balasubramaniam, 1985; Weisz et al., 1988; Turner-Smith et al., 1988; Stokes et al., 1988; Stokes, 1989; Stokes et Moreland, 1989; Drerup et Hierholzer, 1996; Koroivessis et Stamatakis, 1996; Samuelsson et Noren, 1997; Theologis et al., 1997; Goldberg et al., 2001). La régression linéaire simple d'un indice externe à l'angle de Cobb varie largement d'une étude à une autre. Les meilleurs résultats sont comme suit :

1. Le coefficient de détermination  $r^2$  utilisant comme prédicteur le BSR maximal couvre l'intervalle 0,084–0,42 (Koroivessis et Stamatakis, 1996; Huang, 1997; Samuelsson et Noren, 1997). Koroivessis et Stamatakis (1996) ont rapporté une distribution pour l'erreur moyenne entre l'angle de Cobb thoracique mesurée et prédite à partir du BSR thoracique obtenu du scoliomètre de  $5,63^\circ \pm 3,16^\circ$ ; la distribution pour l'angle de Cobb et le BSR lombaires était  $5,79^\circ \pm 2,75^\circ$ . La relation linéaire entre le BSR et l'angle de Cobb maximaux trouvée par Jaremko et al. (2002b) explique 20% de la variabilité du modèle, présentant une distribution d'erreur de  $6,9^\circ \pm 3,2^\circ$ , tandis que Stokes et al. (1988) a expliqué 59% de la variance du modèle avec une méthodologie analogue.
2. En choisissant le pseudo-angle de Cobb comme prédicteur, Goldberg et al. (2001) ont calculé un  $r^2$  de 0,66. Jaremko et al. (2002a), quant à eux, ont rapporté une valeur de 0,27.
3. La régression du *hump sum*, une somme de trois indices d'asymétrie du tronc, à l'angle de Cobb trouvée par Suzuki et al. (1992) présente un  $r^2$  de 0,29.

L'erreur de mesure des indices externes et interne contribue sans doute fortement à cette variabilité dans la qualité des prédictions externe-interne pour la scoliose par régression linéaire simple. À noter qu'un bémol doit être porté à la comparaison froide des résultats présentés. En effet, tel que mentionné dans les sections 1.1 et

1.2, les indices de déformation scoliotique sont mesurées différemment d'une étude à une autre, expliquant une partie de la variabilité trouvée ci-haut. Ceci constitue une problématique en ce qui concerne l'identification d'une solide relation externe-interne.

Du côté de la régression linéaire multiple, Jaremko et al. (2001) ont sélectionné des groupes d'indices utilisés comme variables explicatives pour l'angle de Cobb. Le résultat le plus robuste présente un coefficient de détermination de 0,73 dans l'ensemble d'apprentissage et de 0,72 dans l'ensemble de test. Toutefois, seulement 64% des angles de Cobb de l'ensemble d'apprentissage et 42% dans l'ensemble de test sont prédits avec une erreur de moins de 5 °.

Les relations fortes et robustes recherchées demeurent méconnues, mais ces études demeurent encourageantes, en ce sens qu'il existe au moins de faibles liens entre les géométries externe et interne, suggérant que l'amélioration de la méthodologie de régression pourrait engendrer des meilleures performances. De plus, il est reconnu que l'imposition d'une contrainte de linéarité entre les indices comme modèle de physiologie humaine ne peut représenter toutes les complexités de cette relation. Cette réalité impose donc que des méthodes permettant des liens non linéaires soient tentées.

## 1.7 Réseaux de neurones

Confronté à des relations relativement faibles, et ce malgré l'augmentation de la précision des méthodes d'acquisition et de traitement de données, Jaremko et al. (2002a) ont introduit l'utilisation de techniques d'apprentissage au problème de la prédiction externe-interne scoliotique. Pour ce faire, une machine combinant l'algorithme génétique et un réseau de neurones artificiel fut entraînée dans le but

de relier un ensemble d'indices du tronc à l'angle de Cobb. Le GA choisit les indices utilisés en entrée à l'ANN en étudiant l'effet de choix différents d'indices sur le résultat. Les meilleurs prédicteurs sont retenus par une simulation inspirée de l'évolution génétique (Tsoukalas et Uhrig, 1997). Le ANN, quant à lui, ajuste les coefficients d'un réseau qui imite la capacité d'apprentissage non linéaire d'un système de neurones, de façon à minimiser l'erreur (Tsoukalas et Uhrig, 1997). Cette configuration avait été adoptée dans des études médicales précédentes (Dybowski et al., 1996; Jefferson et al., 1997, 1998a,b). Ce système GA-ANN a identifié une régression avec  $r^2 = 0,94$  dans le groupe d'apprentissage et  $r^2 = 0,87$  dans le groupe de test, soit une amélioration par rapport à MLR, mais démontrant tout de même un manque de robustesse. Les difficultés de cette approche incluaient le bruit dans les valeurs d'entrée et de sortie limitant le processus d'apprentissage, le risque que les variables choisies ne représentaient pas toute l'information utile contenue dans le tronc, et l'exploitation d'un nombre d'échantillons relativement petit contenant des interdépendances (dues aux acquisitions répétées de patients). Finalement, ANN présente le danger de trouver des solutions qui sont localement optimales mais pas globalement extrêmes.

## 1.8 Régression par vecteurs de support

Le besoin de trouver une régression non linéaire à partir de données bruitées par des techniques d'apprentissage possédant une forte capacité de généralisation est clair. Le manque de robustesse dans l'étude de Jaremko et al. (2002a) a donc motivé la recherche d'un nouvel outil d'intelligence artificielle. La méthode de régression par vecteurs de support, présente des propriétés intéressantes de généralisation (Vapnik, 1998; Cristianini et Shawe-Taylor, 2000; Kecman, 2001; Schölkopf et Smola, 2002). Beaucoup d'intérêt pour la classification fait en sorte que peu d'applications

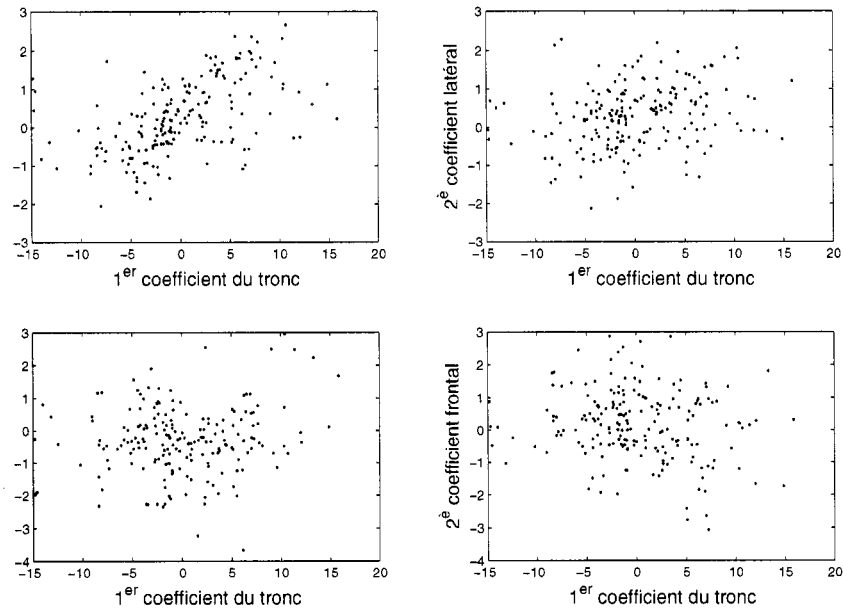


Figure 1.6 Graphes de coefficients externes par rapport aux coefficients internes. Cette figure présente les valeurs des deux premiers coefficients fMNF de la courbe rachidienne en fonction du premier coefficient fMNF de la topographie du tronc. Ces graphes démontrent la forte variabilité dans la relation externe-interne recherchée.

faisant l'objet d'une modélisation par SVR apparaissent dans la littérature ; elle fut toutefois exploitée dans la détermination du dosage d'érythropoietine pour les patients d'anémie secondaire (Martín-Guerrero et al., 2003).

La figure 1.6 illustre clairement la problématique du présent problème : les relations entre les coefficients externes et internes sont très variables. Cette variabilité, qui est traitée comme un bruit par rapport à la relation recherchée, est à la fois interpatient (par exemple, la forme du tronc dépend non seulement de la scoliose mais aussi du patient) et intrapatient (causée par des variations dans la posture lors de l'acquisition, par exemple).

*La méthodologie SVR est un exemple rare pour laquelle l'intuition géométrique,*

*l'élégance mathématique, les garanties théoriques et les algorithmes pratiques s'entrecroisent* (Bennett et Campbell, 2000). L'algorithme SVR est fondé sur sept idées de base, présentées dans les pages qui suivent afin de comprendre la dérivation de cette méthode d'apprentissage :

1. L'adoption de la fonction  $\epsilon$ -insensible, donnée par

$$|y - F(\mathbf{x}, \mathbf{w}, b)|_\epsilon = \max \{0, |y - F(\mathbf{x}, \mathbf{w}, b)| - \epsilon\},$$

comme métrique de l'erreur, retenant la faible pénalité de petites erreurs de la norme-2 ainsi qu'une pénalisation modérée de grandes erreurs de la norme-1 (Kecman, 2001). Le graphique de cette fonction apparaît à la figure 1.7. Utilisant cette métrique, l'erreur empirique est alors donnée par l'expression

$$R_\epsilon^{emp}[F] = \frac{1}{M} \sum_{m=1}^M |y_m - F(\mathbf{x}_m, \mathbf{w}, b)|_\epsilon.$$

Ici,  $M$  est le nombre de données empiriques. La figure 1.8 démontre l'effet d'une telle fonction de pénalité sur l'obtention d'une relation non linéaire robuste.

2. La minimisation d'une fonction objectif donnée par

$$R_\epsilon^{reg}[F] = \frac{1}{2} \|\mathbf{w}\|^2 + C \cdot R_\epsilon^{emp}[F], \quad (1.1)$$

soit une somme pondérée (par le paramètre  $C$  de régularisation) du risque empirique et de la complexité  $\|\mathbf{w}\|$  du modèle. Cette nouvelle mesure de l'erreur  $R_\epsilon^{reg}$  s'appelle le risque structurel (ou régularisé) (Vapnik, 1998). La figure 1.9 illustre l'importance de minimiser la complexité du modèle par quatre interpolations non linéaires de neuf données empiriques. Le résultat le plus simple correspond probablement le mieux à la fonction réelle, sous

## Métriques de distance

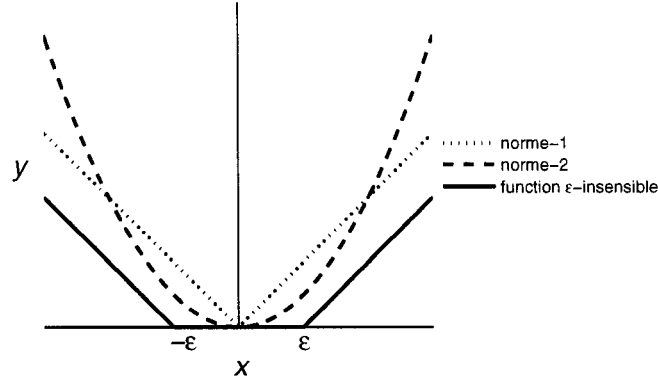


Figure 1.7 Fonctions de pénalité. Cette figure présente les fonctions de pénalité utilisées pour évaluer le risque empirique d'une régression. Ces fonctions sont la norme-1, la norme-2, et la fonction  $\epsilon$ -insensible. Cette dernière combine des propriétés avantageuses des deux premières, soit une pénalisation faible des petites erreurs et modérée des grandes erreurs.

l'hypothèse où celle-ci est lisse.

3. L'expression des paramètres de la régression en fonction des données empiriques. Soit la régression linéaire de la section 1.3

$$F(\mathbf{x}, \mathbf{w}, b) = (\mathbf{w} \cdot \mathbf{x}) + b \quad (1.2)$$

pour laquelle le risque structurel est donné par

$$R_{\epsilon}^{reg}[F] = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{M} \sum_{m=1}^M |y_m - (\mathbf{w} \cdot \mathbf{x}_m) - b|_{\epsilon}.$$

Une condition nécessaire pour que  $\mathbf{w}$  minimise ce risque est

$$\frac{\partial R_{\epsilon}^{reg}}{\partial \mathbf{w}} = 0.$$

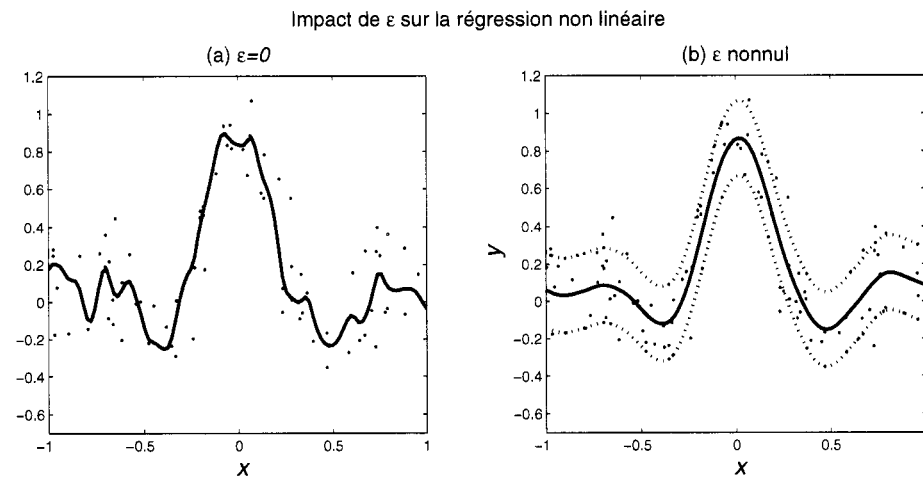


Figure 1.8 Impact du paramètre d'insensibilité. Cette figure démontre l'impact du paramètre d'insensibilité  $\epsilon$  sur une régression non linéaire. Les données sont générées à partir de la fonction lisse  $y(x) = 0,9\text{sinc}(\frac{10x}{\pi})$  à laquelle un bruit gaussien présentant un écart-type de 0,2 a été ajouté. La régression trouvée en (a) ne représente pas bien la tendance recherchée. La bande insensible, définie par les courbes pointillées en (b), dans laquelle les erreurs ne sont pas pénalisées, permet d'identifier une régression qui est plus stable. SVR est donc bien adapté pour identifier des relations malgré la présence de bruit dans les données empiriques. Figure inspirée des tests de Kwok et Tsang (2003).

L'expression de la dérivée partielle est

$$\frac{\partial R_\epsilon^{reg}}{\partial \mathbf{w}} = \mathbf{w} + \frac{C}{M} \sum_{m=1}^M |y_m - (\mathbf{w} \cdot \mathbf{x}_m) - b|_\epsilon \mathbf{x}_m.$$

Combinant ces deux expressions, on trouve :

$$\mathbf{w} = -\frac{C}{M} \sum_{m=1}^M |y_m - (\mathbf{w} \cdot \mathbf{x}_m) - b|_\epsilon \mathbf{x}_m.$$

Il existe donc des  $\beta_m$  tels que

$$\mathbf{w} = \sum_{m=1}^M \beta_m \mathbf{x}_m.$$

Remplaçant cette représentation pour  $\mathbf{w}$  dans l'équation (1.2) donne

$$F(\mathbf{x}, \mathbf{w}, b) = (\mathbf{w} \cdot \mathbf{x}) + b = \sum_{m=1}^M \beta_m (\mathbf{x}_m \cdot \mathbf{x}) + b, \quad (1.3)$$

permettant ainsi de représenter le modèle  $F(\mathbf{x}, \mathbf{w}, b)$  avec un produit scalaire en fonction des données empiriques.

4. La généralisation des machines de régression linéaire en permettant une transformation non linéaire  $\phi : \mathcal{X} \rightarrow \mathcal{H}$  des vecteurs d'entrée. C'est dans cet espace transformé que le produit scalaire, et donc une régression linéaire, est calculé :

$$F(\mathbf{x}, \mathbf{w}, b) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle + b.$$

L'espace  $\mathcal{H}$  est généralement de dimensionnalité plus grande que celle de  $\mathcal{X}$  (Cristianini et Shawe-Taylor, 2000).

5. Le calcul implicite des produits scalaires dans l'espace  $\mathcal{H}$ . Cet aspect est essentiel, puisque la dimensionnalité de ce dernier est habituellement trop

grande pour permettre un temps de calcul réaliste (Cristianini et Shawe-Taylor, 2000). L'approche consiste à définir une fonction noyau  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  telle que

$$k(\mathbf{x}, \mathbf{x}') = (\boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{x}')).$$

Sous cette formulation, si elle existe, le produit scalaire de  $\boldsymbol{\phi}(\mathbf{x})$  avec  $\boldsymbol{\phi}(\mathbf{x}')$  est directement obtenu de  $\mathbf{x}$  et  $\mathbf{x}'$ , sans avoir à travailler dans l'espace  $\mathcal{H}$ . La condition devant être imposée sur  $k(\mathbf{x}, \mathbf{x}')$  afin que celle-ci corresponde au produit scalaire de deux éléments dans un espace  $\mathcal{H}$  quelconque est simple :  $k$  doit être défini positif (Schölkopf et Smola, 2002). Le tableau 1.1 recueille plusieurs noyaux définis positifs.

6. L'exploitation de l'*astuce des noyaux* qui affirme que, dans tout algorithme, un noyau défini positif  $k$  peut être remplacé par un autre noyau défini positif  $\tilde{k}$  (Cristianini et Shawe-Taylor, 2000). Cette astuce permet de dériver une machine de régression linéaire, puis ensuite de remplacer le produit scalaire canonique, par exemple celui de l'équation (1.3), par tout autre noyau défini positif (Schölkopf et Smola, 2002) :

$$F(\mathbf{x}, \mathbf{w}, b) = \sum_{m=1}^M \beta_m k(\mathbf{x}_m, \mathbf{x}) + b, \quad (1.4)$$

7. Finalement, les valeurs optimales pour  $\beta_m$  sont trouvées. Le risque structurel de l'équation (1.1) est minimisé par la formulation d'un problème d'optimisation qui, lorsqu'exprimé sous sa forme duale par multiplicateurs de Lagrange et transformé par les conditions de Karush-Kuhn-Tucker, devient quadratique

Tableau 1.1 Recueil de fonctions noyaux définis positifs

| noyau                                                                                      | formulation mathématique                                                                                                                                                             |
|--------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| produit scalaire canonique                                                                 | $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$                                                                                                                          |
| polynômial<br>(linéaire : $p = 1$ ,<br>quadratique : $p = 2$ , etc.)                       | $k(\mathbf{x}, \mathbf{y}) = ((\mathbf{x} \cdot \mathbf{y}) + 1)^p$                                                                                                                  |
| spline cubique                                                                             | $k(\mathbf{x}, \mathbf{y}) = 1 + (\mathbf{x} \cdot \mathbf{y}) + \frac{1}{2}(\mathbf{x} \cdot \mathbf{y}) \min(\mathbf{x}, \mathbf{y}) - \frac{1}{6} \min(\mathbf{x}, \mathbf{y})^3$ |
| sigmoïde<br>(sgm peut être arctan, sinh, etc.)                                             | $k(\mathbf{x}, \mathbf{y}) = \text{sgm}(\gamma + \sigma(\mathbf{x} \cdot \mathbf{y}))$                                                                                               |
| <i>radial basis function</i> (RBF)<br>(exponentiel : $q = 1$<br>gaussien : $q = 2$ , etc.) | $k(\mathbf{x}, \mathbf{y}) = \exp(-\sigma \ \mathbf{x} - \mathbf{y}\ ^q)$                                                                                                            |
| polynôme de Vovk<br>(réel de degré $r$ ,<br>infini lorsque $r = 0$ )                       | $k(\mathbf{x}, \mathbf{y}) = \frac{1 - (\mathbf{x} \cdot \mathbf{y})^r}{1 - (\mathbf{x} \cdot \mathbf{y})}$                                                                          |
| <i>moderate decreasing</i><br>(KMOD) (Ayat et al., 2002)                                   | $k(\mathbf{x}, \mathbf{y}) = \frac{1}{\exp(\frac{\gamma}{\sigma^2 - 1})} \exp(\frac{\gamma}{\ \mathbf{x} - \mathbf{y}\  + \sigma^2})$                                                |

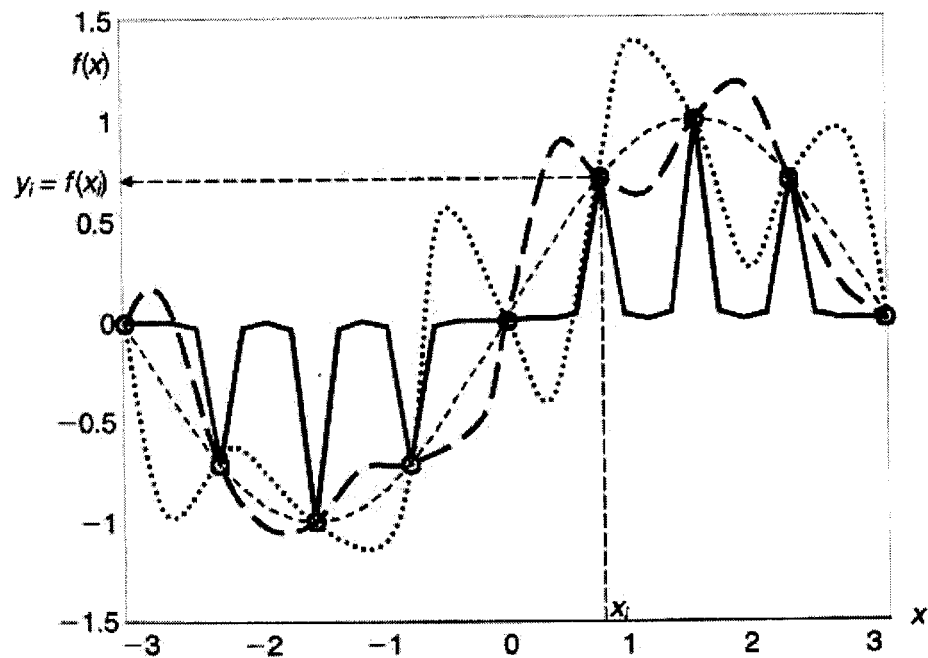


Figure 1.9 Quatre régressions non linéaires de données. Cette figure présente quatre régressions non linéaires de données empiriques  $(x_i, y_i)$  (représentées par des ronds) d'une fonction quelconque  $y = f(x)$  qui est inconnue. En fait, chaque régression est une interpolation, pour laquelle l'erreur empirique est nulle. La diversité des régressions ainsi trouvées démontre la faiblesse du seul critère  $R^{emp}[F]$  dans le choix d'une régression. Admettant que les données fournies sont représentatives de la fonction recherchée et que cette fonction est lisse, c'est la courbe fine tiretée qui devrait être retenue, justifiant la minimisation de la complexité de la régression. Figure tirée de (Kecman, 2001).

et convexe (Schölkopf et Smola, 2002) : Trouver  $\alpha$  et  $\alpha^*$  de façon à maximiser

$$W(\alpha, \alpha^*) = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(\mathbf{x}_i, \mathbf{x}_j) - \epsilon \sum_{i=1}^M y_i (\alpha_i^* - \alpha_i) + \sum_{i=1}^M y_i (\alpha_i^* - \alpha_i)$$

assujetti aux contraintes

$$\begin{aligned} \sum_{i=1}^M (\alpha_i^* - \alpha_i) &= 0 \\ \alpha_i^*, \alpha_i &\in [0, \frac{C}{m}] \quad \forall i. \end{aligned}$$

La solution optimale de ce problème est dénotée  $\bar{\alpha}$  et  $\bar{\alpha}^*$ . Poser  $\beta = \bar{\alpha}^* - \bar{\alpha}$  permet de retrouver les  $\beta_m$  de la régression non linéaire (1.4) :

$$F(\mathbf{x}, \mathbf{w}, b) = \sum_{m=1}^M \beta_m k(\mathbf{x}_m, \mathbf{x}) + b.$$

L'approche SVR évite de nombreuses faiblesses d'autres méthodes d'inférence telles les ANN et les arbres de décision (Bennett et Campbell, 2000) :

1. Aucun problème de minimum local. Des régressions fortement non linéaires peuvent être élaborées, sans danger de se retrouver avec une solution localement extrême mais pas globalement optimale.
2. Peu de paramètres à sélectionner. Par exemple, pour un réseau RBF, seulement trois paramètres doivent être optimisés : le paramètre de régularisation  $C$ , l'insensibilité  $\epsilon$  et la largeur du noyau  $\sigma$ .
3. Les résultats sont stables et reproductibles. Deux utilisateurs appliquant le même modèle SVR obtiennent le même résultat assujetti aux considérations numériques telle la précision machine. Le résultat est largement indépendant

de l'algorithme de résolution du problème d'optimisation choisi.

4. Des algorithmes de résolution robustes existent. En effet, les problèmes d'optimisation quadratique ont fait l'objet de recherches très développées, faisant en sorte que des algorithmes puissants sont existants, permettant de résoudre des problèmes de classification comptant quelques millions de données empiriques. Par exemple, l'algorithme de Dantzig-Wolfe (Fletcher, 1987) constitue une méthode pour laquelle le nombre d'itérations est normalement inférieur à trois fois la somme du nombre de variables et de contraintes (Dantzig, 1963).
5. La méthodologie est relativement simple à exécuter.

Il existe toutefois quelques désavantages associées aux SVR. En particulier (Bennett et Campbell, 2000) :

1. SVR est-il le meilleur outil d'apprentissage pour la régression ? Certes il existe des exemples de données pour lesquels SVR ne trouve pas le meilleur modèle, ces circonstances n'excluent pas la possibilité que SVR soit le meilleur algorithme sur la moyenne d'une grande variété de contextes. Il reste que SVR est peut-être mieux adapté à certains problèmes qu'à d'autres.
2. La difficulté de sélectionner des paramètres du SVR et du noyau demeure problématique. La validation croisée demeure la solution couramment utilisée, mais celle-ci coûte cher en temps d'exécution. Un bémol à cette affirmation peut toutefois être postulé, puisque quelques résultats théoriques concernant l'optimisation automatique de  $\epsilon$  (Kwok et Tsang, 2003) commencent à apparaître dans la littérature.
3. Les connaissances *a priori* sont difficiles à incorporer dans le modèle, surtout étant donné la transformation non linéaire dans un hyperespace.

## 1.9 Positionnement du mémoire

Si le nombre de publications sur la relation externe-interne scoliotique semble croître, la plupart de celles-ci se limitent à l'identification de nouveaux indices cliniques pour lesquelles l'approche traditionnelle de régression simple ou multiple se voit améliorée, sans toutefois obtenir des résultats suffisamment robustes pour usage clinique.

L'objectif général du présent mémoire vise à identifier une relation robuste entre la forme du tronc et la forme de la colonne vertébrale. Pour ce faire, les deux objectifs spécifiques sont :

1. Analyser la représentation fonctionnelle de données afin de représenter la topographie du tronc et la courbe de la colonne vertébrale scoliotique à l'aide de coefficients qui sont globaux, représentatifs de la géométrie et moins sensibles au bruit d'acquisition ; et
2. Étudier la régression par vecteurs de support comme technique d'apprentissage robuste pour relier ces coefficients.

## CHAPITRE 2

### PRÉDICTION DE LA COURBE SCOLIOTIQUE

Ce chapitre contient un premier article intitulé *Prediction of anterior scoliotic spinal curve from trunk surface using support vector regression* pour la publication *Engineering Applications of Artificial Intelligence*. Cet ouvrage fut rédigé en collaboration avec Farida Cheriet (École Polytechnique de Montréal et Hôpital Sainte-Justine), Janet Ronsky (Université de Calgary), Ronald Zernicke (Université de Calgary) et Hubert Labelle (Hôpital Sainte-Justine). Le 15 mars 2005, l'éditeur a confirmé la publication de cet article, dont la citation est (Bergeron et al., accepté 2005). La contribution relative de l'auteur de ce mémoire à cet article est estimée à 90%.

#### 2.1 Synthèse

Cet article fait état de la prédiction de la courbe rachidienne scoliotique projetée dans le plan frontal à partir de la topographie du tronc. L'acquisition et le traitement de données sont traités, et les séries trigonométriques sont privilégiées pour représenter les courbes et surfaces. L'analyse par composantes principales fonctionnelles détermine les coefficients qui deviennent les valeurs d'entrée et de sortie de la régression par vecteurs de support. Le schéma de la figure 2.1 illustre l'interaction entre les objectifs de la section 1.9 et les données et méthodes de cet article. Le modèle résultant est robuste et permet une prédiction avec une fidélité de moins de 4,5 millimètres.

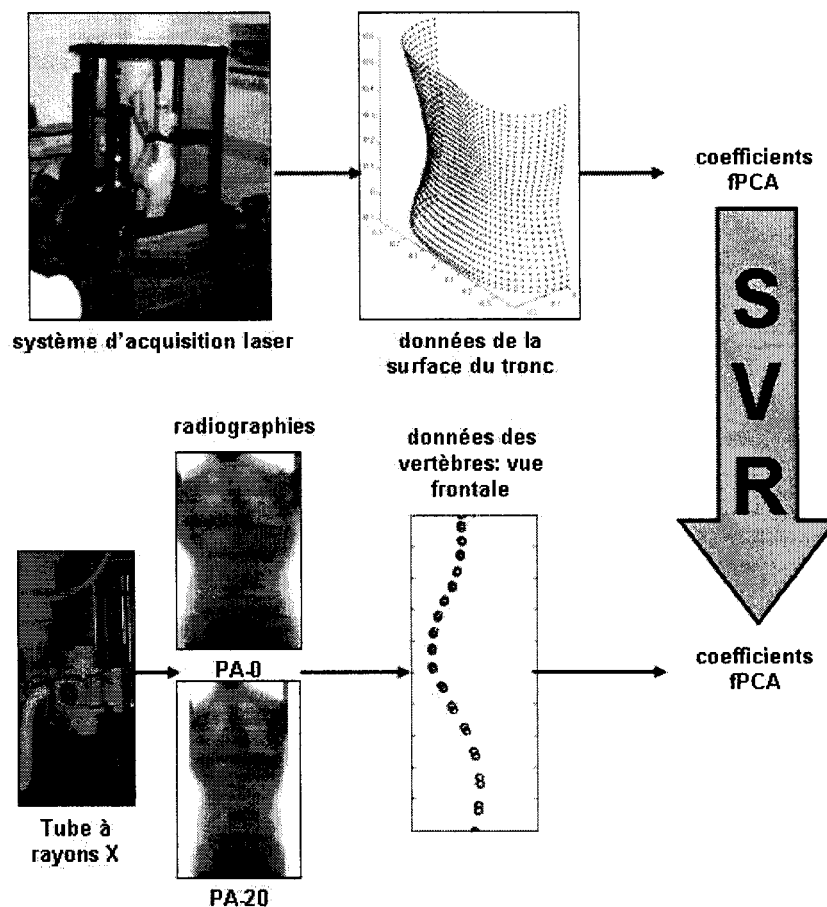


Figure 2.1 Ce schéma présente les liens entre les données et les méthodes de cet article. Les données de la topographie du tronc sont acquises par un système de caméras laser 3D. Quasi-simultanément, deux radiographies sont acquises, et de cette imagerie une reconstruction de la courbe de la colonne vertébrale est obtenue, d'où les données vertébrales. L'analyse par composantes principales fonctionnelles permet d'extraire de ces données un petit nombre de coefficients fPCA qui représentent cette surface et courbe qui sont globaux, moins sensibles au bruit et représentatifs de la géométrie. La régression par vecteurs de support est ensuite invoquée, étant une méthode d'apprentissage robuste, afin d'identifier la relation entre les coefficients externes et les coefficients internes. La combinaison de ces méthodes permet d'adresser les objectifs global et spécifiques de ce mémoire.

## 2.2 Prediction of anterior scoliotic spinal curve from trunk surface using support vector regression

### 2.2.1 Introduction

Learning techniques are powerful methods for modeling when a representative set of input-output examples are available. However, ideal data is often unavailable, requiring that the artificial intelligence (AI) practitioner to take special care in system design and results interpretation. This paper addresses these issues using a dataset of biomedical data from scoliosis patients in which interdependencies, as well as inter- and intra-sample noise, is present.

Idiopathic scoliosis is a three-dimensional (3D) deformation of the spinal column affecting 2–4% of persons and manifesting itself during childhood (3–10 years, 12–21% of cases) or adolescence ( $> 10$  years) (Reamy and Slakey, 2001). Spinal curve magnitude is normally evaluated from a full spinal X-ray in standing position. Given the variability of adolescent growth on an annual (Tuddenham and Snider, 1954) and even daily (Thalange et al., 1996) basis, frequent follow-up is necessary. Reamy and Slakey (2001) recommend X-ray acquisition every 6 months until skeletal maturity for all patients; the limiting factor is the risk associated with serial exposure to harmful radiation. Modern equipment has reduced the dosage, and postero-anterior (PA) views have reduced its impact on sensitive tissues, yet the incidence of cancer of the breast, thyroid, lung, ovary and bone marrow increases by up to 2.4 per thousand patients—comparable to the probability of developing thyroid cancer in the general population (Levy et al., 1996). Given that fewer than 1 in 10 scoliosis patients require treatment beyond periodic monitoring (Reamy and Slakey, 2001), it is desirable to reduce radiation exposure as much as possible. Exploiting magnetic resonance imaging (MRI) would eliminate this risk, but

the high cost and low availability of this technology results in its use being limited to the most severe cases (Carro, 1999). As an additional problem, MRI is acquired while supine (lying down), a position in which scoliosis deformities are relaxed. Thus, an alternate and less invasive technique for scoliosis follow-up is needed.

Trunk deformity is often the first indicator of scoliosis, and remains the most visible effect. Thus, one approach has involved linking the trunk surface asymmetry to spinal curvature, under the assumption that the underlying physiological links are sufficiently consistent among patients. The classical technique of seeking a linear regression between two variables has been extensively explored. The most studied trunk-to-spine relation links maximal back surface rotation to primary anterior Cobb angle (the traditional clinical measure of spinal deformity); recent studies reported coefficients of determination  $r^2$  ranging between 0.084 and 0.42 (Korovesis and Stamatakis, 1996; Huang, 1997; Samuelsson and Noren, 1997; Jaremko et al., 2002a). Two papers reported values for Cobb angle estimation from a different trunk feature: a *quasi-Cobb angle* (analogous to Cobb angle but measured from the posterior spinous process curve on the back surface); values for  $r^2$  were 0.66 by Goldberg et al. (2001) and 0.27 by Jaremko et al. (2002a). These interstudy variabilities are problematic and are probably due to the sensitivity of indices, as local measures, to measurement error, as well as the precision of the data acquisition methods used. For instance, Goldberg et al. (1988) reported a 95% confidence interval for the intra-examiner error of  $5^\circ$  on the Cobb angle, and Bunnell (1984) reported a 95% confidence interval for the intra-examiner measurement error of  $4.8^\circ$  in the back surface rotation. For both measures, inter-examiner errors were higher. Those errors obviated robust external-internal predictors for scoliosis though simple or multiple linear regression (MLR). At the same time, constraining models of human physiology to linearities cannot capture the complexities of the relation. That reality suggested that methods allowing for multiple variables and nonlinear

links should be explored.

Jaremko et al. (2002a) took the lead in introducing learning systems to this research problem. They trained a joint genetic algorithm (GA)-artificial neural network (ANN) to relate a variety of trunk indices to Cobb angle. The GA exploited a survival of the fittest framework to select an input space from a larger candidate input space, while the ANN mimicked the nonlinear learning capacity of a system of neurons. That approach yielded a strong  $r^2 = 0.87$ , an improvement when compared with MLR, but lacked robustness with a 14–19% spread between training and test sets. That overfitting stemmed from working with a very small number of samples with patient repetition. The current paper improves and expands on the approach of Jaremko et al. (2002a) with innovative AI strategies for the identification of complex relations for which limited and interdependent examples were available. Two aspects are studied: the extraction of inputs and outputs, and the choice of AI methods. Firstly, the Cobb angle does not completely characterize the dynamic of spinal curvature in scoliosis, as no scalar measure could. At the same time, the selection of a finite set of trunk indices was prone to the exclusion of critical information. This research modeled external and internal discrete data as surfaces and curves using a functional paradigm. The coefficients thereof modeled available data to the fullest extent, and, as global measures of asymmetry, presented lower error. Being comprehensive, the need to differentiate between effective and poor predictors eliminated the need for the GA. Dimensionality reduction was achieved through principal components analysis (PCA), which constructs new variables as linear combinations of old ones, a capability that the GA does not possess.

Secondly, most learning systems, including ANNs, minimize empirical risk (or error within the learning set) with no further attempt to minimize risk in a more general context. The ANN is prone to returning solutions that are locally but not globally optimal. A support vector regression (SVR) machine returns the globally extremal

solution, subject to minimizing both empirical risk and model complexity, a property referred to as structural risk. Further, SVR requires fewer *a priori*-determined parameters than ANN. These aspects of SVR make it a more generalizable tool, permitting more robust prediction despite a small number of learning samples. As a final advantage of PCA-SVR over GA-ANN, the former is not iterative, resulting in much lower optimization times.

The approaches outlined above were selected with the objective of identifying a robust relation between comprehensive external and internal features, building on a positive preliminary study (Bergeron et al., 2004). Fig. 2.2 summarizes the objective of this study.

The remainder of this paper is organized as follows. Sections 2.2.2 and 2.2.3 discuss the methods used in this study: principal components analysis and support vector regression. Sections 2.2.4 and 2.2.5 describes the acquisition of data. Sections 2.2.6 and 2.2.7 apply a functional data analysis to trunk and spine data. Section 2.2.8 reports results obtained from the learning system developed in this paper, and Section 2.2.9 discusses those results. Finally, Section 2.2.10 provides a conclusion.

### **2.2.2 Principal components analysis**

PCA linearly transforms a set of variables such that the new ones, called principal components, are both uncorrelated and ranked by variation present in the original variables (Jolliffe, 2002). That transform is commonly used to analyze patterns in data. In biomedical engineering, PCA has been used to relate femoral bone shape to arthritis (Ramsay and Silverman, 2001), to analyze growth patterns in adolescence (Ramsay and Silverman, 2001), to detect the boundary of the optic disk of the eye (Li and Chutatape, 2004), and to model the shape of the anterior

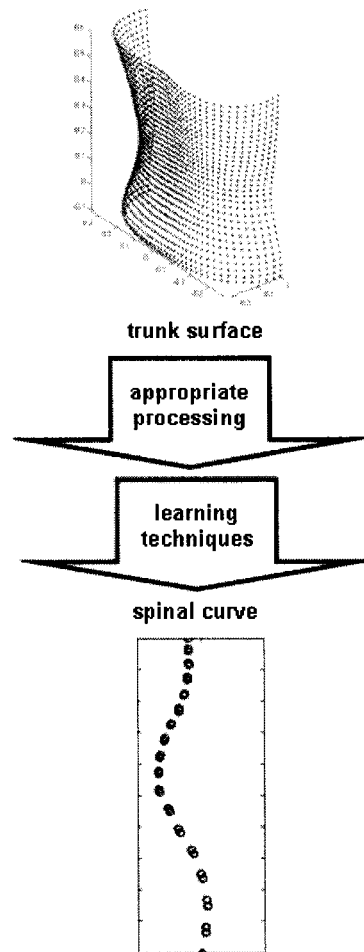


Figure 2.2 Study objective flow chart. This study identified a strong, robust relation for scoliotic spinal curve prediction from the trunk surface, through appropriate processing of data and learning techniques design.

cruciate ligament during surgery (Fleute et al., 1999). Under PCA, the relative variability explained by the  $m$ th component is

$$\frac{\lambda_m}{\sum_{k=1}^K \lambda_k} \quad (2.1)$$

where  $\lambda_k$  are the singular values of a  $K$ -feature covariance matrix  $\Sigma$  estimated from data and whose singular value decomposition is  $V\Lambda V^T$  where  $\Lambda$  is a diagonal matrix containing the singular values in decreasing order of magnitude and  $V$  contains the eigenvectors thereof (Jolliffe, 2002).

PCA may be extended to functional data, a method referred to as functional PCA or fPCA. The simplest implementation is achieved when functional data are represented using a complete orthonormal set, such as any Fourier basis; in such cases, fPCA reduces to PCA applied directly to the coefficients (Ramsay and Silverman, 1997). In theory, fPCA is insensitive to the choice of Fourier basis.

### 2.2.3 Support vector regression

Most learning systems, including ANNs, minimize empirical risk (or error within the learning set) with no further attempt to minimize risk in a more general context. SVR incorporates the learning theory of Vapnik and Chervonenkis to trade-off model complexity with empirical risk. The new measure is referred to as structural risk (Vapnik, 1998) and permits for more robust learning from fewer learning samples. Those strengths were exploited in applications such as the determination of erythropoietin dosage in secondary anemia patients (Martín-Guerrero et al., 2003). Let  $\{(\mathbf{u}_\ell, v_\ell)\}$  be a set of  $L$  known learning examples of inputs  $\mathbf{u}_\ell$  and outputs  $v_\ell$ .

Under SVR, a tube of radius  $\epsilon$  is fitted to the regression

$$f(\mathbf{u}) = (\mathbf{w} \cdot \mathbf{u}) + b \quad (2.2)$$

such that

$$v = f(\mathbf{u}) + R^{emp}$$

where  $\mathbf{w}$  is a vector of weights,  $b$  is the bias, and  $R^{emp}$  is the empirical risk. The latter is evaluated as follows: learning examples falling outside of the tube are penalized according to a linear rule (Fig. 2.3)

$$|x|_\epsilon = \max\{0, |x| - \epsilon\}. \quad (2.3)$$

Empirical risk counterbalances model complexity through *a priori*-established parameter  $C$  called the upper bound, and the resulting problem is formulated as an optimization problem with decision rule

$$\min \frac{1}{2} \|\mathbf{w}\| + C \frac{1}{L} \sum_{\ell=1}^L |v_\ell - f(\mathbf{u}_\ell)|_\epsilon \quad (2.4)$$

for which the first term measures model complexity and the second term is the empirical risk.

This optimization problem is relaxed by Lagrange multipliers, and applying the Karush-Kuhn-Tucker complementary conditions of optimization theory produces a sparse and dual representation problem that is quadratic and convex (possessing a unique local minimum that is global) (Cristianini and Shawe-Taylor, 2000). Eq. (2.2) is thus re-expressed as

$$f(\mathbf{u}) = \sum_{\ell=1}^L \alpha_\ell K(\mathbf{u}, \mathbf{u}_\ell) + \alpha_0. \quad (2.5)$$

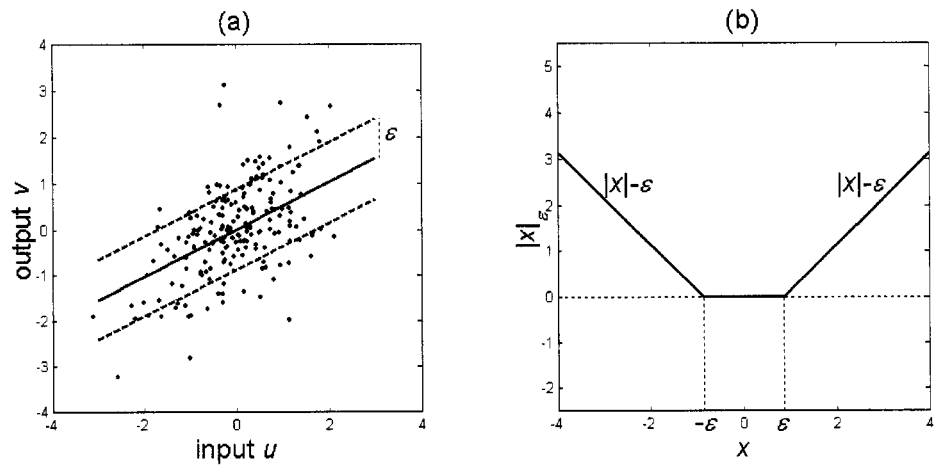


Figure 2.3 Definition of SVR insensitivity parameter  $\epsilon$ . (a) Consider a single-variable linear regression (solid line) of data (points). In determining the optimal regression, empirical errors for data within the insensitive band (bounded by the dashed lines) are not considered. Data falling outside are penalized linearly. The vertical half-width of this tube is defined by  $\epsilon$ . (b) The  $\epsilon$ -insensitive function  $|\cdot|_\epsilon$  linearly penalizes learning data falling outside of the insensitivity band.

Nonlinearities are achieved by the kernel function  $K(\cdot, \cdot)$ , which implicitly maps into a space of (usually) higher dimensionality. The polynomial family of kernels is based on integer powers of the canonical dot product

$$K(\mathbf{u}, \mathbf{u}') = (1 + (\mathbf{u} \cdot \mathbf{u}'))^p, \quad p = 1, 2, \dots \quad (2.6)$$

The case  $p = 1$  is called the linear kernel, the case  $p = 2$  is the quadratic kernel, and so on. Another family of kernels are the radius basis functions (RBF), based on a Euclidean distance metric

$$K(\mathbf{u}, \mathbf{u}') = \exp(-\sigma \|\mathbf{u} - \mathbf{u}'\|^q), \quad q = 1, 2, \dots \quad (2.7)$$

The case  $q = 1$  yields the exponential kernel, while  $q = 2$  generates the Gaussian kernel. Critical choices are the values for SVR parameters  $\epsilon$  and  $C$ , the kernel function, and associated kernel parameters. These are optimized against a validation set. System performance is then evaluated using a test set. A measure of regression performance across a set is the coefficient of determination

$$r^2 = 1 - \frac{\langle f(\mathbf{u}) - v, f(\mathbf{u}) - v \rangle}{\langle v - \bar{v}, v - \bar{v} \rangle} \quad (2.8)$$

where  $\bar{v}$  is the mean value of vector  $v$  of known outputs and  $f(\mathbf{u})$  is the vector of predicted values. A robust quality is sought of an SVR machine, in that regression capacity in the validation set translates well to a test set. Criteria for parameter optimization maximizes validation set  $r^2$  and minimizes the absolute differential between validation and learning set coefficients of determination. Accordingly, the current paper proposes the following decision rule:

$$\max_{\text{parameters}} |r_{\text{validation}}^2 - r_{\text{learning}}^2|. \quad (2.9)$$

As a special case of interest, when empirical error is penalised using a quadratic rule with no insensitivity, an SVR machine trained with a linear kernel, Eq. (2.6) with  $p = 1$ , tends towards the MLR solution as  $C \rightarrow \infty$ .

#### 2.2.4 Spine data acquisition and registration

Data were acquired at the Alberta Children’s Hospital, and consists of 149 samples with repetition at a minimal frequency of 6 months. That database was the same as used by Jaremko et al. (2002a) and Bergeron et al. (2004), but enlarged as new samples became available. Trunk and spine data were quasi-simultaneously acquired from scoliosis patients. The acquisition protocol was described by Poncet et al. (2000) and is summarized in the current and next sections.

Two PA (from the back) full-spinal X-rays were acquired from scoliosis patients. Those views were obtained with the X-ray tube positioned horizontally (PA-0) and at an angle of  $20^\circ$  from the horizontal (PA-20). On each X-ray, a trained technician identified six anatomical landmarks per thoracic and lumbar vertebra. These landmarks, which are geometrical features of the vertebrae that are anatomically relevant, were the upper and lower endplate centroids, and left and right upper and lower pedicle bases. Calibration markers were also identified on each X-ray. Acquisitions were excluded for which the landmark identification process was incomplete. Calibration markers allowed for 3D reconstruction of the landmarks, using a direct linear transformation (DLT). For each anatomical datapoint, the error distribution associated with the reconstruction process in the anterior plane was known (Poncet et al., 2000) to be Gaussian with a standard deviation of 2.2 millimetres, mostly accounted for by the anatomical landmark identification process. That statistic was invariant across anatomical landmarks and acquisition years. However, in the lateral plane, reconstructions were inaccurate as no lateral X-ray was acquired. Thus,

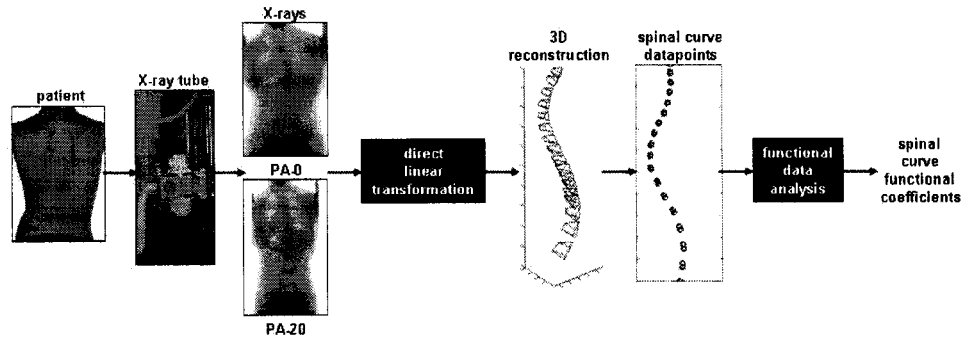


Figure 2.4 Spine data acquisition and processing. Two full spinal X-rays were acquired. On each X-ray, 6 anatomical landmarks on each of 17 vertebrae were identified by a trained technician. These landmarks were 3D-reconstructed using the direct linear transformation. Spinal reconstructions are represented as 17 hexahedrons. The anterior spinal curve defined by the upper and lower endplate centroids was retained. Data underwent functional data analysis to extract coefficients.

the 34 points  $(y, z)$  containing the positions of all endplate centroids were retained to define the spinal curve anterior view. Fig. 2.4 schematically summarises these steps.

### 2.2.5 Trunk data acquisition and registration

Three-dimensional trunk surface coordinates were obtained from each of four *BIRIS* laser scanners mounted on a mobile ring and connected to a computer camera-control and data acquisition system (Poncet et al., 2000) (Fig. 2.5).

A positioning system ensured standard posture. The views were merged into a single 3D dataset through rigid registration computed using calibration points located in the positioning system; the accuracy was  $4.2 \pm 1.7$  millimetres (Tardif et al., 2000). As the scan took approximately 15 seconds, subject movement accounted for an additional tolerated error of up to 1.5 millimetres (Poncet et al.,

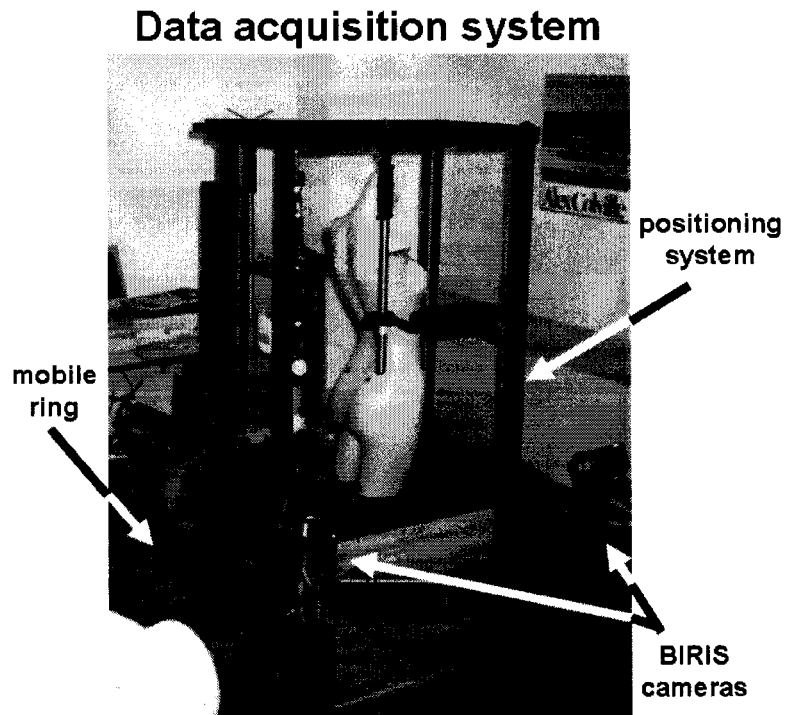


Figure 2.5 Data acquisition system. Three-dimensional trunk surface coordinates were obtained from each of four *BIRIS* laser scanners mounted on a mobile ring and connected to a computer camera-control and data acquisition system. A positioning system ensured standard posture. X-rays were quasi-simultaneously acquired as the patient remained in the same position, using a portable X-ray tube.

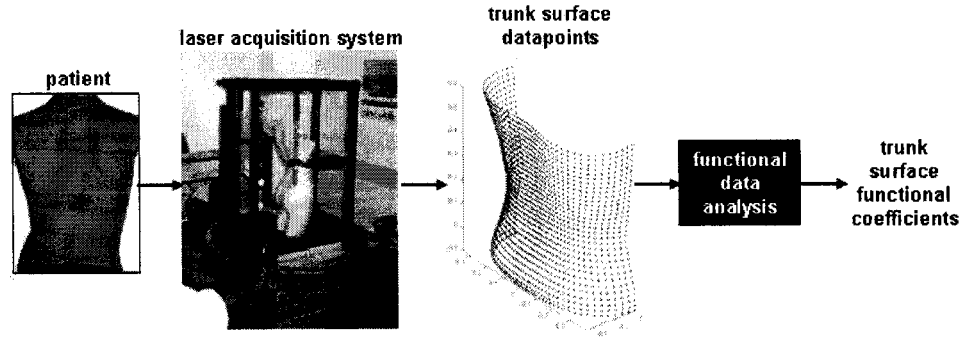


Figure 2.6 Trunk data acquisition and processing. A 3D laser scan of the trunk was acquired quasi-simultaneously to the X-rays. Datapoints of the trunk surface were returned by the scanner, and data underwent functional data analysis to extract coefficients thereof.

2000). Arms were removed using a cylinder, and trunk data were then smoothed using B-splines. Fig. 2.6 explains trunk data acquisition and processing. The result was a dataset consisting of a series of horizontal cross-sectional contours, each containing up to 360 3D coordinates and vertically separated by approximately 10-millimetre intervals.

The database was divided into specific sets (Table 2.1). To simulate a real-world medical implementation, in which previously acquired data were used in the prediction of current queries, the sets were defined according to acquisition date. The general invariance of selected clinical statistics across the sets sustained that decision.

Registration processes data, providing a similar basis for comparison. To begin, 9 surface markers appearing in both spinal and trunk datasets were used to translate those to a common 3D Euclidean space, having as origin the lower endplate centroid of the *L5* vertebra. The *x*-axis was oriented towards the front of the patient, the *y*-axis, towards the patient's left, and the *z*-axis, upwards. That space was scaled to

Table 2.1 Statistics for selected clinical and spinal indices

| set                                                    | learning        | validation      | test           |
|--------------------------------------------------------|-----------------|-----------------|----------------|
| acquisition years                                      | 1998–1999       | 2000            | 2001           |
| number of samples                                      | 72              | 45              | 30             |
| number of distinct patients                            | 36              | 30              | 20             |
| number of samples<br>from new patients                 | —               | 8               | 5              |
| female : male ratio                                    | 57:15           | 39:6            | 27:3           |
| adolescent idiopathic :<br>juvenile idiopathic ratio   | 48:24           | 28:17           | 18:12          |
| age (mean years $\pm$ SD)                              | 12.7 $\pm$ 3.0  | 12.8 $\pm$ 2.2  | 12.6 $\pm$ 1.7 |
| maximal anterior Cobb angle<br>(mean degrees $\pm$ SD) | 34.2 $\pm$ 14.5 | 35.2 $\pm$ 15.1 | 28.4 $\pm$ 7.7 |

unit spinal vertical span, after which the  $T1$  upper endplate centroid was rotated about the origin to point  $(0, 0, 1)$ . In clinical terms, that equated to removing the compensation. Not only did that permit the decorrelation of data orientation and shape, it ensured that the trunk surrounded the  $z$ -axis over the span of the spine. Spinal curves were represented explicitly as the function  $y = y(z)$  such that  $y : [0, 1] \rightarrow \mathfrak{R}$  where  $\mathfrak{R}$  stands for the space of real numbers. The trunk surface was represented explicitly in cylindrical coordinates as the function  $r = r(\theta, z)$  such that  $r : [\frac{\pi}{2}, \frac{3\pi}{2}] \times [-0.1, 0.6]$ . Only a  $180^\circ$  span of the lower back was retained (see Section 2.2.8). Finally, mean functions  $\bar{y}(z)$  and  $\bar{r}(\theta, z)$  were determined, with care taken to verify that those were smooth.

That registration was followed by a functional data analysis (FDA), described in following sections. The goal of FDA was to identify a compact representation for the function spaces  $\mathcal{Y} = \{y(z)\}$  and  $\mathcal{R} = \{r(\theta, z)\}$ , being subsets of the space of smooth functions.

### 2.2.6 Spinal functional data analysis

Here, the identification of a spinal functional representation of low dimensionality is detailed. Fourier series representation (FSR) exploits orthogonal functions under a least squares criterion of fit to discrete data and constitutes a classic choice to approximate smooth functions (Rice, 1964). Convergence is rapid for smooth functions (Davis, 1963). Sinusoids are an ideal choice to represent functions that do not present strong local features (Ramsay and Silverman, 1997). Staib and Duncan (1992) and Székely et al. (1996) both used sinusoidal functions to define object boundaries for MRI segmentation. In scoliosis research, bases of sinusoids were previously used in algorithms for clinical index computation (Dansereau et al., 1993).

Having previously removed the compensation, the resulting boundary conditions suited the basis of sine functions. Accordingly, data were functionally represented as follows:

$$y(z) = \bar{y}(z) + \sum_{m=1}^M b_m \sin(\pi m z) + e_y(z) \quad (2.10)$$

where

$$b_m = 2 \int_0^1 y(z) \sin(\pi m z) dz$$

where  $b_m$  represent the Fourier coefficients and  $e_y$  an error term. Integration was computed using trapezoidal quadrature. Calculations were performed using Matlab (Mathworks, 2001). In theory, coefficients should quickly tend towards zero, and 34 coefficients should represent  $y(z)$  exactly. In practice, numerical instabilities appeared beyond the ninth coefficient. Thus,  $M = 9$ .

Sample Fourier coefficient covariance matrices were estimated from the learning set. Thus, PCA was used to discover a smaller functional basis, replacing the representation of Eq. (2.10) with

$$y(z) = \bar{y}(z) + \sum_{m=1}^{M'} \beta_m y_m(z) + \epsilon_y(z) + e_y(z) \quad (2.11)$$

where  $\beta_m$  are functional PCA coefficients,  $y_m$  are the PCA functions, subscripts refer to functional component numbers, and  $\epsilon_y(z)$  an error term for a model of size  $M' < M$ . Table 2.2 reports for each component the relative variance (2.1). The modeling error was evaluated across the learning set using the  $L_2$ -norm

$$\|\epsilon_y(z)\| = \sqrt{\int_0^1 \epsilon_y(z)^2 dz}, \quad (2.12)$$

where numerical integration was invoked once again. Statistics of Table 2.2 translated well to validation and test sets.

Table 2.2 Properties of spinal curve principal components

| component<br>number<br>$M'$ | relative<br>variance<br>(%) | cumulative modeling<br>$L_2$ -error<br>(mean mm $\pm$ SD) |
|-----------------------------|-----------------------------|-----------------------------------------------------------|
| 1                           | 67                          | $7.99 \pm 5.61$                                           |
| 2                           | 29                          | $2.63 \pm 1.73$                                           |
| 3                           | 3.0                         | $1.52 \pm 0.83$                                           |
| 4                           | 0.62                        | $1.00 \pm 0.45$                                           |
| 5                           | 0.23                        | $0.72 \pm 0.28$                                           |

The components (Table 2.2) defined space  $\mathcal{Y}$  that is a subset to the set of smooth functions. Fig. 2.7 presents that basis; curves represent a variability of 3 standard deviations across the learning set. Those components served as outputs to support vector regression. A model dimension of 3 was retained, matching modeling error with spinal data error.

### 2.2.7 Trunk functional data analysis

Here, FDA on trunk data is presented, with the goal of using function coefficients as inputs to predict spinal curve coefficients derived in the previous section. The looser, zero derivative boundary condition on  $r(\theta, z)$  was accepted; thus, a cosine basis was retained. Accordingly, data were functionally represented as follows:

$$r(\theta, z) = \bar{r}(\theta, z) + \sum_{i=1}^I \sum_{j=1}^J c_{i,j} \cos(i\theta) \cos\left(\frac{\pi j z}{0.7}\right) + e_r(z) \quad (2.13)$$

where  $c_{i,j}$  represent the approximation coefficients and  $e_r(\theta, z)$  an error term. Trunk data had been interpolated to a  $48 \times 32$  regular grid: piecewise cubic Hermite interpolating polynomials ensured monotonic interpolation between data points in the angular orientation, and cubic splines were suitable in the vertical direction.

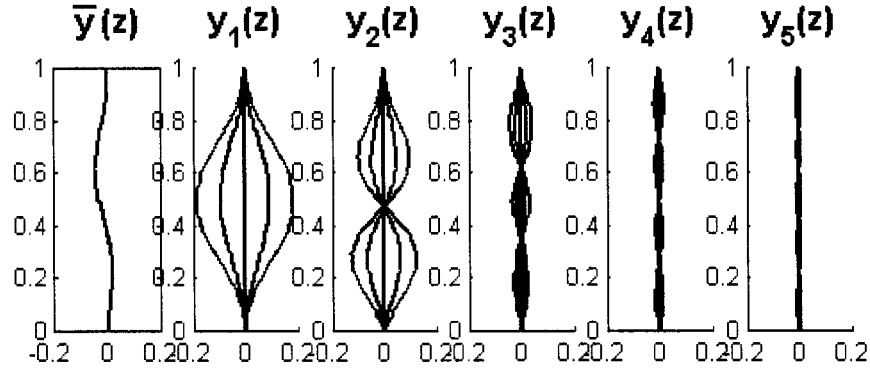


Figure 2.7 Spinal curve functional basis. The leftmost curve is the mean curve. Remaining curves are the functional principal components, illustrated to represent a variability of 3 standard deviations across the learning set. These functions define functional space  $\mathcal{Y}$  of anterior view spinal curves. Component 1 appears to account for overall spinal curvature, while component 2 adjusts the relative curve deviation between the thoracic and lumbar regions. Higher-order components are harder to interpret clinically.

Coefficients were computed as the 2D discrete cosine transform (DCT), an efficient algorithm. Implementing functional PCA resulted in the following representation:

$$r(\theta, z) = \bar{r}(\theta, z) + \sum_{n=1}^{N'} \gamma_n r_n(\theta, z) + \epsilon_r(\theta, z) + e_r(\theta, z) \quad (2.14)$$

where  $\gamma_n$  are functional PCA coefficients,  $r_n$  are the PCA functions, subscripts refer to functional component numbers, and  $\epsilon_r(\theta, z)$  an error term for a model of size  $N' < N = IJ$ . Table 2.3 reports leading components. The modeling error was determined using an expression analogous to (2.12).

Table 2.3 Properties of trunk surface principal components

| component<br>number<br>$N'$ | relative<br>variance<br>(%) | cumulative modeling<br>$L_2$ -error<br>(mean mm $\pm$ SD) |
|-----------------------------|-----------------------------|-----------------------------------------------------------|
| 1                           | 57                          | $7.32 \pm 3.06$                                           |
| 2                           | 23                          | $4.99 \pm 2.24$                                           |
| 3                           | 8.7                         | $3.60 \pm 1.73$                                           |
| 4                           | 3.9                         | $3.01 \pm 0.79$                                           |
| 5                           | 2.1                         | $2.55 \pm 0.65$                                           |

### 2.2.8 Results

An array of SVRs was trained to predict spinal coefficients from anterior basis coefficients (Fig. 2.8). The implementation by Gunn (1998) was used, whereby one SVR was assigned to each output, and outputs were normalized to distributions possessing zero mean and unit variance. A parameter space for SVR insensitivity  $\epsilon$  and upper bound  $C$  defined over intervals  $[0.005, 1]$  and  $[0.005, 2000]$  were found to be sufficiently large. As the magnitude of coefficients  $\gamma_n$  (2.14) decreased as  $n$  increased, the prediction was insensitive to the choice for  $N'$  beyond a small threshold. Fourteen inputs, for an inputs-to-learning set samples ratio of 1 : 5, worked well. For each output, prediction capability was evaluated across each set using the coefficient of determination (2.8).

For each output, optimal SVR parameters for the kernels (2.6,2.7) were determined using decision rule (2.9). Results for SVR prediction of spinal components for different kernel choices are compared (Tables 2.4-2.6).

Predictions for the three leading spinal components were used to generate the predicted spinal curves. Results thereof are presented as Table 2.7. For comparison purposes, SVR prediction was contrasted with a maximum likelihood (ML) predic-

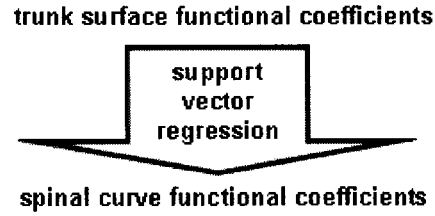


Figure 2.8 System flow chart. Leading trunk surface functional coefficients were used to predict those representing the spinal curve. To this end, learning set samples were used to train support vector regression machines. Model performance was assessed using a test set of previously unseen samples.

Table 2.4 SVR prediction of first spinal coefficient  $\beta_1$

| kernel                     | optimal<br>kernel<br>parameters | optimal<br>insensitivity<br>$\epsilon$ | optimal<br>upper bound<br>$C$ | learning<br>set<br>$r^2$ | validation<br>set<br>$r^2$ | test<br>set<br>$r^2$ |
|----------------------------|---------------------------------|----------------------------------------|-------------------------------|--------------------------|----------------------------|----------------------|
| linear<br>( $p = 1$ )      | —                               | 0.97                                   | 0.92                          | 0.70                     | 0.71                       | 0.70                 |
| quadratic<br>( $p = 2$ )   | —                               | 0.36                                   | 0.089                         | 0.86                     | 0.76                       | 0.71                 |
| exponential<br>( $q = 1$ ) | $\sigma = 530$                  | 0.82                                   | $\infty$                      | 0.67                     | 0.66                       | 0.63                 |
| Gaussian<br>( $q = 2$ )    | $\sigma = 890$                  | 0.96                                   | $\infty$                      | 0.70                     | 0.71                       | 0.70                 |

Table 2.5 SVR prediction of second spinal coefficient  $\beta_2$

| kernel                     | optimal<br>kernel<br>parameters | optimal<br>insensitivity<br>$\epsilon$ | optimal<br>upper bound<br>$C$ | learning<br>set<br>$r^2$ | validation<br>set<br>$r^2$ | test<br>set<br>$r^2$ |
|----------------------------|---------------------------------|----------------------------------------|-------------------------------|--------------------------|----------------------------|----------------------|
| linear<br>( $p = 1$ )      | —                               | 0.093                                  | 50                            | 0.77                     | 0.77                       | 0.82                 |
| quadratic<br>( $p = 2$ )   | —                               | 0                                      | 1.0                           | 0.96                     | 0.84                       | 0.83                 |
| exponential<br>( $q = 1$ ) | $\sigma = 250$                  | 0                                      | $\infty$                      | 1                        | 0.62                       | 0.72                 |
| Gaussian<br>( $q = 2$ )    | $\sigma = 34$                   | 0.29                                   | $\infty$                      | 0.93                     | 0.89                       | 0.75                 |

Table 2.6 SVR prediction of third spinal coefficient  $\beta_3$ 

| kernel                     | optimal<br>kernel<br>parameters | optimal<br>insensitivity<br>$\epsilon$ | optimal<br>upper bound<br>$C$ | learning<br>set<br>$r^2$ | validation<br>set<br>$r^2$ | test<br>set<br>$r^2$ |
|----------------------------|---------------------------------|----------------------------------------|-------------------------------|--------------------------|----------------------------|----------------------|
| linear<br>( $p = 1$ )      | —                               | 1.0                                    | 0.86                          | 0.03                     | 0.04                       | 0.04                 |
| quadratic<br>( $p = 2$ )   | —                               | 1.1                                    | 0.13                          | 0.21                     | 0.14                       | 0.09                 |
| exponential<br>( $q = 1$ ) | $\sigma = 1100$                 | 1.4                                    | $\infty$                      | 0.05                     | 0.06                       | -0.05                |
| Gaussian<br>( $q = 2$ )    | $\sigma = 51$                   | 1.3                                    | $\infty$                      | 0.17                     | 0.16                       | 0.05                 |

Table 2.7 Prediction of spinal curve using support vector regression

| set        | support vector regression<br>prediction $L_2$ -error<br>(mean mm $\pm$ SD) | maximum-likelihood<br>prediction $L_2$ -error<br>(mean mm $\pm$ SD) |
|------------|----------------------------------------------------------------------------|---------------------------------------------------------------------|
| learning   | $4.50 \pm 2.74$                                                            | $8.12 \pm 4.67$                                                     |
| validation | $4.10 \pm 2.81$                                                            | $8.67 \pm 4.76$                                                     |
| test       | $3.61 \pm 1.91$                                                            | $6.83 \pm 3.15$                                                     |

tion (mean spinal curve of Fig. 2.7).

A qualitative comparison is presented (Fig. 2.9) between acquired and predicted spines across a representative subset of the test set-graphed in order of increasing SVR prediction  $L_2$ -error.

### 2.2.9 Discussion

Data revealed a strong support vector regression capability for the two leading spinal coefficients that constitute the spinal curve model (Tables 2.4 and 2.5), with coefficients of determination 0.70 and 0.82, respectively. When compared with  $r^2$ 's reported from the linear regressions reported in Section 2.2.1, these correlations were strong. Errors were attributed to unexplained variables including gender,

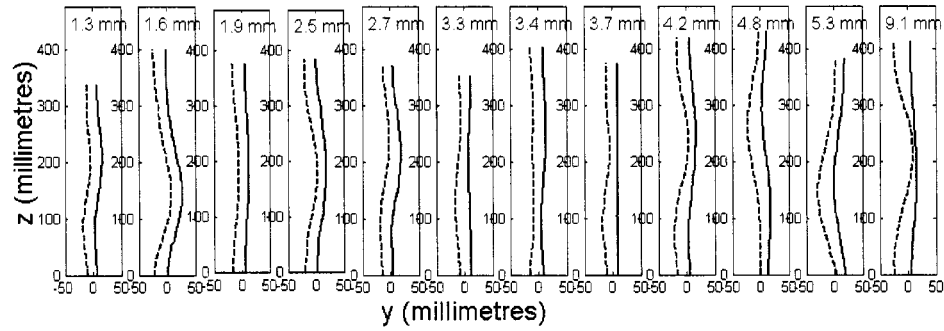


Figure 2.9 Comparison of spinal data with SVR-predicted spinal curves. X-ray-reconstructed spinal curves are dashed, while SVR-predicted ones are solid and shifted right. Test set samples were sorted in order of increasing prediction  $L_2$ -error; roughly every third sample is displayed. Numerical values thereof are indicated.

body morphology type, and stage of physical development. More importantly, validation set statistics extended very well to the test set, a robust prediction capability over the results from the GA-ANN approach. The current study reported a first strong and robust prediction from learning methods in this application area. With respect to spinal component 3, a weak yet robust predictive capability was found (Table 2.6). For higher-order components, negligible or ungeneralizable regressions were found. These observations sustained the decision to retain three components to define the spinal model used in this study.

The strength of this methodology was the achievement of a robust regression between externality and internality. The functional data analysis was very effective at implicitly modeling the nonlinearities between trunk and spine, allowing for a relation to be found without resorting to nonlinear kernels that induced overfitting as available data were available. Training attempts with the quadratic, exponential and Gaussian kernels resulted in the self-imposed constraint of robust prediction returning systems presenting weaker generalization capabilities. However, it is reasonable to suggest that slight nonlinearities exist between the inputs and outputs;

the SVR framework would permit those more subtle relations to be found as the size of the learning set increased over time. A larger learning set size might also have permitted for the entire trunk to participate in the prediction, as opposed to a  $180^\circ$  span of the lower back. Indeed, the back contributed quality information for spinal prediction, whereas the front, separated from the spine by the rib cage and varying levels of fat, presented higher levels of potential physiological noise. That does not, however, imply that the front did not contain distinct information that would improve prediction capability. As the number of inputs to SVR was constrained by an overfitting consideration, increasing the surface area represented by those inputs resulted in a model in which the back was represented less accurately—a loss for which the front could not compensate. Other sources of physiological noise were the breasts in female patients and the holes remaining from arm removal.

Spinal reconstructions from SVR-predicted coefficients yielded mean  $L_2$ -errors of 4.50, 4.10 and 3.61 millimetres across the three sets (Table 2.7). The higher error observed in the validation set was consistent with the overall stronger spinal deviations, as verified by the ML prediction error of Table 2.7 and the Cobb angle statistics (Table 2.1). In all sets, when SVR performance was compared to the ML prediction, the  $L_2$ -error was reduced by a factor slightly greater than 2. Those prediction errors fared well when compared to combined trunk acquisition errors of 4.7 millimetres and the spine reconstruction error of 2.2 millimetres.

The database used in this study contained samples from repeated patients. While those samples were considered to be distinct, based on growth and scoliosis progression considerations, they were not independent. Redundancy was circumvented in available data by assigning as few as 48% of available samples to learning. Incidentally, prediction of spinal curves for new patients in the validation and test sets presented comparable predictions with  $L_2$ -errors of 3.40 and 3.93 millimetres,

respectively—results that were statistically insignificant given the small sample sizes. While a database of distinct samples would have been more representative, scoliosis clinical protocols would most certainly have required one set of X-rays to be diagnostic. Thus, set definition was in-line with the clinical intentions of the current study, in which earlier-acquired data were used to make predictions about new data.

#### **2.2.10 Conclusion**

The purpose of this study was to investigate the extent to which prediction of the shape of spinal deformity in scoliosis patients based on trunk shape asymmetries could be accurately performed, with the aim of finding a stronger and more robust predictive capability than had been reported in previous work. The biological complexity of the sought relation was high. However, the limited number of available data imposed the reliance on the learning system to identify the nonlinearities between inputs and outputs. The novelty of this study was the joint fPCA-SVR approach in which a functional data analysis on the trunk surface and the spinal curve were performed, resulting in a relatively small number of comprehensive coefficients that were linked using the learning technique. The strength of the approach was found in the trunk and spine coefficients that effectively modeled the nonlinearities and contained much of the complexity in data. Hence, support vector regression was effective using a linear kernel, avoiding the inevitable overfitting that can result with unprocessed data in concert with a higher-order kernel. Results were significant to two reasons. First, this is the first robust predictive capability to be reported in this research application to date. Second, that predictive capability was strong, comprehensive, and more highly innovative than past results.

The current results are encouraging, but an improvement in the predictive strength

of the system would be required prior to its implementation in a clinical setting. That is particularly evident given the few cases (Fig. 2.9) in which the SVR-prediction is poor. Nonetheless, this approach shows strong potential eventual X-ray replacement. A clinical study would constitute the best manner in which to introduce the technology to clinical staff, allowing for an ongoing clinical validation process. At the same time, the acquisition of further samples would add to the learning system's knowledge and thus improve its prediction. Following clinical verification, the authors envision that this system could reduce the need for scoliosis patient exposure to X-rays beyond a first diagnostic acquisition.

## CHAPITRE 3

### PRÉDICTION TRIDIMENSIONNELLE AMÉLIORÉE

Ce chapitre contient un second article dont le titre est *Noninvasive assessment of scoliotic spinal curve using robust learning techniques*, soumis à la publication *Medical Image Analysis* le 14 octobre 2004. Les coauteurs sont Farida Cheriet (École Polytechnique de Montréal et Hôpital Sainte-Justine), Janet Ronsky (Université de Calgary), Ronald Zernicke (Université de Calgary) et Hubert Labelle (Hôpital Sainte-Justine), et la citation est (Bergeron et al., soumis 2004). La contribution relative du premier auteur à cet article est estimée à 90%.

#### 3.1 Synthèse

L'article du chapitre précédent présente une première régression robuste utilisant des techniques d'apprentissage, et une première utilisation d'indices externes et internes qui sont compréhensives de leur géométrie, soient des coefficients de fonctions orthonormales. Toutefois, le niveau de bruit dans chaque composante fonctionnelle n'est pas quantifié, et la modélisation de la courbe rachidienne dans le plan latéral n'a pas lieu, puisque les données sont plus bruitées dans ce plan.

Ce nouvel article étend la méthode par la représentation par séries de fonctions Haar et une généralisation à la méthode des fractions de bruit minimales fonctionnelles. C'est cette généralisation, de concert avec une base de données vertébrales indépendante, qui permet d'étendre la prédiction au plan latéral de la courbe de la colonne vertébrale, permettant une modélisation 3D. De plus, l'implémentation de la validation croisée permet d'augmenter la confiance envers les résultats présentés

Tableau 3.1 Comparaison des deux articles

| article                                    | 1                                                                                                       | 2                                                                                        |
|--------------------------------------------|---------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| titre                                      | <i>Prediction of anterior scoliotic spinal curve from trunk surface using support vector regression</i> | <i>Noninvasive assessment of scoliotic spinal curve using robust learning techniques</i> |
| revue                                      | <i>Engineering Applications of Artificial Intelligence</i>                                              | <i>Medical Image Analysis</i>                                                            |
| nombre d'échantillons                      | 147                                                                                                     | 196                                                                                      |
| vues rachidiennes considérées              | frontale seulement                                                                                      | latérale et frontale                                                                     |
| méthode d'analyse fonctionnelle de données | analyse par composantes principales fonctionnelles                                                      | transformée des fractions fonctionnelles de bruit minimal                                |
| choix des paramètres SVR                   | optimisation contre le sous-ensemble de validation                                                      | résultats et arguments théoriques                                                        |
| assurance de robustesse                    | sous-ensembles d'apprentissage (1998-1999), de validation (2000) et de test (2001)                      | validation croisée avec cinq sous-ensembles (1998, 1999, 2000, 2001 et 2002)             |
| évaluation des prédictions                 | erreur- $L_2$                                                                                           | coefficient de détermination fonctionnel $R^2$                                           |

pour le système proposé, puisque celui-ci est validé sur un ensemble de test étendu. Finalement, les données acquises pendant l'année 2002 devenues disponibles, celles-ci sont rajoutées à la base de données, représentant l'addition d'une cinquantaine d'échantillons (soit une augmentation du tiers). Le schéma de la figure 3.1 présente une vue d'ensemble sur la méthodologie de cet article, tandis que le tableau 3.1 fait ressortir les différences entre les deux articles présentés dans ce mémoire. Le modèle retenu par cette formulation permet d'expliquer plus de 60% de la variabilité de la courbe rachidienne à partir seulement de la topographie du tronc. Cette capacité prédictive est augmentée à plus de 72% dans la section Addendum de ce chapitre, section qui présente une mise-à-jour des meilleurs résultats obtenus.

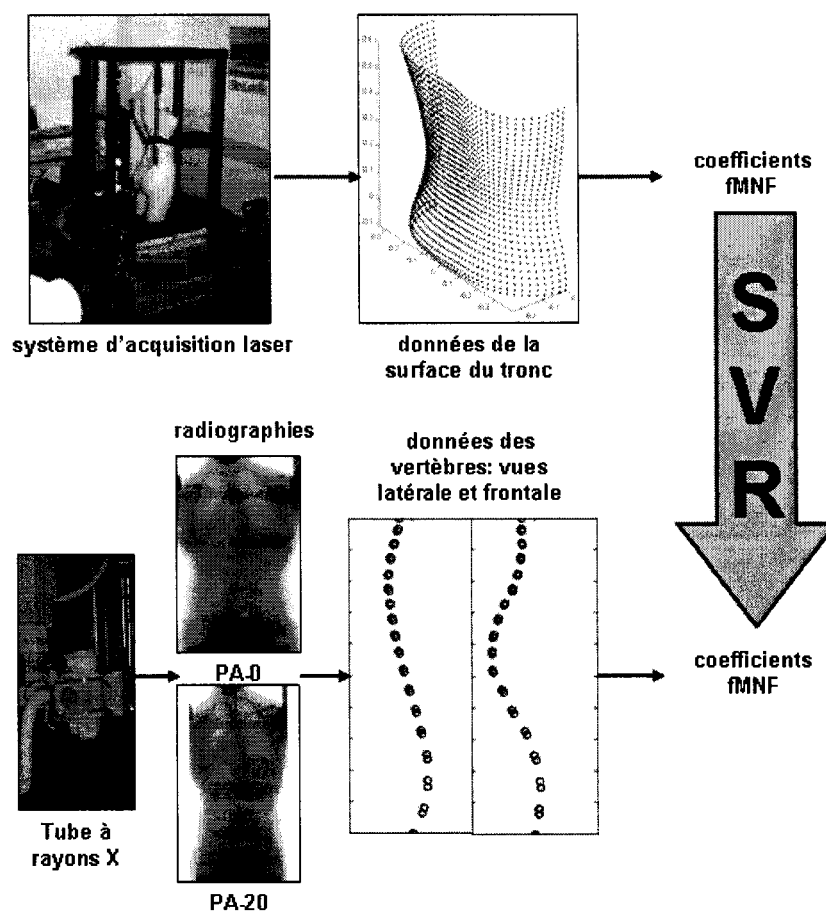


Figure 3.1 Ce schéma présente les liens entre les données et les méthodes de cet article. Cette figure présente deux différences par rapport à la figure 2.1. En premier lieu, les deux radiographies sont reconstruites en 3D, permettant de travailler sur les vues frontale et latérale de la courbe de la colonne vertébrale. En second lieu, c'est une transformée par fractions fonctionnelles de bruit minimal qui est utilisée pour extraire des coefficients représentatifs de la forme de la surface du tronc et des courbes rachidiennes. Ces coefficients fMNF externes et internes sont ensuite utilisés pour entraîner SVR.

## 3.2 Noninvasive assessment of scoliotic spinal curve using robust learning techniques

### 3.2.1 Introduction

Scoliosis is a complex three-dimensional (3D) deformation of the spine, involving the axial rotation of vertebrae, curvature in the anterior view, and alterations to the sagittal view. Resulting in primary and secondary health problems, this medical condition affects up to 4% of the population. The cause of idiopathic scoliosis is unknown, manifesting itself during infancy ( $< 3$  years,  $< 1\%$  of cases), childhood (3–10 years, 12–21% of cases) or adolescence ( $> 10$  years) (Reamy and Slakey), 2001). A smaller number of instances are genetic in nature, referred to as congenital. The main determinants of progression of this medical condition are gender (females have a risk of curve progression 10 times higher than males), growth potential, curve apex height, and curve magnitude (Reamy and Slakey), 2001; Greiner, 2002). The latter two are evaluated from a full spinal X-ray. Given the variability of adolescent growth on an annual (Tuddenham and Snider, 1954) and even daily (Thalange et al., 1996) basis, frequent follow-up is necessary. However, the diagnosis and monitoring of scoliosis remain controversial. For example, Reamy and Slakey) (2001) recommend X-ray acquisition every 6 months until skeletal maturity. By contrast, Greiner (2002) only considers X-rays for patients presenting a curve severe enough to warrant treatment or believed to progress in that direction. Either way, the limiting factor is the risk associated with serial exposure to harmful X-radiation. Modern equipment has reduced the dosage, and postero-anterior (PA) views have reduced its impact on sensitive tissues, yet the incidence of cancer of the breast, thyroid, lung, ovary and bone marrow increases by up to 2.4 per thousand patients-comparable to the probability of developing thyroid cancer in the general population (Levy et al., 1996). Given that fewer than 1 in 10 scoliosis patients

require treatment beyond periodic monitoring (Reamy and Slakey, 2001). it is desirable to reduce radiation exposure as much as possible. Thus, an alternate and less invasive technique for scoliosis follow-up is needed.

Magnetic resonance imaging (MRI) technology is a possible alternative. However, its use is limited to the most severe cases due to high cost and low availability (Carro, 1999). Further, MRI is less effective than X-radiation for distinguishing bone from other tissues. As an additional problem, MRI is acquired while supine, a position in which scoliosis deformities are relaxed.

Another possibility combines trunk surface scanning technology with a suitable mathematical model to relate external asymmetry to internal deformity. This approach presupposes that the underlying physiological links are sufficiently consistent among scoliosis patients to permit for a relation to be identified. The classical strategy has relied on simple or multiple linear regressions to relate trunk features to the maximal anterior Cobb angle, a traditional clinical measure of spinal curvature measured directly from an X-ray. Results are highly variable between studies. For example, coefficients of determination  $r^2$  for the maximum back surface rotation predictor have been reported over the range 0.17–0.42 (Korovessis and Stamatakis, 1996; Samuelsson and Noren, 1997; Jaremko et al., 2002a). Another example is found in the quasi-Cobb angle, analogously measured from the spinous process curve on the trunk surface, with values for  $r^2$  covering the interval 0.27–0.66 (Goldberg et al., 2001; Jaremko et al., 2002a).

Jaremko et al. (2002a) introduced artificial intelligence (AI) to this application area and found a nonlinear regression from an input set of trunk features that explained 87% of Cobb angle variability, using a hybrid genetic algorithm (GA)-artificial neural network (ANN). However, results lacked robustness with a 14–19% spread between training and test set results. Difficulties included high levels of

measurement noise in inputs and outputs hindering the learning process, the risk that selected inputs did not encapsulate the useful information contained in the trunk scans, and working with a relatively small number of training examples containing interdependencies (due to repeated patient samples). Bergeron et al. (2004) and Bergeron et al. (accepté 2005) remedied these weaknesses by representing the trunk surface and anterior view spinal curve using functional principal components analysis (PCA) whose leading coefficients were comprehensive and noise-insensitive, and choosing support vector regression (SVR), a learning method that penalized model complexity. That approach yielded a first nonlinear and robust prediction in this application area. Further, the prediction was comprehensive and strong, with correlations of 0.85 and 0.80. This current paper addresses 2 weaknesses of that methodology, based on a positive preliminary study (Bergeron et al., 2005a).

First, spinal curve prediction in the sagittal view was difficult, as 2 PA X-rays were acquired (with the tube positioned horizontally (PA-0) and at an angle of  $20^\circ$  from the horizontal (PA-20)), yet no lateral (LAT) X-rays were. Lateral spinal configuration could be simulated as the PA-20 imagery contained some sagittal information, but those simulations were noisy. The resulting functional principal components (PCs) were noisy and lacked interpretability, hindering trunk-spine prediction. Under those circumstances, Ramsay and Silverman (1997) propose regularized PCA, a generalization that incorporates a smoothness constraint on the PCs. However, their method requires *a priori* determination of a smoothing parameter, a route that was undesirable given the small number of available samples. Rather, spinal PCs were learned from a second database containing better quality sagittal spinal information. Noise-insensitivity of retained PCs was ensured by the method of minimum noise fractions (MNF), a generalization of PCA.

Second, the implementation of this solution resulted in mild to moderate overfitting, in that a differential was observed between validation and test sets. Two

factors could explain this situation: high input dimensionality and/or learning set unrepresentativeness towards the modeled population. This paper discounts the former, and resorts to cross-validation (CV) to increase confidence in the generalization of testing results to address the latter.

This research expanded and improved the previously developed support vector regression learning system using minimum noise fractions and cross-validation to permit prediction of the spinal curve. The remainder of this paper is organized as follows. Section 3.2.2 discusses the methods used in this study. Section 3.2.3 describes data acquisition process. Section 3.2.4 reports and Section 3.2.5 analyzes the results. Finally, conclusions are presented in Section 3.2.6.

### **3.2.2 Methods**

This section describes three methods that constitute the approach of this research: minimum noise fractions (MNF), support vector regression (SVR), and functional model assessment.

#### **3.2.2.1 Minimum noise fractions**

Principal components analysis (PCA) linearly transforms a set of variables such that the new ones, called principal components, are both uncorrelated and ranked by variation present in the original variables (Jolliffe, 2002). That transform is a commonly used strategy to analyze patterns in data: in biomedical engineering, PCA was used to relate femoral bone shape to arthritis (Ramsay and Silverman,

2001). Consider  $K$  measured signals  $\{\mathbf{s}_k\}_{k=1}^K$ . Under PCA, signals obey the model

$$\mathbf{s} = \sum_{k=1}^K c_k \boldsymbol{\omega}_k \quad (3.1)$$

where  $\boldsymbol{\omega}_k$  are the principal components which correspond to the eigenvectors of the signal covariance matrix  $\Sigma_S = \text{Cov}\{\mathbf{s}\}$  estimated from the signals. The relative variability explained by the  $m$ th component is

$$\frac{\lambda_m}{\sum_{k=1}^K \lambda_k} \quad (3.2)$$

where  $\lambda_k$  are the latent values. When the measured signals  $\mathbf{z}_k = \mathbf{s}_k + \mathbf{n}_k$  are polluted with additive white noise with standard deviation (SD)  $\sigma_N$  such that  $\mathbf{s}_k$  and  $\mathbf{n}_k$  are decorrelated, the PCA model (3.1) becomes

$$\mathbf{z} = \sum_{k=1}^M c_k \boldsymbol{\omega}_k + \mathbf{n}$$

for which the noise covariance matrix is  $\Sigma_N = \text{Cov}\{\mathbf{n}\} = E\{\mathbf{n}^T \mathbf{n}\} = \sigma_N^2 I$  where  $I$  is the identity matrix. Exploiting the additive property of covariance,  $\Sigma_Z = \Sigma_S + \Sigma_N$ , and the  $M$  leading latent values of  $\Sigma_Z$  form a decreasing sequence and the remaining ones are equal (Hyvärinen et al., 2001):

$$\lambda_1 > \lambda_2 > \cdots > \lambda_M > \lambda_{M+1} = \lambda_{M+2} = \cdots = \lambda_K = \sigma_N^2 \quad (3.3)$$

The MNF transform generalizes these results, being a PCA in a non-Euclidean metric space defined by an estimated noise covariance matrix. Thereupon, MNF maximizes the signal-to-noise ratio (SNR) rather than overall variance. That method was proposed by Green et al. (1988), in the context of multispectral satellite image decomposition. Larsen and Hilger (2003) has since applied MNF to a biomedical

application: the annotation of lungs in chest X-rays. SNR is defined as the ratio of signal variance to noise variance:

$$SNR = \frac{Var\{\boldsymbol{\omega}^T \mathbf{s}_k\}}{Var\{\boldsymbol{\omega}^T \mathbf{n}_k\}} = \frac{\boldsymbol{\omega}^T \Sigma_S \boldsymbol{\omega}}{\boldsymbol{\omega}^T \Sigma_N \boldsymbol{\omega}} = \frac{\boldsymbol{\omega}^T \Sigma_Z \boldsymbol{\omega}}{\boldsymbol{\omega}^T \Sigma_N \boldsymbol{\omega}} - 1 \quad (3.4)$$

Therefore, the MNFs, or noise-adjusted components, are given by the generalized eigenvectors  $\boldsymbol{\omega}_k$  of  $\Sigma_Z$  with respect to  $\Sigma_N$  corresponding to singular values  $\lambda_k$ . By design, the  $m$ th MNF is the linear combination that exhibits the highest SNR subject to being uncorrelated to lower-order MNFs. The SNR of component  $m$  is

$$\lambda_m - 1 \quad (3.5)$$

As a special case of interest, when the noise covariance matrix is of the form  $\Sigma_N = \sigma_N^2 I$ , PCA and MNF components are the same, and their singular values are directly proportional. Hence, MNF singular values possess double significance.

PCA may be extended to functional data, a method referred to as functional PCA. The simplest implementation is achieved when functional data are represented using a complete orthonormal set, such as any wavelet basis; in such cases, functional PCA reduces to PCA applied directly to the coefficients thereof (Ramsay and Silverman, 1997). Functional MNF is similarly implemented.

### 3.2.2.2 Support vector regression

Most learning systems, including ANNs, minimize empirical risk (or error within the learning set) with no further attempt to minimize risk in a more general context. Support vector regression (SVR) incorporates the learning theory of Vapnik and Chervonenkis to trade-off model complexity with empirical risk (Vapnik, 1998). This new measure is referred to as structural risk and permits for more robust

learning from fewer learning samples. Let  $\{(\mathbf{u}_\ell, v_\ell)\}$  be a set of  $L$  learning examples of inputs  $\mathbf{u}_\ell \in \mathbb{R}^n$  and outputs  $v_\ell \in \mathbb{R}$  where  $\mathbb{R}$  stands for the space of real numbers. Inputs are (usually nonlinearly) mapped to  $\mathbf{w} = \psi(\mathbf{u})$  such that  $\psi : \mathbb{R}^n \rightarrow \mathcal{F}$  where  $\mathcal{F}$  is a Hilbert space of (usually) higher or infinite dimensionality with inner product  $\langle \cdot, \cdot \rangle$ . Then, a tube of radius  $\epsilon$  is linearly fitted to the regression

$$v = \langle \mathbf{w}, \psi(\mathbf{u}) \rangle + b + \text{error} \quad (3.6)$$

where  $\mathbf{w}$  contains the weights and  $b$  is the bias. Learning examples falling outside of the tube are penalized according to a linear rule, and trade-off between this error and model complexity is achieved through *a priori*-established parameter  $C$  called the regularization parameter. The resulting optimization problem is relaxed by Lagrange multipliers, and applying the Karush-Kuhn-Tucker complementary conditions of optimization theory produces a sparse and dual representation that is quadratic and convex (possessing a global minimum but no other local minima) (Cristianini and Shawe-Taylor, 2000). Eq. (3.6) is thus re-expressed as

$$\hat{v} = \sum_{\ell=1}^L \alpha_\ell K(\mathbf{u}, \mathbf{u}_\ell) + \alpha_0. \quad (3.7)$$

The kernel function  $K(\mathbf{u}, \mathbf{u}_\ell) = \langle \psi(\mathbf{u}), \psi(\mathbf{u}_\ell) \rangle$  is used to implicitly evaluate the map  $\psi$  in  $\mathcal{F}$ . Table 3.2 lists several kernels for regression. Choosing the linear regression with  $C \rightarrow \infty$  reduces to multiple linear regression using the  $\epsilon$ -insensitive function to penalize error. The cubic spline kernel can be an advantageous choice for nonlinear regression as it does not impose additional parameter setting.

Critical choices are the values for SVR parameters  $\epsilon$  and  $C$  and the kernel function. These are optimized against a validation set, although Kwok and Tsang (2003)

Table 3.2 Kernel functions

| kernel                                                         | mathematical formulation                                                                                                                                                                                |
|----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| polynomial<br>(linear: $p = 1$ ,<br>quadratic: $p = 2$ , etc.) | $k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + 1)^p$                                                                                                                            |
| cubic spline                                                   | $k(\mathbf{x}, \mathbf{y}) = 1 + \langle \mathbf{x}, \mathbf{y} \rangle + \frac{1}{2} \langle \mathbf{x}, \mathbf{y} \rangle \min(\mathbf{x}, \mathbf{y}) - \frac{1}{6} \min(\mathbf{x}, \mathbf{y})^3$ |

recently demonstrated the optimal relationship

$$\epsilon = 1.0043\sigma \quad (3.8)$$

where  $\sigma$  is the SD of the estimated output noise distribution under an assumption of normality, a relation that avoids additional computation and/or data costs. System performance is then evaluated using a test set. A measure of multiple regression performance across a set is the coefficient of determination

$$r^2 = 1 - \frac{\langle \hat{v} - v, \hat{v} - v \rangle}{\langle v - \bar{v}, v - \bar{v} \rangle} \quad (3.9)$$

where  $\bar{v}$  is the learning set mean value of vector  $v$  of known outputs and  $\hat{v}$  is the vector of corresponding predicted values. When calculated for a training set, statistic (3.9) could be misleading: increasing the number of variables increases the value for  $r^2$ , although the model itself may be worsening. An adjusted coefficient of determination that accounts for the number  $L$  of learning samples and number  $P$  of variables is

$$r_{adjusted}^2 = \frac{(L - 1)r^2 - P}{L - 1 - P} \quad (3.10)$$

For a linear model, any difference between the learning set  $r_{adjusted}^2$  and a test set  $r^2$  indicates an insufficient learning set representativeness of the population being

modeled. A robust quality is sought of an SVR machine, in that regression capacity in the validation set translates well to a test set. Dual criteria for parameter optimization are the maximization of validation set  $r^2$  and minimization of the absolute differential between validation and learning set  $r^2$ 's. Accordingly, this paper proposes the following decision rule:

$$\max_{parameters} r_{validation}^2 - |r_{validation}^2 - r_{adjusted}^2|. \quad (3.11)$$

When the total number of samples is small,  $K$ -fold cross-validation (CV) is a standard method for learning system evaluation. Samples are divided into  $K$  subsets, and the training algorithm executes  $K$  times. For each trial,  $K - 1$  subsets form a training set and a different remaining subset acts as test set. System performance measures are then averaged across all  $K$  trials.

### 3.2.2.3 Functional model assessment

Subsection 3.2.2.1 presented a framework for functional feature extraction. Subsection 3.2.2.2 outlined a method to relate those features. This subsection defines a metric for the assessment of goodness of fit of a functional model that accepts functions in input and output. Let  $x(z)$  be a known, desired output of a learning system and  $\hat{x}(z)$  the actual prediction. Further, let  $\bar{x}(z)$  be the mean of the distribution made up of all learning example outputs. The functional coefficient of determination is analogous to (3.9):

$$R^2 = 1 - \frac{(L_2 R)^2}{(L_2 DM)^2} \quad (3.12)$$

where  $L_2R$  is the  $L_2$ -norm of the residuals given by

$$L_2R = \sqrt{\int (\hat{x}(z) - x(z))^2 dz} \quad (3.13)$$

and  $L_2DM$  is the  $L_2$ -deviation from the mean

$$L_2DM = \sqrt{\int (x(z) - \bar{x}(z))^2 dz} \quad (3.14)$$

For each sample, (3.12) is interpreted as the proportion of variance of the original function  $x(z)$  explained by the learning system's prediction  $\hat{x}(z)$ . This is done by comparing the error using the functional model  $L_2R$  against the error under a null hypothesis  $L_2DM$  in which the prediction is the ML estimate  $\bar{x}(z)$ . Values for  $R^2$  fall in the interval  $(-\infty, 1]$ . Negative values indicate that the mean function is a better estimate than  $\hat{x}(z)$ —this occurs when a small prediction error  $L_2R$  is compared to an even smaller deviation from the mean  $L_2DM$ . Thus, care must be taken in inferring from the functional coefficient of determination as negative values do not necessarily imply poor predictive capability.

### 3.2.3 Materials

This section describes data acquisition for this study and the representation thereof by a relatively small number of functional MNF coefficients that become inputs and outputs to SVR.

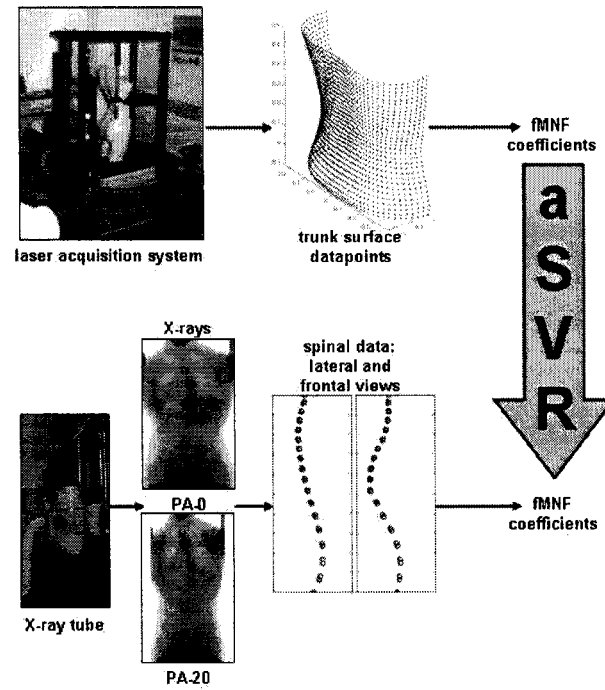


Figure 3.2 Learning system flow chart. Two postero-anterior full spinal X-rays were acquired quasi-simultaneously along with a 3D laser scan of the trunk. Anatomical landmarks were identified on the X-rays, permitting the generation of a 3D reconstruction. Data were registered, after which distinct functional data analyzes were performed on the trunk surface and both spinal curve projections. Resulting functional MNF coefficients became inputs and outputs to support vector regression.

### 3.2.3.1 Data acquisition

A database from Calgary was used in this study. Acquisition is described in this subsection. Data were acquired at the Alberta Children's Hospital, and consisted of 196 samples with repetition at a minimal frequency of 6 months. That database was the same as used by Jaremko et al. (2002a), Bergeron et al. (2004) and Bergeron et al. (2005a), but enlarged as new samples became available and patients diagnosed with congenital scoliosis were included. Trunk and spine data were quasi-simultaneously acquired from scoliosis patients. The acquisition protocol was described by Poncet et al. (2000) and is summarized here.

Two postero-anterior full spinal X-rays were acquired from scoliosis patients. The X-ray tube was positioned horizontally and at an angle of  $20^\circ$  from the horizontal (PA-20). On each X-ray, a trained technician manually identified six anatomical landmarks (the upper and lower endplate centroids, and left and right upper and lower pedicle bases) per thoracic and lumbar vertebra, as well as calibration markers. Acquisitions for which the landmark identification process was incomplete were excluded. The markers allowed for the 3D reconstruction of the landmarks, using a direct linear transformation (DLT). For each marker, the error associated with the reconstruction process was known. In the anterior plane, this error presented a SD of 2.2 millimetres, mostly explained by the anatomical landmark identification process. In the sagittal plane, the simulated data presented an additional error distribution with a SD of 5.1 millimetres, explained by the lack of quality of sagittal information contained in PA X-rays. These statistics were invariant across anatomical landmarks and acquisition years. In clinical terms, the set of 34 points  $(x, y, z)$  representing the position of the endplate centroids defines the vertebral body line, referred to in the current paper as the spinal curve.

Three-dimensional trunk surface coordinates were also obtained from each of four

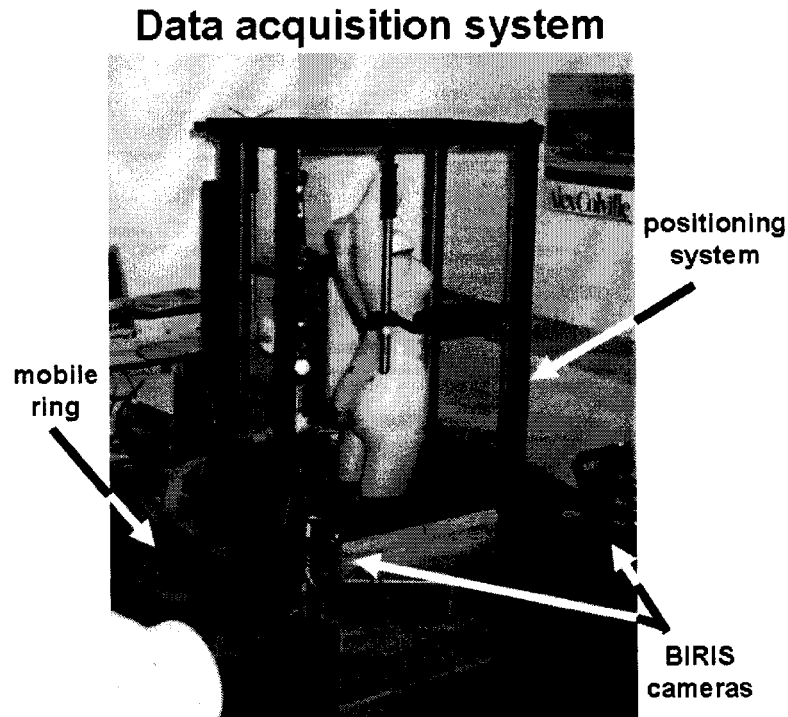


Figure 3.3 Data acquisition system. 3D trunk surface coordinates were obtained from each of four laser scanners mounted on a mobile ring and connected to a computer camera-control and data acquisition system. A positioning system ensured standard posture. X-rays were quasi-simultaneously acquired as the patient remained in the same position, using a portable X-ray tube.

*BIRIS* laser scanners mounted on a mobile ring and connected to a computer camera-control and data acquisition system (Poncet et al., 2000), illustrated in Fig. 3.3. A positioning system assisted with ensuring a consistent posture in patients.

Those views were merged into a single 3D dataset through rigid registration computed using calibration points located in the positioning system. The accuracy of that system was  $4.2 \pm 1.7$  millimetres (Tardif et al., 2000). As the scan took

Table 3.3 Statistics for selected clinical and spinal indices

| set                                                                | first           | second          | third           | fourth          | fifth           |
|--------------------------------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| acquisition year                                                   | 1998            | 1999            | 2000            | 2001            | 2002            |
| number of samples                                                  | 37              | 44              | 51              | 35              | 29              |
| number of distinct patients                                        | 29              | 34              | 35              | 23              | 19              |
| female : male ratio                                                | 26 : 11         | 33 : 11         | 41 : 10         | 29 : 6          | 25 : 4          |
| adolescent idiopathic:<br>juvenile idiopathic:<br>congenital ratio | 21 : 13 : 3     | 27 : 11 : 6     | 28 : 18 : 5     | 18 : 12 : 5     | 18 : 10 : 1     |
| age (mean years $\pm$ SD)                                          | 12.3 $\pm$ 3.0  | 13.0 $\pm$ 2.8  | 12.7 $\pm$ 2.1  | 12.6 $\pm$ 1.6  | 12.7 $\pm$ 1.6  |
| maximal anterior<br>Cobb angle<br>(mean degrees $\pm$ SD)          | 33.1 $\pm$ 13.0 | 35.6 $\pm$ 15.1 | 35.4 $\pm$ 14.9 | 29.4 $\pm$ 8.9  | 27.9 $\pm$ 8.3  |
| lordosis<br>(mean degrees $\pm$ SD)                                | 26.6 $\pm$ 15.4 | 23.5 $\pm$ 12.3 | 26.0 $\pm$ 15.6 | 25.1 $\pm$ 19.1 | 23.7 $\pm$ 19.2 |
| cyphosis<br>(mean degrees $\pm$ SD)                                | 28.1 $\pm$ 13.7 | 31.7 $\pm$ 13.9 | 30.8 $\pm$ 10.2 | 29.5 $\pm$ 11.9 | 24.3 $\pm$ 12.4 |

approximately 15 seconds, subject movement accounted for an additional tolerated error of up to 1.5 millimetres (Poncet et al., 2000). Arms were removed using a cylinder, and trunk data were then smoothed using B-splines. The result was a dataset consisting of a series of horizontal cross-sectional contours, each containing up to 360 3D coordinates and vertically separated by approximately 10-millimetre intervals.

The database was divided into 5 sets (Table 3.3) used in 5-fold CV. Defining these sets based on acquisition year ensured that each set contained no more than 2 repeated samples per patient.

### 3.2.3.2 Data registration

Registration processes data, providing a similar basis for comparison. Nine surface markers appearing in both spinal and trunk datasets were used to translate those to a common 3D Euclidean space with the origin  $(0, 0, 0)$  defined as the lower endplate centroid of the fifth lumbar ( $L5$ ) vertebra. Coordinates  $(0, 0, 1)$  were defined as the upper endplate centroid of the first thoracic vertebra ( $T1$ ). Consequently, data was scaled to unit spinal vertical span, and the decorrelation of spinal shape and orientation was achieved. The  $x$ -axis was oriented towards the front of the patient, the  $y$ -axis, towards the patient's left, and the  $z$ -axis, upwards. Hence, the right sagittal view is the projection onto the  $xz$ -plane as seen from the negative  $y$ -axis and the anterior view is the projection onto the  $yz$ -plane viewed from the positive  $x$ -axis.

Spinal curves were represented explicitly as the functions  $x = x(z)$  and  $y = y(z)$  such that  $x, y : [0, 1] \rightarrow \mathbb{R}$ . The trunk surface was represented explicitly in cylindrical coordinates as the function  $r = r(\theta, z)$  such that  $r : [0, 2\pi] \times [-0.1, 0.9] \rightarrow \mathbb{R}$ .

It is noted that the retained vertical span for trunk data begins below  $L5$  as data was available, but is cut below  $T1$  as the horizontal shoulders interfered with later modeling. Finally, mean functions  $\bar{x}(z)$ ,  $\bar{y}(z)$  and  $\bar{r}(\theta, z)$  were determined; care was taken to verify that they were smooth.

Registration was followed by a functional data analysis (FDA), described in the following subsections. The goal of FDA was clear: to identify a compact representation for the function spaces  $\mathcal{X} = \{x(z)\}$ ,  $\mathcal{Y} = \{y(z)\}$  and  $\mathcal{R} = \{r(\theta, z)\}$ , being subsets of the space of smooth functions.

### 3.2.3.3 Functional data analysis: spinal curve

As stated earlier in this section, sagittal plane spinal curve simulations were very noisy. Consequently, the FDA was performed on a spinal database of better quality. Montréal data were acquired at *Hôpital Sainte-Justine* and consisted of 369 samples from distinct scoliosis patients. The spinal acquisition protocol was similar as that from Calgary data, save that PA-0 and LAT X-rays were acquired. Errors in the 3D spinal reconstructions thereof were 2.2 millimetres in each of the anterior and right sagittal planes. These data were invoked to identify a spinal curve functional representation of low dimensionality.

Wavelets, defining sets that are complete and orthonormal, were chosen for their broad representation capability; the basis simply defined by the Haar mother wavelet worked well. Accordingly, data were functionally represented as follows:

$$x(z) = \bar{x}(z) + \sum_{m=1}^M a_m \phi_m(z) \quad (3.15)$$

$$y(z) = \bar{y}(z) + \sum_{m=1}^M b_m \phi_m(z)$$

where  $a_m$  and  $b_m$  represent the wavelet coefficients, and  $\phi_m$ , the Haar wavelets. Coefficients were computed from the discrete wavelet transform (DWT), an efficient algorithm. Periodic extension padding was used. Calculations were performed using Matlab (Mathworks, 2001). The discretely-defined functions  $x$  and  $y$  were not regularly sampled over  $z$ , and, being noisy, the definition of a smooth interpolant would have been difficult. Fortunately, mechanical constraints derived from the size of the vertebrae and their general rigidity, coupled with the weakness of the curvature induced by scoliosis, guarantee fairly consistent relative heights of data points (its standard deviation of the middle vertebra,  $T9$ , expressed relative

to spinal vertical span, was 1.5%). For each spinal function, 17 approximation coefficients and 17 detail coefficients were retained. Thus,  $M = 34$ . For these coefficients, a covariance matrix could be estimated.

MNF also requires an estimate for the noise covariance matrix. To this end, the propagation of the 3D spinal reconstruction noise to the wavelet coefficients needed to be determined. To this end, 5000 34-vectors with random entries observing a Gaussian distribution with the error SD were generated, and the wavelet coefficients thereof were calculated. This propagated error was uniform across all coefficients, and permitted the estimation of a noise covariance matrix. This protocol assumes that this noise is white and additive, as imposed by the MNF derivation—Fig. 3.6(a) confirms that this assumption is valid, a logical finding as the 3D spinal reconstruction error was mostly explained human error in the manual identification of anatomic landmarks. Thus, MNF discovered a smaller functional basis, replacing the functional representation of (3.15) with

$$x(z) = \bar{x}(z) + \sum_{m=1}^{M'} \alpha_m x_m(z) + \epsilon_x(z) \quad (3.16)$$

$$y(z) = \bar{y}(z) + \sum_{m=1}^{M'} \beta_m y_m(z) + \epsilon_y(z)$$

where  $\alpha_m$  and  $\beta_m$  are functional MNF coefficients,  $x_m$  and  $y_m$  are the MNF functions, subscripts refer to functional component numbers, and  $\epsilon(z)$  are noise terms for a model of size  $M' < M$ . The integer  $M'$  was determined based on SVR predictive performance. Table 3.4 reports relative variability (3.2) and SNR (3.5) for leading fMNF components; the modeling error is evaluated using the  $L_2$ -norm (the integration was computed numerically using a trapezoidal method):

$$\|\epsilon(z)\| = \sqrt{\int_0^1 \epsilon^2(z) dz}. \quad (3.17)$$

Table 3.4 Properties of spinal basis functions.

| spinal component | relative variability (%) | signal-to-noise ratio | modeling $L_2$ -error (mean mm $\pm$ SD) |
|------------------|--------------------------|-----------------------|------------------------------------------|
| average sagittal | —                        | —                     | $11.16 \pm 4.42$                         |
| first sagittal   | 85                       | 160                   | $9.50 \pm 3.74$                          |
| second sagittal  | 12                       | 21                    | $5.19 \pm 1.78$                          |
| third sagittal   | 1.5                      | 1.7                   | $4.33 \pm 1.42$                          |
| fourth sagittal  | 0.41                     | 0.12                  | $3.71 \pm 1.08$                          |
| average anterior | —                        | —                     | $8.02 \pm 4.66$                          |
| first anterior   | 74                       | 640                   | $5.30 \pm 3.30$                          |
| second anterior  | 22                       | 190                   | $2.32 \pm 1.17$                          |
| third anterior   | 2.7                      | 22                    | $1.44 \pm 0.73$                          |
| fourth anterior  | 0.72                     | 5.2                   | $1.03 \pm 0.62$                          |

It should be noted that Table 3.4 reports statistics for Calgary data.

The components appearing in Table 3.4 define spaces  $\mathcal{X}$  and  $\mathcal{Y}$  that are subsets to the set of smooth functions. These basis functions are illustrated in Fig. 3.4.

### 3.2.3.4 Functional data analysis: trunk surface

This subsection performs an FDA on trunk data, with the goal of using function coefficients as inputs towards the prediction of the spinal curve. For sagittal spinal prediction, trunk coefficients were computed to represent a narrow column of the back surface for the full available vertical span (corresponding to domain  $[135^\circ, 225^\circ] \times [-0.1, 0.9]$ ). For anterior spinal prediction, a wider portion of the back was retained, but the upper portion was removed (corresponding to domain  $[90^\circ, 270^\circ] \times [-0.1, 0.6]$ ). Fig. 3.5 illustrates those trunk regions.

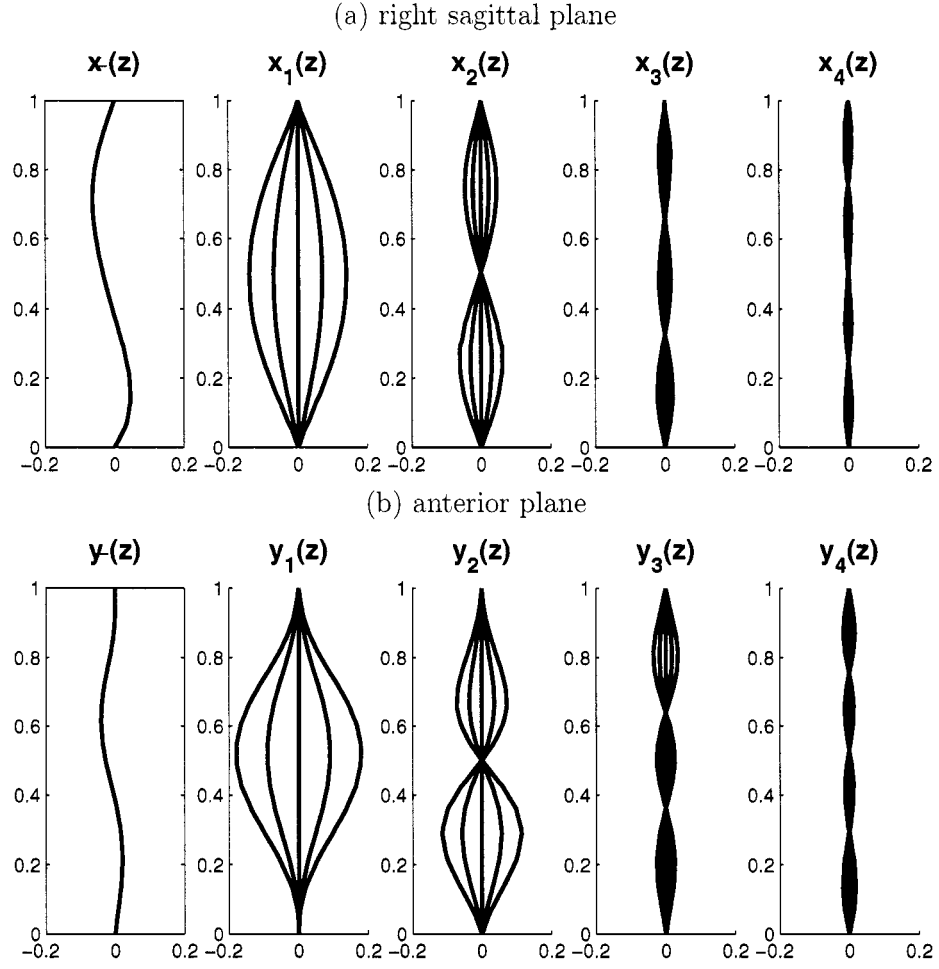


Figure 3.4 Spinal functional bases. Mean functions and leading functional minimum noise fractions for spinal curve representation. These basis functions define functional spaces  $\mathcal{X}$  and  $\mathcal{Y}$  of spinal curves in the right sagittal and anterior planes.

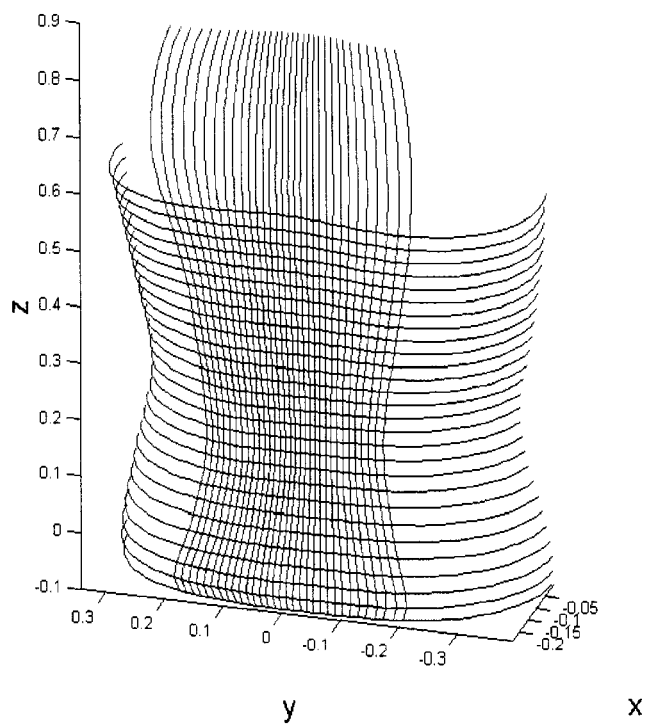


Figure 3.5 Retained trunk sections. For sagittal spinal curve prediction, a  $90^\circ$  angular span of the back was retained, presented here by vertical curves. For anterior spinal curve prediction, a  $180^\circ$  span of the back cut off at the height of the lower sternum (thus excluding the holes resulting from arm removal) was retained, illustrated here with horizontal curves.

Once again, a wavelet basis defined by the Haar mother wavelet was retained. Trunk data were interpolated to a  $128 \times 40$  regular grid: piecewise cubic Hermite interpolating polynomials (PCHIP) ensured monotonic interpolation between data points in the angular orientation, and natural cubic splines were then suitable in the vertical direction. As trunk data was smoothened as part of the acquisition process, detail coefficients were small and neglected. Accordingly, data were functionally represented as follows:

$$r(\theta, z) = \bar{r}(\theta, z) + \sum_{i=1}^I \sum_{j=1}^J c_{i,j} \phi_i(\theta) \phi_j(z) \quad (3.18)$$

where  $c_{i,j}$  represent the approximation coefficients and  $\phi_i$  the Haar wavelets. Repeated interpolation of trunk data made difficult the definition of a noise covariance matrix; the impact of the joint laser acquisition accuracy/subject movement error reported earlier in this section, assumed to be additive and white, on the wavelet coefficients was determined to estimate a noise covariance matrix. Fig. 3.6(b) verifies the extent of this assumption by plotting the latent values of the trunk wavelet coefficient covariance matrix, and comparing the sequence with the theory (3.3).

The trunk functional MNF representation of (3.18) is

$$r(\theta, z) = \bar{r}(\theta, z) + \sum_{n=1}^{N'} \gamma_n r_n(\theta, z) + e(\theta, z) \quad (3.19)$$

where  $\gamma_n$  are functional MNF coefficients,  $r_n$  are the MNF functions, subscripts refer to functional component numbers, and  $e(\theta, z)$  is a noise term for a model of size  $N' < N = IJ$ . The choice for  $N'$  was based on predictive performance, but no MNF component presenting an SNR below 60 was retained.

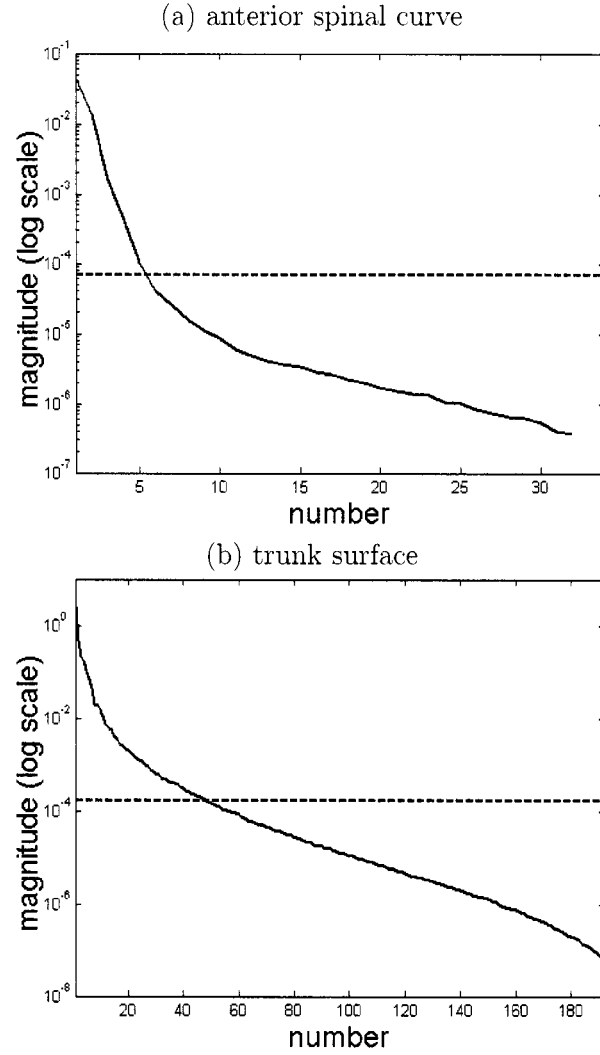


Figure 3.6 Decreasing sequences of latent values. These sequence of decreasing latent values were computed for comparison with the theory (3). (a) With respect to the covariance matrix estimated from anterior spinal curve wavelet coefficients (16), a plateau is observed near the variance of the estimated reconstruction error (dashed line). This graph supported the assumption that noise in spinal data was white and additive. A few latent values were found below the machine precision (dotted line), and result from numerical computation errors. (b) With respect to the covariance matrix estimated from trunk surface wavelet coefficients (16), the plateau is less flat than in (a), yet it occurs near the variance of the estimated acquisition error, suggesting that the assumption of white and additive noise was less weaker, and that indeed some interdependent noise was present. A second plateau of latent values below the machine precision was also found.

### 3.2.3.5 Support vector regression learning

An array of SVRs was trained to predict spinal coefficients from trunk coefficients. One SVR was assigned to each output. The implementation by Gunn (1998) was used.

Five-fold cross-validation was implemented, using the 5 subsets of Table 3.3. Thirty trials was required to execute all combinations of a learning set made up of 2 subsets, a validation set formed from 2 subsets, and a testing set corresponding to 1 subset. Varying subset dimensionalities was not problematic, as each sample was tested 6 times. Keeping the training set relatively small and a validation set of roughly same size was necessary to prevent overfitting due to interdependencies in data stemming from repeated patient samples.

The value for SVR insensitivity  $\epsilon$  was determined according to the relation (3.8), for which  $\sigma$  had to be estimated. For each output, this was done by defining  $\sigma$  as the SD of the residuals of a linear SVR with  $C \rightarrow \infty$  and  $\epsilon = 0$ .

Protocol A was used to optimize, for each output  $\alpha_m$  and  $\beta_m$  (3.16), SVR input dimensionality  $N'$ . A linear SVR with  $C \rightarrow \infty$  and  $\epsilon = 0$  was 5-fold cross-validated for integer values for  $N'$  (3.19) ranging between 1 and 18 inclusively. For each output, the retained  $N'$  maximized  $r_{test}^2$ .

Then, protocol B was used to optimize the remaining parameters: the regularization parameter  $C$  and kernel parameters for kernels appearing in Table 3.2. To this end, a 5-fold cross-validation was performed according to decision rule (3.11). A parameter space  $[0.01, 1000]$  for  $C$  was comprehensive.

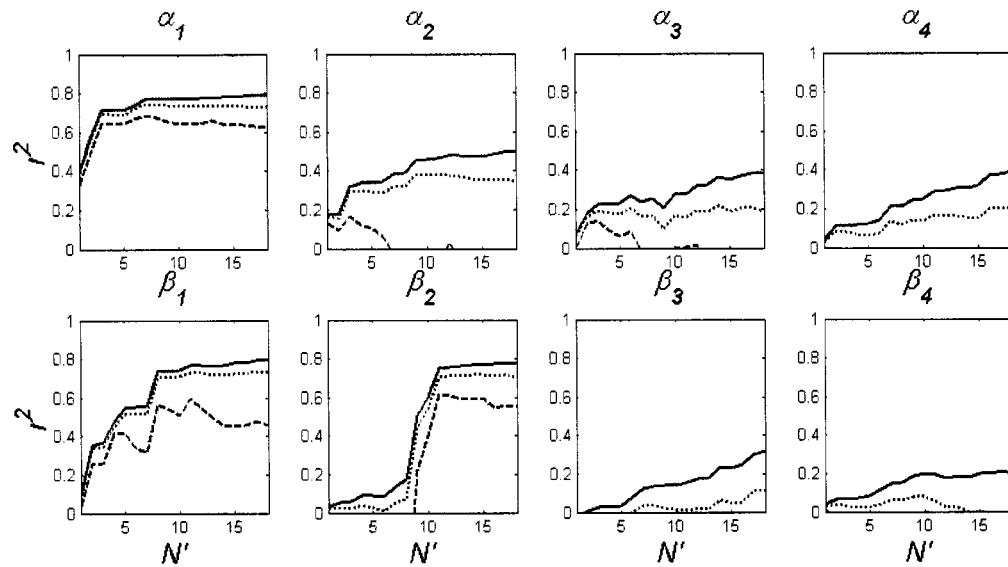


Figure 3.7 SVR input dimensionality optimization. Graphs present the variation of the coefficient of determination as input dimensionality increases. The solid line represents the learning set, the dotted line represents the adjusted  $r^2$  in the learning set, and the dashed line represents the validation set. Protocol A was used, and the decision rule maximized  $r^2$  in the validation set. For some outputs, validation set predictive capability is not displayed as it is negative.

### 3.2.4 Results

Fig. 3.7 presents the results of Protocol A to determine optimal input dimensionalities for each output. For each graph, 4 coefficients of determination (3.9, 3.10) are plotted, corresponding to learning, learning (adjusted value), validation and test sets. Table 3.5 presents, for each spinal component, the number of inputs retained for SVR.

Table 3.5 also presents the prediction capability resulting from Protocol B. For all outputs, the linear kernel was optimal.

Table 3.5 Support vector regression prediction of spinal minimum noise fraction coefficients

| spinal component | number of inputs | training $r^2$ | training $r^2_{adj}$ | validation $r^2$ | test $r^2$ |
|------------------|------------------|----------------|----------------------|------------------|------------|
| first sagittal   | 7                | 0.74           | 0.72                 | 0.70             | 0.69       |
| second sagittal  | 3                | 0.26           | 0.23                 | 0.23             | 0.19       |
| third sagittal   | 3                | 0.14           | 0.11                 | 0.14             | 0.09       |
| first anterior   | 11               | 0.64           | 0.60                 | 0.59             | 0.51       |
| second anterior  | 11               | 0.69           | 0.64                 | 0.56             | 0.50       |

The test predictions for spinal curve coefficients for each CV trial were used to generate the corresponding spinal curves. Those SVR-predicted spinal curves were compared to a blind, maximum-likelihood (ML), prediction, in which the mean spinal curve constituted the prediction for all samples. Error distributions for those predictions are presented as Fig. 3.8; statistics describing these are also included.

Fig. 3.9 permits a qualitative comparison between acquired and predicted spines; they are graphed in order of increasing  $L_2$ -errors.

### 3.2.5 Discussion

The first major contribution of this paper was the implementation of MNF on the Montréal spinal data to learn the underlying global functional components and, in particular, the leading noise-insensitive ones. Fig. 3.6 verified that the noise being estimated was sufficiently white and additive, sustaining the choice of MNF. Working with 34 points per spinal curve permitted to cut through the particularly noisy simulations of the sagittal spinal curve in Calgary data, and the dual interpretations of relative variance of the latent values permitted to better understand the contribution of each component. The result is 3 sagittal view MNF components that present an SNR greater than unity in the Calgary simulations.

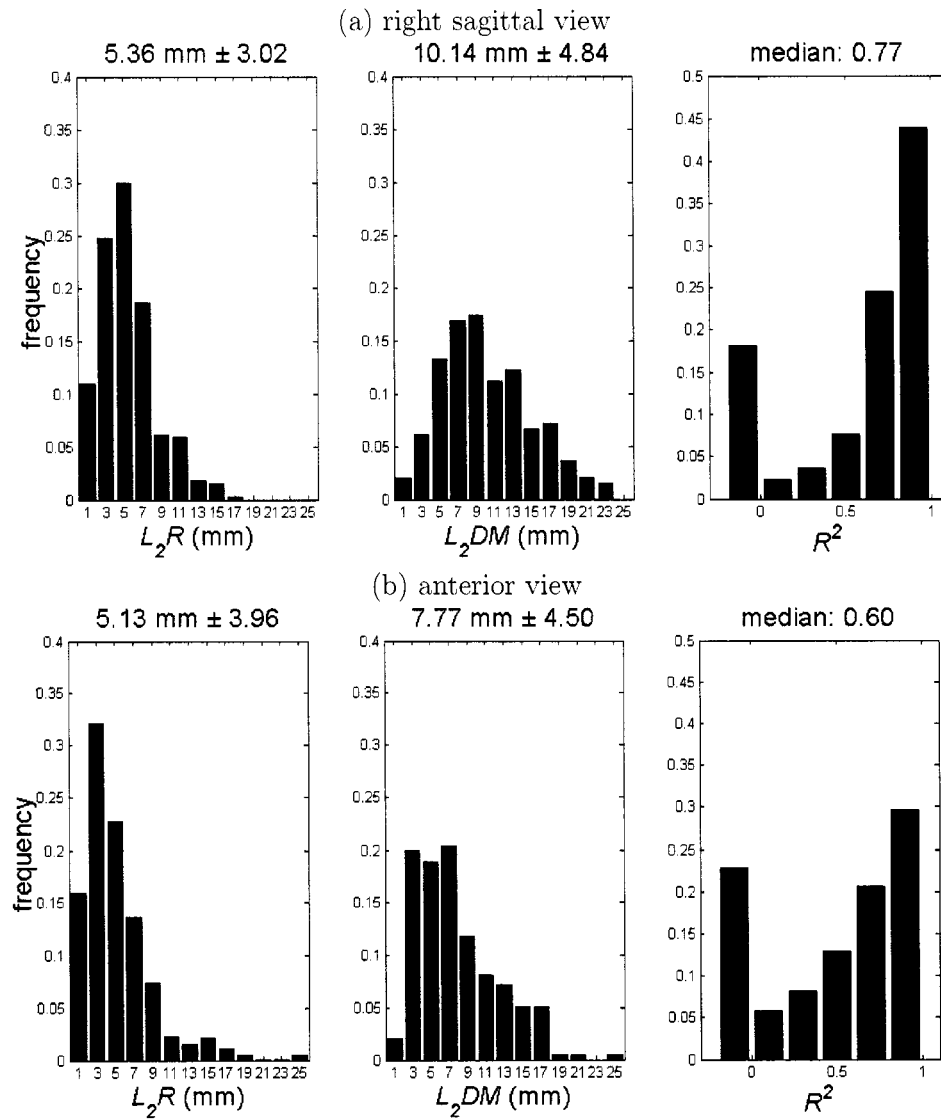


Figure 3.8 Comparison of support vector regression (SVR) and maximum-likelihood (ML) spinal curve predictions in the (a) right sagittal and (b) anterior planes. Test results are presented.  $L_2R$  is the  $L_2$ -norm of the residuals (3.13) from SVR.  $L_2DM$  is the  $L_2$ -deviation from the mean (3.14), corresponding to the  $L_2$ -norm of the residuals for an ML prediction. For both predictions, the mean and standard deviation are reported. The expression for the functional coefficient of determination  $R^2$  is given by (3.12)—the leftmost bar in histogrammes for  $R^2$  includes all negative cases. For each spinal projection, the median value is indicated.

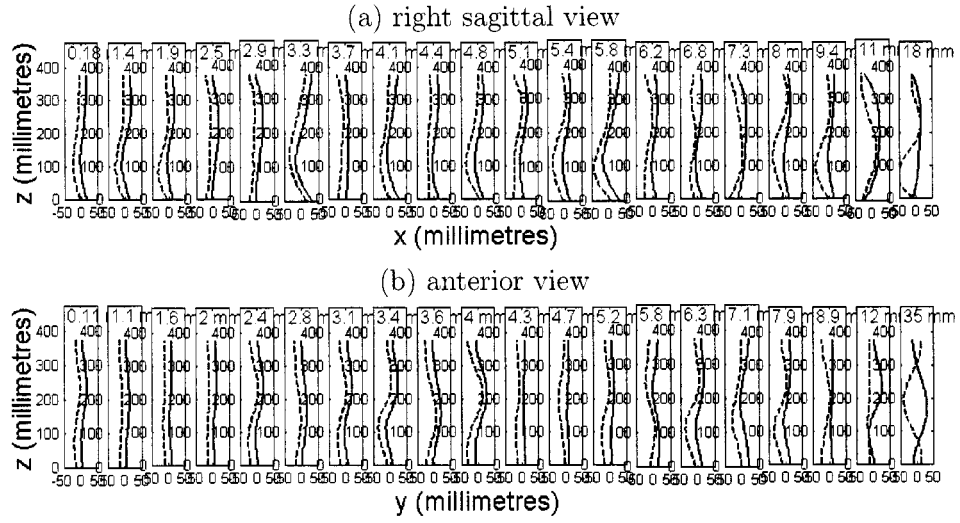


Figure 3.9 Comparison of X-ray-reconstructed and support vector regression (SVR)-predicted spinal curves in the (a) right sagittal and (b) anterior projections. X-ray-reconstructed spinal curves are dashed, while SVR-predicted spinal curves are solid and shifted right. 196 samples were tested 6 times, and testing results were sorted in order of increasing  $L_2$ -errors, which are reported. Roughly every 59th result is displayed. Numerical values thereof are indicated.

The same protocol yielded 4 anterior view functional components for the spinal curve with  $SNR > 1$ . The search for a relation between these uncorrelated external and internal features could then proceed.

The second major contribution of this paper was the introduction of cross-validation to artificial intelligence approaches in this application area. Fig. 3.7 clearly illustrates the mild-to-moderate overfitting of the models. For selected SVR input dimensionalities, a comparison of the learning set  $r^2$  and the adjusted value suggest that the number of model variables was not the cause of this lack of generalization. Instead, it is the constitution of training sets that were too small to encapsulate the diversity of the population that caused this overfitting. The classic fix involves increasing the number of cross-validation subsets to permit larger training set sizes. That approach would not have succeeded, as interdependencies in the database caused by repeated patient samples over time imposed working with training and validation sets of similar size to ensure that the validation and test sets contain samples from unseen patients.

Cross-validation required repeated testing on all 194 samples using different learning and validation sets permitted for high confidence in reported test set predictive capabilities, despite the slight-to-moderate lack of robustness in results. This study reports a model for leading sagittal spinal coefficients that explains 69, 19 and 9% of the variability, and a model for leading anterior coefficients that explains 51 and 50% of the variability (Table 3.5). The coefficient of determination increased as the relative variance increased, suggesting that more subtle characteristics of spinal shape are more difficult to predict.

Fig. 3.7 also highlights the nil predictive capability for  $\alpha_4$ ,  $\beta_3$  and  $\beta_4$ . Thus, these components do not appear in Table 3.5. This was not problematic as the relative variability thereof were small (Table 3.4), indicating that their impact of the spinal

curve prediction is very small. It should be noted that working with spinal bases of dimension 3 and 2, respectively, for the sagittal and anterior planes result in a modeling error below that found in data (Table 3.4).

When training with higher-order kernels was attempted, learning system overfitting was clear. Over time, as further samples are acquired, nonlinear kernels might better model the biological complexity of scoliosis geometry.

Fig. 3.8 presents the results when predicted spinal coefficients are used to construct the spinal curve. Values for the  $L_2$ -norm of the residuals are compared with the blind, ML prediction in which the mean spinal curve constitutes the prediction for all samples. The distributions for the functional coefficient of determination (3.12) are also displayed. Median values for  $R^2$  are reported for two reasons. First, distributions are clearly not Gaussian, resulting in the mean being a poor measure of central tendency. Second, because many negative values for  $R^2$  should not be interpreted as poor predictions, as discussed in Subsection 2.3—the median is less sensitive to their presence than the mean. This study concludes that the joint MNF-SVR model explains 77% of right sagittal view and 60% of anterior view spinal curve variability. Values for  $L_2R$  of 5.36 and 5.13 millimetres compare favourably with the combined error in data.

As noted earlier in this section, the database used in this study contained samples from repeated patients. While those samples were considered to be distinct, based on growth and scoliosis progression considerations, they were not independent. While a database of distinct samples would have been more representative, scoliosis clinical protocols will most certainly require one set of X-rays upon diagnostic in the foreseeable future. Thus, working with a database containing repeated patient scans is consistent with the clinical intentions of this study.

### 3.2.6 Conclusion

The purpose of this study was to investigate the extent to which prediction of the 3D shape of spinal deformity in scoliosis patients based on trunk shape asymmetries could be performed, despite working with noisy spinal data in the sagittal plane and a relatively small database of interdependent samples leading to overfitting complications. Minimum noise fractions and cross-validation enabled this study to nonetheless make strides in this application area. As further data is acquired, larger learning set dimensionalities should better reflect the variability in scoliosis trunk and spine shapes, reduce the mild to moderate observed lack in robustness, and permit more complex relations to be identified.

This approach permitted for a first predictive capability for sagittal spinal features in this research application to be reported to date. Other innovations include the strength of this predictive capability, and the comprehensiveness of the predicted outputs in describing spinal shape. This progress is important towards the possibility of reducing X-ray acquisitions in scoliosis patients without sacrificing critical information required during clinical follow-up.

This paper proposes the clinical adoption of the system developed in this study. The non-invasiveness of the trunk laser scan permitted, say, monthly follow-up visits. Over the period of a few years, a clinical study could validate SVR predictions at the current X-ray-acquisition frequency of 6 months. At the same time, that new data could add to the learning system's knowledge and thus improve its prediction. Following clinical verification, the authors envision this system reducing the need for scoliosis patient exposure to X-rays beyond a first acquisition upon diagnostic.

Currently, at *Hôpital Sainte-Justine* in Montréal, scoliosis data consisting of 3D reconstructed spines from X-rays as well as a full torso scan using a Inspec system,

based on shifted Moiré projections, is being acquired, albeit not simultaneously. Provided elastic registration methods to link these datasets are successful, the impact of improved external resolution and precision on spinal predictions would be examined. Further, the combination of Calgary and Montréal databases could allow for a sufficiently large number of samples to involve the entire trunk surface and/or higher-order kernels in spinal curve prediction.

The treatment of scoliosis depends on the progression of the deformity. The monthly acquiring of external data might allow for the identification of surface attributes indicating change in the internal structures. A similar study, with the incorporation of a temporal variable, might result in a useful tool for making clinical decisions. Hence, the use of artificial intelligence methods to predict scoliosis progression could be envisioned.

Of primary concern to patients with scoliosis is their appearance, and thus, the reverse relationship is also of interest. An evaluation of the bijectiveness of the external-internal relation discovered in this paper might constitute a first step towards understanding the nature of changes resulting from bracing or orthopaedic treatment.

### **3.3 Addendum**

Cette section présente l'état actuel des résultats obtenus en utilisant les méthodes de l'article présenté dans le présent chapitre.

### 3.3.1 Implémentation SVR

Cette sous-section reprend la méthodologie d'apprentissage et de sélection automatique des paramètres de la sous-section 3.2.3.

Un vecteur de SVR fut entraîné afin de prédire les coefficients rachidiens à partir des coefficients du tronc. Un SVR fut a été assigné à chaque valeur de sortie. L'implémentation par Gunn (1998) a été adaptée pour utiliser l'algorithme de Dantzig-Wolfe (Fletcher, 1987; Dantzig, 1963) pour résoudre le problème d'optimisation quadratique. Le paramètre d'insensibilité  $\epsilon$  fut déterminée selon la relation (3.8), où  $\sigma_N$  fut estimé comme étant l'écart-type de l'erreur d'un SVR linéaire entraîné sur tous les échantillons avec  $\epsilon = 0$ . Une distribution normale pour cette erreur est supposée.

Les paramètres de SVR linéaire ont été optimisés comme suit. Sous un noyau linéaire, le modèle présente une complexité très petite. Il n'est donc pas nécessaire de pénaliser cet attribut. Par conséquent, la valeur  $1 \cdot 10^6$  pour  $C$  fut retenue pour modéliser  $C \rightarrow \infty$ . Le nombre de valeurs d'entrée retenu (choisi sur l'intervalle  $1 - 18$ ) a été déterminée par CV d'un SVR linéaire. Pour chaque valeur de sortie, le choix fut effectué en fonction de l'optimisation du critère de l'équation (3.11). La validation croisée fut implémentée pour toutes les combinaisons possibles d'un groupe d'apprentissage formé de trois sous-ensembles et d'une groupe de test formé des deux sous-ensembles restants. La grandeur variable de chaque sous-ensemble n'a pas été problématique, puisque chaque échantillon fut testé six fois.

Les SVR utilisant le noyau de la spline cubique n'a requis que l'optimisation *a priori* du paramètre de régularisation. Le nombre de valeurs d'entrée retenu a été déterminée par CV de tels SVR—optimisant la règle de décision (3.11) pour  $C$  contre un groupe de validation. La validation croisée fut implémentée pour toutes

Tableau 3.6 Prédiction SVR linéaire de coefficients fMNF rachidiens

| coefficient rachidien | valeurs d'entrée | paramètre $\epsilon = \sigma_N$ | apprentissage $r^2$ | apprentissage $r^2_{adj}$ | test $r^2$ |
|-----------------------|------------------|---------------------------------|---------------------|---------------------------|------------|
| premier latéral       | 3                | 0.53                            | 0.71                | 0.70                      | 0.69       |
| deuxième latéral      | 3                | 0.71                            | 0.35                | 0.34                      | 0.26       |
| troisième latéral     | 2                | 0.80                            | 0.17                | 0.16                      | 0.11       |
| quatrième latéral     | 2                | 0.87                            | 0.13                | 0.11                      | 0.055      |
| premier frontal       | 8                | 0.60                            | 0.74                | 0.72                      | 0.62       |
| deuxième frontal      | 11               | 0.55                            | 0.75                | 0.72                      | 0.62       |
| troisième frontal     | 1                | 1.14                            | 0.0000              | 0.0000                    | -0.027     |
| quatrième frontal     | 1                | 1.08                            | 0.0032              | 0.024                     | 0.0052     |

Tableau 3.7 Prédiction SVR spline cubique de coefficients fMNF rachidiens

| coefficient rachidien | valeurs d'entrée | paramètre $\epsilon = \sigma_N$ | apprentissage $r^2$ | apprentissage $r^2_{adj}$ | validation $r^2$ | test $r^2$ |
|-----------------------|------------------|---------------------------------|---------------------|---------------------------|------------------|------------|
| premier latéral       | 1                | 0.78                            | 0.52                | 0.51                      | 0.39             | 0.38       |
| deuxième latéral      | 1                | 0.79                            | 0.23                | 0.22                      | 0.062            | 0.046      |
| troisième latéral     | 1                | 0.83                            | 0.21                | 0.20                      | -0.030           | -0.032     |
| quatrième latéral     | 1                | 0.90                            | 0.12                | 0.11                      | -0.15            | -0.22      |
| premier frontal       | 1                | 1.04                            | 0.21                | 0.19                      | -0.15            | -0.14      |
| deuxième frontal      | 1                | 1.08                            | 0.11                | 0.10                      | -0.40            | -0.36      |
| troisième frontal     | 1                | 1.14                            | 0.045               | 0.033                     | -0.49            | -0.46      |
| quatrième frontal     | 1                | 1.08                            | 0.12                | 0.11                      | -0.077           | -0.16      |

les combinaisons d'un groupe d'apprentissage formée de deux sous-ensembles, un groupe de validation formé de deux autres sous-ensembles et un groupe de test formé du sous-ensemble restant.

### 3.3.2 Résultats

Cette sous-section met à jour les résultats de la sous-section 3.2.4.

Le tableau 3.6 présente la capacité de prédiction de SVR linéaire pour chaque coefficient rachidien. Le tableau 3.7 fait de même pour SVR spline cubique.

Tableau 3.8 Prédiction SVR de la courbe rachidienne (apprentissage et test)

| <b>vue<br/>rachidienne</b> | <b>erreur-<math>L_2</math><br/>(moyenne mm)</b> |      | <b>erreur-<math>L_2</math><br/>(écart-type mm)</b> |      |
|----------------------------|-------------------------------------------------|------|----------------------------------------------------|------|
| latérale $x(z)$            | 5,03                                            | 4,97 | 2,78                                               | 2,72 |
| frontale $y(z)$            | 3,87                                            | 4,38 | 2,53                                               | 3,46 |

Les prédictions des coefficients rachidiens obtenus par les SVR linéaire ont été utilisées pour générer les courbes rachidiennes correspondantes. Les résultats apparaissant dans le tableau 3.8, et sont rapportés pour les groupes d'apprentissage et de test, respectivement.

La figure 3.10 présente les distributions pour le coefficient de détermination fonctionnel de ces prédictions.

La figure 3.11 permet la comparaison qualitative entre les reconstructions radiographiques et les prédictions SVR de la colonne vertébrale.

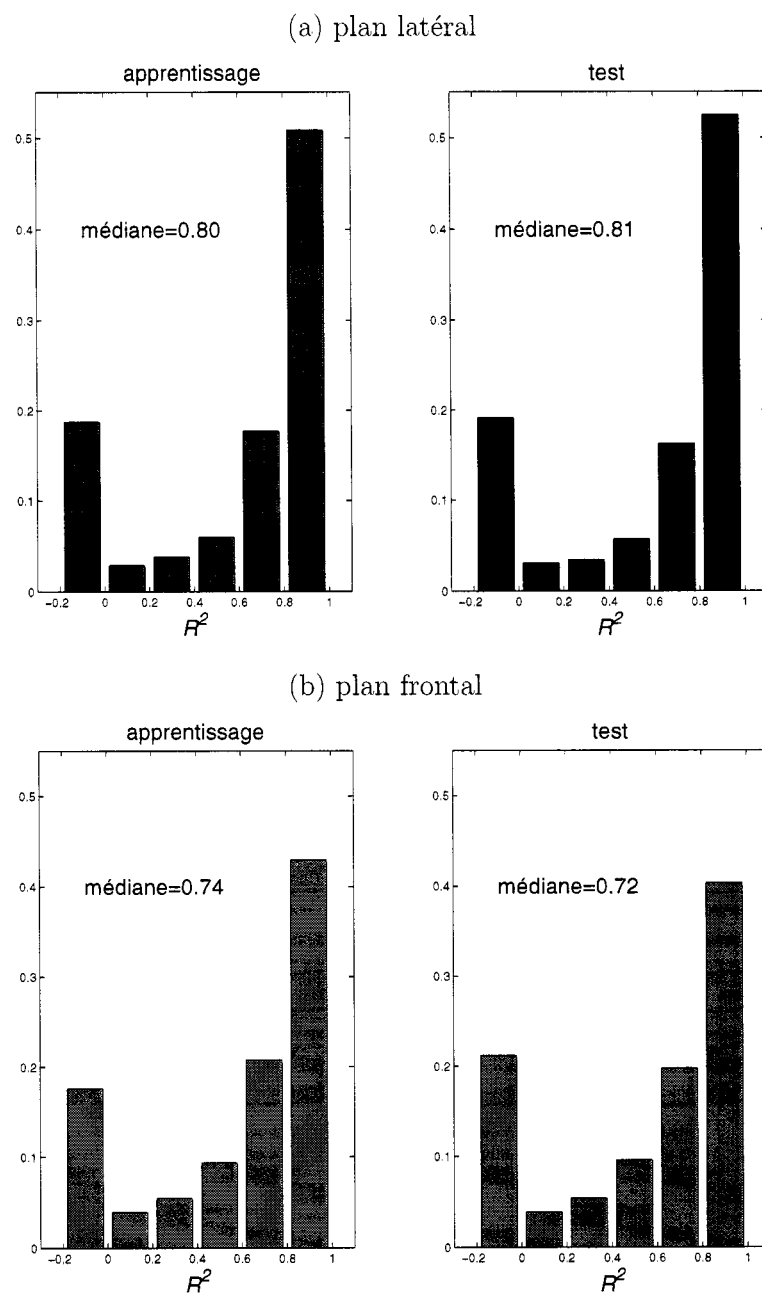


Figure 3.10 Ces figures présentent les distributions de  $R^2$  pour les plans rachidiens latéral et frontal pour les groupes d'apprentissage et de test—la barre de gauche dans chaque histogramme représente l'ensemble des valeurs négatives. Pour chaque graphique, la valeur médiane pour  $R^2$  est indiquée. Dans chaque plan, les distributions pour  $R^2$  sont très semblables, d'où l'affirmation que ces prédictions sont robustes.

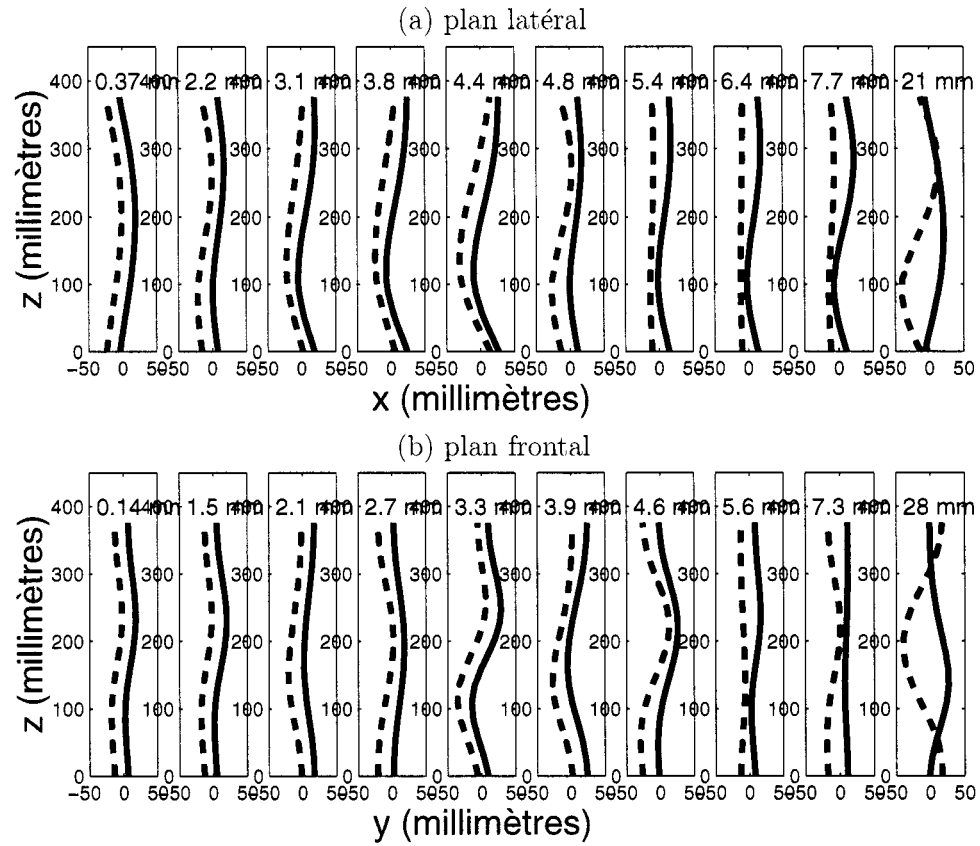


Figure 3.11 Cette figure compare les reconstructions radiographiques (courbes tiretées) aux prédictions SVR (courbes pleines) de la colonne vertébrale dans (a) le plan latéral et (b) le plan frontal. Ces prédictions du groupe de test sont présentées en ordre croissant d'erreur- $L_2$ . Des échantillonnages représentatifs sont présentés.

## DISCUSSION GÉNÉRALE ET CONCLUSION

L'objectif global du présent mémoire visait à identifier une relation robuste entre la forme du tronc et la forme de la colonne vertébrale. Pour ce faire, ce mémoire promettait :

1. L'étude de la représentation fonctionnelle de données afin de représenter la topographie du tronc et la courbe de la colonne vertébrale scoliotique à l'aide de coefficients qui sont globaux, représentatifs de la géométrie et moins sensibles au bruit d'acquisition ; et
2. L'implémentation d'une technique d'apprentissage robuste pour relier ces coefficients.

L'article du chapitre 2 définit une nouvelle approche pour exprimer l'asymétrie de la topographie du tronc et la déformation de la courbure du rachis. Le choix d'une modélisation globale par série de Fourier généralisée permet d'extraire un nombre relativement petit de coefficients qui sont à la fois compréhensifs de la géométrie et moins sensibles au bruit. Ce nombre de coefficients est ensuite réduit par l'analyse par composantes fonctionnelles. Le résultat est un modèle rachidien comportant 3 ou 4 paramètres pour chacune des projections frontale et latérale, et un modèle topographique du tronc comportant jusqu'à une quinzaine de paramètres.

L'exploitation de techniques d'apprentissage robustes, soit la régression par vecteurs de support, a ensuite permis de relier ces coefficients. Par son petit nombre de paramètres *a priori* à déterminer, sa pénalisation conjointe de l'erreur empirique et de la complexité du modèle et sa formulation en problème d'optimisation quadratique possédant une solution globale extrême, SVR est un nouvel algorithme puissant pour identifier des relations pour lesquelles les données sont bruitées. Cet apprentissage a permis l'identification de modèles externe-interne

expliquant entre 72% et 81% de la variabilité totale. Cette capacité est très encourageante, étant donné les sources de variabilité dans les données qui ne sont pas traitées explicitement par le modèle, telles le sexe, la morphologie corporelle, et l'état du développement physique (par exemple, la grosseur des seins). Cette capacité est également supérieure aux tentatives précédentes de recherche de relation. Finalement, ces modèles présentent une généralisation plus robuste que celle obtenue par l'approche GA-ANN de (Jaremko et al., 2002a).

Il est important de noter que l'analyse fonctionnelle de données fut très efficace dans la représentation implicite des non linéarités entre le tronc et la colonne, permettant l'identification de relations externe-interne sans avoir recours à des noyaux non linéaires. Cette possibilité a empêché le surapprentissage, un résultat critique en considérant la quantité limitée de données disponibles. Toutefois, il est raisonnable de suggérer que, au minimum, de légères non linéarités sont présentes dans la relation recherchée. L'approche SVR est l'outil idéal pour identifier ces non linéarités, étant donné un plus grand nombre de données d'apprentissage.

L'article de revue du chapitre 3 apporte des améliorations sur l'approche proposée dans le premier article. Tout d'abord, les fractions fonctionnelles de bruit minimales permettent d'évaluer le niveau de bruit dans chaque coefficient, afin de simplifier le choix du nombre de coefficients à retenir pour les modèles de la surface du tronc et de la courbe rachidienne dans les plans frontal et latéral. Ensuite, l'utilisation d'une vaste base indépendante de données sur la colonne vertébrale pour laquelle les données dans le plan latéral sont moins bruitées permet de définir un modèle rachidien plus juste, permettant un apprentissage de la relation externe-latérale. L'exécution de la validation croisée a permis d'accroître la confiance du système par une augmentation du nombre d'échantillons dans l'ensemble de test.

La méthode principale du présent mémoire, soit la régression par vecteurs de support, est une approche assez récente (Vapnik, 1998). La combinaison des idées menant à cette façon de faire évoluera sans doute au cours des prochaines années dans le but d'améliorer son implémentation à des problèmes concrets. En particulier, la recherche théorique sur la question d'optimisation automatique des paramètres SVR et du noyau permettra peut-être le développement d'algorithmes encore plus puissants que ceux faisant l'objet de cette étude. De plus, il serait intéressant d'incorporer les idées du *Partial Least Squares Regression* (McIntosh et al., 1996) à l'algorithme SVR afin de développer un outil d'intelligence artificielle à plusieurs valeurs de sortie.

Plusieurs questions sont soulevées par le présent mémoire. Faisant référence au tableau 3.3, les acquisitions utilisées dans cette étude comprennent d'importantes variabilités intersujet. Les paragraphes qui suivent discutent de quelques-unes de ces variabilités.

Tout d'abord, la base de données inclut des patients des deux sexes. L'importance relative du sexe et des effets scoliotiques sur la topographie du tronc a donc un impact sur la capacité de prédiction développée dans ce document. D'une part, l'algorithme génétique de (Jaremko et al., 2002a) avait effectivement retenu le sexe comme indice d'entrée au réseau de neurones prédisant l'angle de Cobb ; par contre, l'ajout de ce prédicteur aux coefficients topographiques fonctionnels de cette étude n'a pas modifié la performance prédictive des SVR. L'approche classique d'élimination de cette variabilité garçons-filles consisterait à entraîner des systèmes d'intelligence artificielle distincts pour chaque sexe. Du côté féminin, une légère détérioration de la performance prédictive a été observée. Du côté masculin, une étude distincte n'était pas possible vu le petit nombre d'échantillons. Une augmentation du nombre d'acquisitions simultanées externe-interne permettrait toutefois d'étudier cet aspect avec plus de profondeur.

Ensuite, la base de données comprend des patients dont l'âge couvre l'intervalle inclusif 8–17 ans, une période pendant laquelle d'importants changements corporels interviennent, soient la croissance et la puberté. Le changement le plus visible sur la topographie du tronc est sans doute l'apparition et le développement des seins chez les jeunes filles. Dans cette étude, l'utilisation de la surface du tronc entier a résulté en de prédictions bien moindres que ceux présentés dans ce mémoire utilisant le dos seulement. Devrait-on conclure que l'utilisation du tronc en entier introduit des problèmes de surapprentissage causé par un volume d'informations trop important, ou bien que les seins constituent une source de bruit inutile à la prédiction des déformations scoliotiques internes ? Cette étude semble tendre vers l'hypothèse selon laquelle les seins constituent de bruit dans la relation recherchée, une hypothèse contredisant l'avis du Prof. Janet Ronsky de l'Université de Calgary sur cette question. Encore une fois, une plus grande quantité d'échantillons permettrait peut-être de départager à cette question.

Finalement, d'autres différences existent entre les patients : la taille, la grandeur relative des membres, le poids, et ainsi de suite. Aucun de ces facteurs ne fut retenu par l'algorithme génétique de (Jaremko et al., 2002a), mais ceci n'exclut pas la possibilité qu'une décorrélation de ceux-ci soit possible ou pertinente. En particulier, dans les articles des chapitres 2 et 3, les données externes et internes furent normalisées à une longueur verticale commune de la colonne vertébrale, dans le but de standardiser le calcul des coefficients fonctionnels. L'impact de cette transformation n'a pas été étudiée.

Il est essentiel de rappeler que les prédictions présentées dans cette étude sont basées sur des reconstructions calculées à partir de deux radiographies PA. Cette approche de l'équipe de Calgary est motivée par le désir de reconstruire la cage thoracique en entier. En échange, la qualité de reconstruction 3D des repères anatomiques vertébraux est moins bonne que si des radiographies PA et LAT avaient été acquises.

Cette étude s'inscrit donc dans un scénario où les données ne sont pas de la meilleure précision possible.

Les données utilisées dans cette étude comportent des échantillons de patients répétés. Ces échantillons sont traités comme étant distincts, étant donné la croissance et la progression scoliotique entre les acquisitions, mais ceux-ci ne sont certes pas indépendants. Cet échantillonnage interdépendant est-il problématique ? Dans le contexte du développement d'un protocole de suivi scoliotique basé sur la topographie de l'externe, l'objectif vise à éliminer l'exposition *répétée* aux rayons X qui est nocive à la santé. Tant que l'acquisition de radiographies demeure *ponctuelle*, l'incidence accrue de développement de cancer est négligeable. Par conséquent, un nouveau protocole clinique de suivi continuerait sans doute à prévoir la prise d'imagerie radiologique lors du diagnostic de la scoliose, et l'utilisation de méthodes de prédiction de l'interne à partir de l'acquisition non invasive de la topographie externe lors de consultations subséquentes. La présence d'échantillons répétés dans la base de données utilisée est donc de mise avec l'application visée.

Il est de plus important de mentionner une autre problématique en ce qui concerne l'acquisition de données, soit la variabilité intrasujet. L'utilisation d'appareils de positionnement de patients lors d'acquisition de données externes et internes vise à minimiser cette variabilité. Toutefois, il semble que des variations dans la posture lors des acquisitions fasse en sorte qu'aucune relation robuste ne puisse être identifiée entre les projections frontale et latérale de la courbe rachidienne scoliotique (Bergeron et al., 2005b). La décorrélation de ces différences dans la posture lors d'acquisitions constitue une problématique difficile à contourner.

La question clinique *Existe-t-il une relation tronc-rachis scoliotique définitive ?* ou, reformulée dans le langage de l'Introduction de ce mémoire, *La supposition fondamentale selon laquelle les liens physiologiques externes-internes sont suffisamment*

*réguliers entre les patients pour permettre l'identification d'une relation est-elle justifiée ?* est posée. Ce mémoire répond, qu'au minimum, une relation partielle entre la topographie du tronc et la courbe de la colonne vertébrale scoliotique existe. La supposition fondamentale serait donc vraie dans une proportion de la variabilité variant entre 72 et 81%. L'introduction de nouveaux prédicteurs non considérés dans ce mémoire permettraient possiblement d'améliorer cette réponse. Toutefois, il serait raisonnable de postuler qu'une prédiction externe-interne exacte n'est probablement pas réaliste, puisque la diversité biologique chez l'être humain serait très difficile à complètement modéliser en utilisant un nombre fini de variables. Cette possibilité est soutenue par les dires du Dr Hubert Labelle de l'Hôpital Sainte-Justine lorsqu'il mentionne les patients scoliotiques *aberrants*, pour lesquels la réalité biologique fait contradiction à toute intuition clinique (par exemple, une gibbosité masurée gauche face à une scoliose thoracique droite). Ces cas aberrants pourraient expliquer certains des pires prédictions présentés dans les articles de ce mémoire.

Les résultats actuels sont très encourageants, mais une amélioration de la capacité prédictive du système serait nécessaire avant de permettre son implémentation clinique. Cette réalité est particulièrement évidente étant donné les situations où la prédiction est grossière. Toutefois, cette approche démontre un potentiel de remplacement éventuel des rayons X. Une étude clinique constituerait le meilleur moyen d'introduire cette technologie dans les hôpitaux, d'augmenter le nombre d'échantillons disponibles pour l'apprentissage, de raffiner la capacité de prédiction, et de valider son utilisation. Suite à cette étude, ce système pourrait réduire le besoin d'acquisition radiologique au-delà du diagnostic initial.

De plus, une analyse clinique des composantes fonctionnelles identifiées dans ce mémoire pourrait constituer une piste de solution pour relier ceux-ci aux indices traditionnels utilisés toujours en clinique. Ce transfert de l'expertise mathématique

développée dans ce mémoire au langage clinique traditionnel mettrait en valeur le travail ici présenté.

Présentement, à l'Hôpital Sainte-Justine de Montréal, de nouvelles données scoliotiques formées de reconstructions rachidiennes 3D à partir de radiographies de même que la topographie du tronc utilisant le système *Inspeck* sont en cours d'acquisition, quoique non simultanément. Pourvu que les méthodes de recalage élastique efficaces soient développées, l'impact d'une meilleure résolution externe et d'une meilleure précision interne pourra être étudiée. De plus, la combinaison des données de Montréal et de Calgary pourrait permettre l'utilisation de noyaux non linéaires tout en évitant le surapprentissage.

Le traitement de la scoliose est très dépendant de la progression de cette maladie. L'acquisition mensuelle de la topographie du tronc permettrait un suivi plus fréquent de la déformation des structures internes. Une étude semblable à celle-ci, mais incorporant une variable temporelle, pourrait faciliter les décisions cliniques de traitement. Les méthodes d'intelligence artificielle dans la prédiction de la progression scoliotique doivent donc être envisagées.

## RÉFÉRENCES

ADAIR, I.V., VAN WIJK, M.C., ARMSTRONG, G.W.D. (1977). Moire topography in scoliosis screening. *Clinical Orthopaedics*, 129, 165–171.

ANDRÉ, B., DANSEREAU, J., LABELLE, H. (1992). Effect of radiographic landmark identification errors on the accuracy of three-dimensional reconstruction of the human spine. *Med & Biol Eng & Comput*, 30, 569–575.

AYAT, N.E., CHERIET, M., SUEN, C.Y. (2002). Optimization of the SVM Kernels using an empirical error minimization scheme. *Lecture Notes in Computer Science*, 2388, 354–369.

BATOUCE, M., BENLAMRI, R., KHOLLADI, M.K. (1996). A computer vision system for diagnosing scoliosis using Moire images. *Comput Biol Med*, 26(4), 339–353.

BENAMEUR, S., MIGNOTTE, M., PARENT, S., LABELLE, H., SKALLI, W., DE GUISE, J. (2001). 3D Biplanar Reconstruction of Scoliotic Vertebrae Using Statistical Models. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. II 577–582.

BENNETT, K.P., CAMPBELL, C. (2000). Support Vector Machines : Hype or Hallelujah ? *SIGKDD Explorations*.

BERGERON, C., CHERIET, F., RONSKY, J., ZERNICKE, R., LABELLE, H. (2004). A Robust Methodology for Non-Invasive Follow-Up of Scoliotic Spinal Curve from Three-Dimensional Surface. *Proceedings of the International Research Society on Spinal Deformities Biannual Meeting*. Vancouver, 10–12 June 2004.

BERGERON, C., LABELLE, H., RONSKY, J., ZERNICKE, R. (2005a). Robust prediction of three-dimensional spinal curve from back surface for non-invasive follow-up of scoliosis. *Proceedings of the SPIE Medical Imaging Symposium*. San Diego, 12–17 February 2005.

BERGERON, C., CHERIET, F., MAC-THIONG, J.M., LABELLE, H. (2005b). Independence of the relationship between coronal and sagittal plane deformity assessed from spinal X-rays in adolescent idiopathic scoliosis. *Proceedings of the Canadian Orthopaedics Association Annual Meeting*. Montreal, 2–4 June 2005.

BERGERON, C., CHERIET, F., RONSKY, J., ZERNICKE, R., LABELLE, H. (Accepté pour publication en 2005). Prediction of Anterior Scoliotic Spinal Curve from Trunk Surface Using Support Vector Regression. *Engineering Applications of Artificial Intelligence*.

BERGERON, C., CHERIET, F., RONSKY, J., ZERNICKE, R., LABELLE, H. (Soumis en 2004). Non invasive assessment of scoliotic spinal curve using robust learning techniques. *Medical Image Analysis*.

BUNNELL, W.P. (1984). An Objective Criterion for Scoliosis Screening. *J Bone Joint Surg Am*, 66(9), 1381–1387.

CARMAN, D.L., BROWNE, R.H., BIRCH, J.G. (1990). Measurement of scoliosis and kyphosis radiographs. *J Bone Joint Surg Am*, 72, 328–333.

CARRO, P.A. (1999). Magnetic Resonance Imaging in Children with Scoliosis. *Semin Musculoskelet Radiol*, 3(3), 257–266.

CHOCKALINGAM, N., DANGERFIELD, P.H., GIAKAS, G., COCKRANE, T., DORGAN, J.C. (2002). Computer-assisted Cobb measurement of scoliosis. *Eur Spine J*, 11, 353–357.

COBB, J. (1948). Outline for the study of scoliosis. American Academy of Orthopaedic Surgeons : Instructional Course Lectures. Ann Arbor, Michigan. 5, 251.

CHRISTIANINI, N., SHAW-TAYLOR, J. (2000). An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University, Cambridge.

DANSEREAU, J., STOKES, I.A.F. (1987). Measurements of the three-dimensional shape of the rib cage. *J Biomechanics*, 21, 893–901.

DANSEREAU, J., BEAUCHAMP, A., DE GUISE, J., LABELLE, H. (1990). Three-dimensional reconstruction of the spine and the rib cage from stereoradiographic and imaging techniques. *16th Conference of the Canadian Society of Mechanical Engineering*, Toronto, Canada, 2, 61–64.

- DANSEREAU, J., LABELLE, H., AUBIN, C.É. (1993). 3-D Personalized Parametric Modelling of Reconstructed Scoliotic Spines, *Proc. IV<sup>th</sup> Symposium On Computer Simulation in Biomechanics*, Montlignon, France, 1.6–1.9.
- DANTZIG, G.B. (1963). *Linear Programming and Extensions*. Princeton University Press, Princeton.
- DARUWALLA, J.S., BALASUBRAMANIAM, P.P. (1985). Moire topography in scoliosis : Its accuracy in detecting the site and size of the curve. *Journal of Bone & Joint Surgery*, 67B, 211–213.
- DAVIS, H.F. (1963). *Fourier Series and Orthogonal Functions*. Allyn & Bacon, Boston.
- DRERUP, B., HIERHOLZER, E. (1996). Assessment of scoliotic deformity from back shape asymmetry using an improved mathematical model. *Clin Biomech*, 11(7), 376–383.
- DYBOWSKI, R., WELLER, P., CHANG, R., GRANT, V. (1996). Prediction of Outcome in Critically Ill Patients Using Artifical Neural Network Synthesized By Genetic Algorithm. *Lancet*, 347(9009), 1146–1150.
- FLEUTE, M., LAVALLÉE, S., JULLIARD, R. (1999). Incorporating a statistically based shape model into a system for computer-assisted anterior cruciate ligament surgery. *Medical Image Analysis*, 3(3), 209–222.

FLETCHER, R. (1987). Practical Methods of Optimization. Second Edition. Wiley, Chichester.

FROBIN, W., HIERHOLZER, E. (1983). Rasterstereography : a photogrammetric method for measurement of body surfaces. *Journal of Biological Photography*, 51(1), 11–17.

GOLDBERG, M.S., POITRAS, B., MAYO, N.E., LABELLE, H., BOURASSA, R., CLOUTIER, R. (1988). Observer variation in assessing spinal curvature and skeletal development in adolescent idiopathic scoliosis. *Spine*, 13(12), 1371–1377.

GOLDBERG, C.J., LALISZER, M., MOORE, D.P., FOGARTY, E.E., DOWLING, F.E. (2001). Surface Topography, Cobb Angles, and Cosmetic Change in Scoliosis. *Spine*, 26(4), E55–E63.

GREEN, A.A., BERMAN, M., SWITZER, P., CRAIG, M.D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans Geosci Remote Sens*, 26(1), 65–74.

GREINER, K.A. (2002). Adolescent Idiopathic Scoliosis : Radiologic Decision-Making. *Am Fam Physician*, 65(9), 1817–1822.

GUNN, S.R. (1998). Support Vector Machines for Classification and Regression. Technical Report, School of Electronics and Computer Science. University of Southampton, Southampton, U.K.

HILGER, K.B., LARSEN, R., WROBEL, M.C. (2003). Statistical shape analysis using non-Euclidean metrics. *Medical Image Analysis*, 7, 417–423.

HOWELL, F.R., DICKSON, R.A. (1989) The deformity of idiopathic scoliosis made visible by computer graphics. *J Bone Joint Surg Br*, 71(3), 399–403.

HUANG, S.C. (1997). Cut-off point of the scoliometer in school scoliosis screening. *Spine*, 22(7), 1985–1989.

HYVÄRINEN, A., KARHUNEN, J., OJA, E. (2001). Independent Component Analysis. Wiley, New York.

JAREMKO, J.L. (2001a). Estimation of Scoliosis Severity from the Torso Surface by Neural Networks. Doctoral thesis, University of Calgary.

JAREMKO, J.L., PONCET, P., RONSKY, J.L., HARDER, J., DANSEREAU, J., LABELLE, H., ZERNICKE, R.F. (2001b). Estimation of Spinal Deformity in Scoliosis From Torso Surface Cross Sections. *Spine*, 26(14), 1583–1591.

JAREMKO, J.L., PONCET, P., RONSKY, J.L., HARDER, J., DANSEREAU, J., LABELLE, H., ZERNICKE, R.F. (2002a). Genetic algorithm-neural network estimation of Cobb angle from torso asymmetry in scoliosis. *J Biomech Eng*, 124 : 496–503.

JAREMKO, J.L., PONCET, P., RONSKY, J.L., HARDER, J., DANSEREAU, J., LABELLE, H., ZERNICKE, R.F. (2002b). Indices of torso asymmetry related to spinal deformity in scoliosis. *Clin Biomech*, 17, 559–568.

JEFFERSON, M.F., PENDLETON, N., LUCAS, S.B., HORAN, M.A. (1997). Comparison of a Genetic Algorithm Neural Network with Logistic Regression for Predicting Outcome After Surgery for Patients with Nonsmall Cell Lung Carcinoma. *Cancer*, 79(7), 1338–1342.

JEFFERSON, M.F., PENDLETON, N., MOHAMED, S., KIRKMAN, E., LITTLE, R.A., LUCAS, S.B., HORAN, M.A. (1998a). Prediction of Hemorrhagic Blood Loss with a Genetic Algorithm Neural Network. *J Appl Physiol*, 84(1), 357–361.

JEFFERSON, M.F., PENDLETON, N., LUCAS, C.P., LUCAS, S.B., HORAN, M.A. (1998b). Evolution of Artificial Neural Network Architecture : Prediction of Depression After Mania. *Meth Inf Med*, 37(3), 220–225.

JOLLIFFE, I.T. (2002). Principal Component Analysis. Second Edition. Springer-Verlag, New York.

KECMAN, V. (2001). Learning and Soft Computing : Support Vector Machines, Neural Networks, and Fuzzy Logic Models. MIT Press, Cambridge, Massachusetts.

KOROVESSIS, P.G., STAMATAKIS, M.V. (1996). Prediction of scoliotic Cobb angle with the use of the scoliometer. *Spine*, 21(14), 1661–1666.

KWOK, J.T., TSANG, I.W. (2003). Linear Dependency Between  $\epsilon$  and the Input Noise in  $\epsilon$ -Support Vector Regression. *IEEE Transactions on Neural Networks*, 14(3), 544–553.

LABELLE, H., DANSEREAU, J., BELLEFLEUR, C., JEQUIER, J.C. (1995). Variability of geometric measurements from three-dimensional reconstructions of scoliotic spines and rib cages. *Eur Spine J*, 4(2), 88–94.

LARSEN, R., HILGER, K.B. (2003). Statistical shape analysis using non-Euclidean metrics. *Medical Image Analysis*, 7, 417–423.

LAULUND, T., SOJBJERG, J.O., HORLYCK, E. (1982). Moire topography in school screening for structural scoliosis. *Acta orthopaedica Scandinavia*, 53, 765–768.

LEVY, A.R., GOLDBERG, M.S., MAYO, N.E., HANLEY, J.A., POITRAS, B. (1996). Reducing the Lifetime Risk of Cancer from Spinal Radiographs Among People With Adolescent Idiopathic Scoliosis. *Spine*, 21(13), 1540–1548.

LI, H., CHUTATAPE, O. (2004). Automated Feature Extraction in Color Retinal Images by a Model Based Approach. *IEEE Trans. PAMI*, 51(2), 246–254.

LUI, X., THOMETZ, J., LYON, R., KLEIN, J. (2001). Functional classification of patients with idiopathic scoliosis assessed by the Quantec system. *Spine*, 26(11), 1274–1279.

MATHWORKS, The, (2001). Matlab : The Language of Technical Computing, Version 6. The Mathworks inc., Natick, MA.

MARTÍN-GUERRERO, J.D., CAMPS-VALLS, G., SORIA-OLIVAS, E., SERRANO-LÓPEZ, A.J., PÉREZ-RUIXO, J.J., JIMÉNEZ-TORRES, N.V. (2003). Dosage Individualization of Erythropoietin Using a Profile-Dependent Support Vector Regression. *IEEE Trans. Biomedical Engineering*, 50(1), 1136–1142.

MARZAN, G.T. (1976). Rational design for close-range photogrammetry. Doctoral thesis, Department of Civil Engineering, University of Illinois at Urbana–Champaign.

MATSUMOTO, T., KITAHARA, H., MINAMI, S., TAKAHASHI, K., YAMAGATA, M., MORIYA, H., TAMAKI, T. (1997). Flexibility in the scoliotic spine : three-dimensional analysis. *J Spinal Disord*, 10(2), 125–131.

MCINTOSH, A.R., BOOKSTEIN, F.L., HAXBY, J.V., GRADY, C.L. (1996). Spatial Pattern Analysis of Functional Brain Images Using Partial Least Squares. *Neuroimage*, 3, 143–157.

MORELAND, M.S., POPE, M.H., WILDER, D.G., STOKES, I.A.F., FRYMOTER, J.W. (1981). Moire fringe topography of the human body. *Medical Instrumentation*, 15(2), 129–132.

MORRISSEY, R.T., GOLDSMITH, G.S., HALL, E.C., KEHL, D., COWIE, G.H. (1990). Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. *J Bone Joint Surg Am*, 72(3), 320–327.

PEARCY, M.J. (1985). Stereo radiography of lumbar spine motion. *Acta Orthop Scand (suppl)*, 56, 1–45.

PLAMONDON, A., GAGNON, M. (1990). Evaluation of Euler's angles with a least squares method for the study of lumbar spine motion. *J Biomed Eng*, 12, 143–149.

PONCET, P., DELORME, S., RONSKY, J.L., DANSEREAU, J., HARDER, J., CLYNCH, G., DEWAR, R.O., LABELLE, H., GU, P.H., ZERNICKE, R.F. (2000). Reconstruction of Laser-Scanned 3D Torso Topography and Stereo-Radiographical Spine and Rib-Cage Geometry in Scoliosis. *Computer Methods in Biomechanics & Biomedical Engineering*, 4(1), 59–75.

PONCET, P. (2001). La torsion géométrique dans la scoliose idiopathique. Thèse de doctorat, École Polytechnique de Montréal.

RAMSAY, J.O., SILVERMAN, B.W. (1997). *Functional Data Analysis*. Springer, New York.

RAMSAY, J.O., SILVERMAN, B.W. (2001). *Applied Functional Data Analysis : Methods and Case Studies*. Springer, New York.

REAMY, B.V., SLAKEY, J.B. (2001). Adolescent Idiopathic Scoliosis : Review and Current Concepts. *Am Fam Physician*, 64(1), 111–116.

RICE, J.R. (1964). The Approximation of Functions. Volume 1 : Linear Theory. Addison-Wesley, Reading.

RISSER, J.C. (1958). The iliac apophysis : an invaluable sign in the management of scoliosis. *Clin Orthop*, 11, 111–118.

SAMUELSSON, L., NOREN, L. (1997). Trunk rotation in scoliosis. The influence of curve type and direction in 150 children. *Acta Orthop Scand*, 63(3), 273–276.

SCHÖLKOPF, B., SMOLA, A.J. (2002). Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, Massachusetts.

SLUPSKY, S.H., DURDLE, N.G., RASO, V.J. PETERSON, A.E. (1992). A video digitizer for analysis of trunk deformity in scoliosis. *J Biomed Eng*, 14, 405–408.

STAIB, L.H., DUNCAN, J.S. (1992). Boundary Finding with Parametrically Deformable Models. *IEEE Trans. PAMI*, 14(11), 1061–1075.

STOKES, I.A.F., BIGALOW, L.C., MORELAND, M.S. (1987). Three-dimensional spinal curvature in idiopathic scoliosis. *Journal of Orthopaedic Research*, 5, 102–113.

STOKES, I.A.F., ARMSTRONG, J.G., MORELAND, M.S. (1988). Spinal deformity and back surface asymmetry in scoliosis. *Journal of Orthopaedic Research*, 6, 129–137.

STOKES, I.A.F. (1989). Axial rotation component of thoracic scoliosis. *Journal of Orthopaedic Research*, 7, 702–708.

STOKES, I.A.F., MORELAND, M.S. (1989). Concordance of back surface asymmetry and spine shape in idiopathic scoliosis. *Spine*, 14(1), 73–78.

SUZUKI, N., ONO, T., TEZUKA, M., KAMIISHI, S. (1992). Moiré Topography and Back Shape Analysis : Clinical Application. *Proceedings of the International Symposium on 3D Scoliotic Deformities*, Montréal, 124–130.

SWAMY, M.N.S. (2002). Engineering Analysis. Concordia University, Montréal.

SZÉKELY, G., KELEMAN, A., BRECHBÜHLER, C., GERIG, G. (1996). Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models. *Medical Image Analysis*, 1(1), 19–34.

TARDIF, N., PONCET, P., RONSKY, J.L., DANSEREAU, J., ZERNICKE, R.F. (2000). Evaluation of an integrated laser imaging/X-ray technique for torso asymmetry measurement in scoliosis. *Arch Physiol Biochem*, 108–200.

THALANGE, N.K., FOSTER, P.J., GILL, M.S., PRICE, D.A., CLAYTON, P.E. (1996). Model of normal prepubertal growth. *Archives of Disease in Childhood*, 75, 427–431.

THEOLOGIS, T., FAIRBANK, J., TURNER-SMITH, A., PANTASOPOULOS, T. (1997). Early detection of progression in adolescent idiopathic scoliosis by measurement of changes in back shape with the Integrated Shape Imaging System scanner. *Spine*, 22, 1223–1228.

THOMETZ, J.G., LAMDAN, R., LIU, X.C., LYON, R. (2000). Relationship between Quantec measurement and Cobb angle in patients with idiopathic scoliosis. *J Pediatr Orthop*, 20(4), 512–516.

TREDWELL, S.J., BANNON, M. (1988). The use of the ISIS optical scanner in the management of the braced adolescent idiopathic scoliosis patient. *Spine*, 13(10), 1104–1105.

TSOUKALAS, L.H., UHRIG, R.E. (1997). *Fuzzy and Neural Approaches in Engineering*. Wiley, New York.

TUDDENHAM, R.D., SNIDER, M.M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development*, 1, 183–364.

TURNER-SMITH, A.R., HARRIS, J.D., HOUGHTON, G.R., JEFFERSON, R.J. (1988). A method for analysis of back shape in scoliosis. *Journal of Biomechanics*, 21(6), 497–509.

UPADHYAY, S.S., BURWELL, R.G., WEBB, J.K. (1988). Hump changes on forward flexion of the lumbar spine in patients with idiopathic scoliosis : A study using ISIS and the Scoliometer in two standard positions. *Spine*, 13(2), 146–151.

VANDEGRIEND, B., HILL, D., RASO, J., DURDLE, N., ZHANG, Z. (1995). Application of computer graphics for assessment of spinal deformities. *Med Biol Eng Comput*, 33(2), 163–166.

VAPNIK, V.N. (1998). Statistical learning theory. Wiley, New York.

WANG, L. (2003). Classification of scoliotic deformities from external surface of the trunk using support vector machines. Mémoire de maîtrise, École Polytechnique de Montréal.

WEISZ, I., JEFFERSON, R.J., TURNER-SMITH, A.R., HOUGHTON, G.R., HARRIS, J.D. (1988). ISIS Scanning : A useful assessment technique in the management of scoliosis. *Spine*, 13(4), 405–408.

YLIKOSKY, M., TALLROTH, K. (1990). Measurement variations in scoliotic angle, vertebral rotation, vertebral body height, and intervertebral disc space height. *J Spinal Disord*, 3, 387–391.