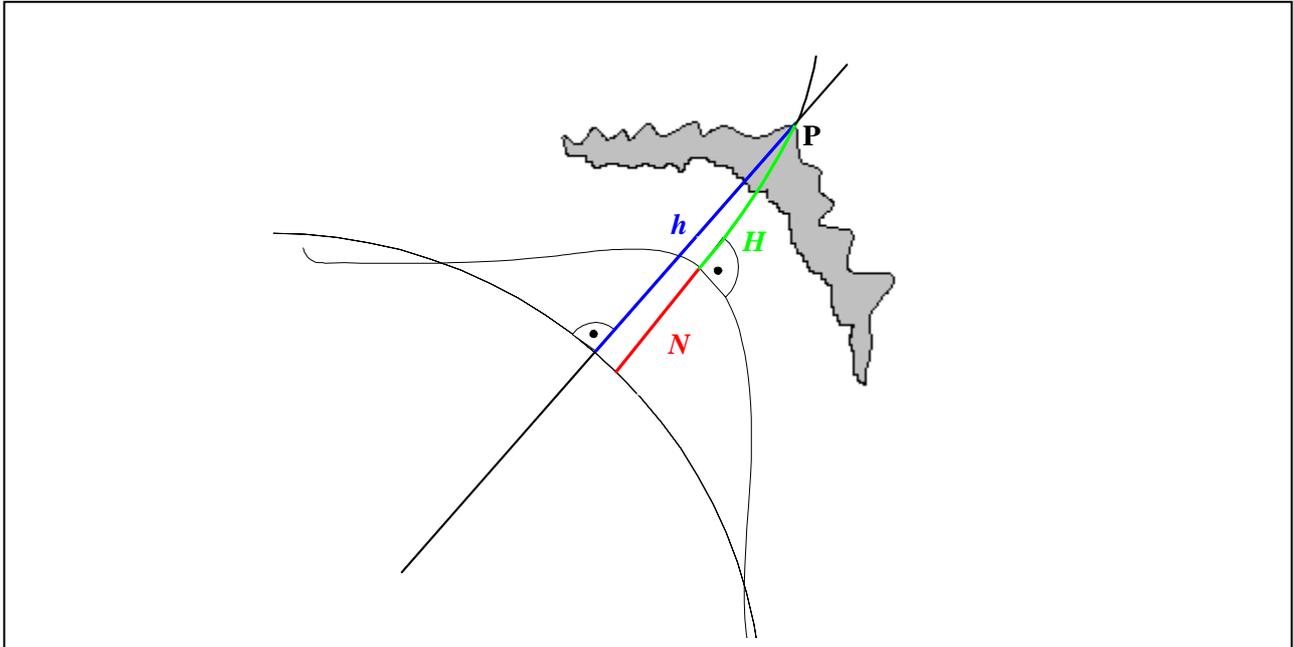


Universität Stuttgart

Geodätisches Institut



A regional analysis of GNSS-levelling



Diplomarbeit im Studiengang
Geodäsie und Geoinformatik
an der Universität Stuttgart

Carina Raizner

Stuttgart, Februar 2008

Betreuer:

Prof. Dr.-Ing. Georgia Fotopoulos
University of Toronto, Canada

Dr. Mohammad Sharifi
University of Tehran, Iran

Prüfer:

Prof. Dr.-Ing. Nico Sneeuw
Universität Stuttgart

Selbständigkeitserklärung

Hiermit erkläre ich, Carina Raizner, dass ich die von mir eingereichte Diplomarbeit zum Thema

A regional analysis of GNSS-levelling

selbständig verfasst und ausschließlich die angegebenen Hilfsmittel und Quellen verwendet habe.

Datum, Ort: _____

Unterschrift: _____
(Carina Raizner)

Abstract

The traditional technique used to obtain precise height information is known as spirit-levelling. With the improved accuracy of the geoid provided, in part, by the latest geopotential models derived from CHAMP and GRACE (and in the future GOCE), a modern height measurement technique called GNSS-levelling can be contemplated as an alternative for practical height applications. The proposed methodology for determining orthometric heights using a new vertical datum based on a geoid model (regional or global) and ellipsoidal heights from Global Navigation Satellite Systems (GNSS) is described. The main interest lies in a detailed analysis of the optimal combination of heterogeneous data consisting of ellipsoidal, orthometric and geoid heights of a vertical control network. Theoretically, the relationship between the height triplet is based on a simple linear equation. However, due to numerous factors that cause discrepancies among the combined height data such as datum inconsistencies and systematic errors, a parametric “corrector” surface model is incorporated. The performance of selected parametric models is assessed by several empirical and statistical tests. In order to properly account for correlation among model parameters, the Gram-Schmidt orthonormalization procedure was applied and test results were compared with orthogonal and non-orthogonal base functions. In addition, variance component estimation is applied to the common adjustment of the height triplet to test the suitability of the stochastic model, i.e., to separate random errors and to estimate the variance components for each height type. A detailed study of the appropriate data screening methods for the sake of gross error detection and elimination is also performed.

Finally, the study includes an implementation of radial basis function interpolants. They provide an orthonormal basis which determines a parametric surface with uncorrelated parameters. At the same time, the corresponding performance of the surface is not dependent on the degree of development of the base function. The optimization of the shape parameter inherent in the interpolation kernel is realized by means of a genetic algorithm (GA) in order to minimize the interpolation error. This methodology has the advantage that it is differentiable. Thus, the data distribution can be taken into account and any possible oscillations in the parametric surface are avoided. Among others, the performance of this inverse multiquadric function is assessed by a special cross-validation procedure which has a higher performance than the classic ones.

The described approaches are evaluated using the most current geoid model, GPS and levelling data in Canada and its regions.

Keywords:

GNSS-levelling · variance component estimation · gross error detection · radial basis function · genetic algorithm

Zusammenfassung

Die herkömmliche Methode zur präzisen Höhenbestimmung ist das Nivellement. Durch die verbesserte Genauigkeit des Geoids, das von den neuesten und von CHAMP und GRACE (und in Zukunft GOCE) abgeleiteten Modellen des Gravitationspotentials bereitgestellt wird, kann das sogenannte „GNSS-levelling“ als eine Alternative für praktische Höhenanwendungen in Betracht gezogen werden. Der Schwerpunkt dieser Arbeit liegt in einer detaillierten Analyse einer optimalen Kombination von heterogenen Daten, die sich aus Ellipsoid-, orthometrischen und Geoidhöhen eines Höhenfestpunktnetzes zusammensetzen. Das Verhältnis zwischen diesen drei verschiedenartigen Höhen basiert theoretisch auf einer einfachen linearen Gleichung. Nachdem aber zahlreiche Faktoren, wie z.B. Datumsinkonsistenzen und systematische Fehler, zu Abweichungen zwischen diesen kombinierten Höhendaten führen, wird ein Modell als Korrekturfläche integriert. Die Eignung ausgewählter Modelle wird anhand verschiedener empirischer und statistischer Tests beurteilt. Um der Korrelation zwischen den Modellparametern entgegenzuwirken, wird das Gram-Schmidtsche Orthonormalisierungsverfahren angewandt und die Testergebnisse mit jenen von orthogonalen und nicht-orthogonalen Basisfunktionen verglichen. Zusätzlich wird eine Varianzkomponentenschätzung auf die kombinierte Ausgleichung der drei Höhen angewandt, um die Eignung des stochastischen Modells zu überprüfen, d.h. um zufällige Fehler zu separieren und die zugehörigen Varianzkomponenten für jeden Höhentyp zu schätzen. Außerdem wird eine detaillierte Studie über geeignete Methoden zur Überprüfung der Daten durchgeführt, um grobe Fehler zu entdecken und zu eliminieren. Schließlich beinhaltet diese Arbeit auch die Implementierung einer Interpolation mit radialen Basisfunktionen. Diese stellen eine orthonormale Basis dar, mit der eine Oberfläche mit unabhängigen Parametern bestimmt werden kann. Gleichzeitig hängt die entsprechende Effizienz der Fläche nicht vom Entwicklungsgrad der Basisfunktionen ab. Die Optimierung des in dem Interpolationskern enthaltenen Formparameters wird mittels eines genetischen Algorithmus (GA) realisiert, um den Interpolationsfehler zu minimieren. Diese Methode bietet den Vorteil, dass sie differenzierbar ist und somit kann die Verteilung der Daten berücksichtigt werden. Mögliche Oszillationen in der Parameterfläche können vermieden werden. Die Wirkungsweise dieser inversen multiquadratischen Funktion wird durch eine spezielle Kreuzvalidierung beurteilt, welche eine höhere Effizienz aufweist als das klassische Verfahren. Die beschriebenen Ansätze werden anhand des aktuellen Geoidmodells und anhand von GPS und Nivellierdaten in Kanada und einzelnen Teilregionen evaluiert.

Schlüsselwörter:

GNSS-levelling · Varianzkomponentenschätzung · Ausreißertest · radiale Basisfunktionen · genetischer Algorithmus

Preface

During my study term at the University of Calgary, Canada, I met Dr. G. Fotopoulos who offered me the excellent opportunity to join her as summer research student at the Department of Civil Engineering, University of Toronto. Thus, the research for the first part of the thesis, i.e., GNSS-levelling with parametric models and the development of the software *OrthoPlus*, was conducted under her supervision from May until August 2006. I also had the unique chance to attend the 1st International Symposium of the International Gravity Field Service in Istanbul, Turkey, from 28th of August until 1st of September 2006. I presented the results as a poster under the title “The modern vertical datum and GNSS-levelling”. Two months later, I gave an oral presentation about “An investigation on the use of GNSS-levelling for practical height determination” at the Geodetic Week in Munich in October 2006. At this conference, I met Dr. M. Sharifi who had the idea to combine radial basis functions and a genetic algorithm with GNSS-levelling. Under his supervision and with the support of my supervisors, I was able to continue the thesis about “A regional analysis of GNSS-levelling” at the University of Stuttgart.

First of all, I want to express my deepest gratitude and appreciation to my supervisor at the University of Toronto, Dr. G. Fotopoulos, for her continuous support, her advices and answers to countless questions. By arranging a summer research work term at the UofT, she made the studies for this thesis possible and provided me with an unforgettable summer. I want to extend my thanks to my supervisor at the University of Stuttgart, Dr. M. Sharifi (University of Tehran). His guidance and encouragement for the second part of this work are highly appreciated. Furthermore, I want to sincerely thank my professor at the University of Stuttgart, Dr. N. Sneeuw, Head of Institute of Geodesy, for his additional comments, patience and his supervision.

I am also very grateful to my dear friends, Dr. M. Weigelt and B. Devaraju. Their continuous encouragement, useful comments and discussions as well as the thorough proofreading for improving the quality of the study are highly appreciated. I especially want to thank them and all my other friends from Calgary, Toronto and Stuttgart for their friendship as well as the good times and memories.

Last but not least, I am very grateful to my dear parents Rainer and Barbara, my sister Julia and my brother Peter. With their unceasing love, endless encouragement, moral support and faith in me, they gave me enough strength to finish this thesis as well as my diploma studies.

List of Tables

Table 2.1: Defining parameters of the GRS 1980.....	8
Table 2.2: Summary of the main characteristics of the orthometric and normal heights	16
Table 3.1: Overview and descriptions of functions called in <i>OrthoPlus.m</i>	20
Table 3.2: Summary of parametric surface models	42
Table 4.1: Comparison between global test & data snooping and τ -test	59
Table 4.2: RMS values for classical empirical approach and cross-validation procedure.....	65
Table 4.3: Number of insignificant parameters with and without Gram-Schmidt orthonormalization ..	67
Table 4.4: Characteristics of the initial covariance matrices	71
Table 4.5: Final variance components for IMINQUE after 200 iterations	74
Table 4.6: Final variance components for IAUE with N separated in components.....	76
Table 4.7: Final variance components for IAUE with summarized N	77
Table 4.8: Final variance components and average standard deviations of CV matrices.....	79
Table 4.9: Effect of correlations on final variance components	79
Table 4.10: Basic network characteristics of four different regions in Canada.....	81
Table 4.11: Comparison of the results from the regional analysis.....	90
Table 5.1: Examples of more commonly employed RBF kernels	94
Table 5.2: Key terms within biological and numerical optimization (GA)	102
Table 5.3: Specifications of algorithm options	109
Table 5.4: Comparison of RMS values for cross-validation (one-leave-out method)	113
Table 5.5: Comparison of RMS values for check-point validation	114
Table 5.6: Comparison of RMS values for single-step scheme.....	115
Table 5.7: Comparison of RMS values for combined evaluation.....	115
Table 5.8: Comparison of RMS values from single-step scheme (with GA) and <i>OrthoPlus</i>	120
Table 5.9: Comparison of RMS values for combined evaluation (with GA)	121

List of Figures

Figure 1.1: Relationship between ellipsoidal, orthometric and geoid heights	2
Figure 2.1: Reference ellipsoid and geodetic coordinates	8
Figure 2.2: Definition of the orthometric height.....	12
Figure 2.3: Definition of the normal height.....	14
Figure 3.1: Summary of modules in <i>OrthoPlus</i>	19
Figure 3.2: Flowchart for 3-RMS test.....	23
Figure 3.3: Flowchart for post-adjustment data screening.....	25
Figure 3.4: Flowchart of major steps of parametric surface fit.....	35
Figure 3.5: 3D parametric model... ..	35
Figure 3.6: Projection of model in area map.....	45
Figure 3.7: Example of height misclosures before and after parametric model fit.....	45
Figure 3.8: Iterative VCE computational scheme.....	53
Figure 3.9: Convergence behaviour of variance estimates	54
Figure 4.1: GPS-levelling benchmark network of Canada	57
Figure 4.2: Possible blunders detected in the Canadian dataset	60
Figure 4.3: parametric models A – I and surface fit of models on grid	63
Figure 4.4: RMS values for classical empirical approach and cross-validation procedure	65
Figure 4.5: Statistical measures of goodness of fit	66
Figure 4.6: Specification of insignificant terms.....	67
Figure 4.7: Parametric surface fit before and after Gram-Schmidt orthonormalization	68
Figure 4.8: Condition number of parametric models.....	69
Figure 4.9: Initial covariance matrix for GPS heights	72
Figure 4.10: Initial covariance matrix for orthometric heights.....	72
Figure 4.11: Initial covariance matrix for geoid heights.....	72
Figure 4.12: Variance estimates of IMINQUE algorithm.....	73
Figure 4.13: Variance estimates of IAUE algorithm	76
Figure 4.14: Variance estimates and iterations (fully-populated CV matrices).....	80
Figure 4.15: Variance estimates and iterations (diagonal CV matrices).....	80
Figure 4.16: Division of Canadian dataset into four different regions	82
Figure 4.17: Values for precision and accuracy within Western Canada	82
Figure 4.18: Statistical measures of goodness of fit for Western Canada	83
Figure 4.19: Third-order polynomial model and its parametric surface fit to Western Canada	84
Figure 4.20: Values for precision and accuracy within Central Canada.....	85
Figure 4.21: Statistical measures of goodness of fit for Central Canada.....	85

Figure 4.22: Classic four-parameter model and its parametric surface fit to Central Canada	86
Figure 4.23: Values for precision and accuracy within Eastern Canada	87
Figure 4.24: Statistical measures of goodness of fit for Eastern Canada.....	87
Figure 4.25: Classic four-parameter model and its parametric surface fit to Eastern Canada.....	87
Figure 4.26: Values for precision and accuracy within the Maritimes	88
Figure 4.27: Statistical measures of goodness of fit for the Maritimes	89
Figure 4.28: Classic four-parameter model and its parametric surface fit to the Maritimes	89
Figure 5.1: Flowchart of a continuous GA	104
Figure 5.2: Two-point crossover.....	107
Figure 5.3: Reproduction options	110
Figure 5.4: Scheme of applying radial basis functions without optimization.....	112
Figure 5.5: RMS (combined evaluation, varying single shape parameters) for whole Canada.....	116
Figure 5.6: RMS (combined evaluation, varying single shape parameters) for Western Canada	117
Figure 5.7: RMS (combined evaluation, varying single shape parameters) for Central Canada.....	117
Figure 5.8: RMS (combined evaluation, varying single shape parameters) for Eastern Canada.....	118
Figure 5.9: RMS (combined evaluation, varying single shape parameters) for the Maritimes	118
Figure 5.10: Scheme of applying radial basis functions with GA optimization	119
Figure 5.11: RBF model with GA optimized shape parameter and fit to the Canadian network	122
Figure 5.12: RBF model with GA optimized shape parameter and fit to Western Canada.....	123
Figure 5.13: RBF model with GA optimized shape parameter and fit to Central Canada	123
Figure 5.14: RBF model with GA optimized shape parameter and fit to Eastern Canada	123
Figure 5.15: RBF model with GA optimized shape parameter and fit to the Maritimes	124

Table of Contents

SELBSTÄNDIGKEITSERKLÄRUNG	II
ABSTRACT	III
ZUSAMMENFASSUNG	IV
PREFACE.....	V
LIST OF TABLES	VI
LIST OF FIGURES	VII
1 INTRODUCTION.....	1
1.1 MOTIVATION.....	1
1.2 STATEMENT OF PROBLEM - THESIS OBJECTIVES.....	4
1.3 THESIS OUTLINE	5
2 BACKGROUND: HEIGHTS.....	7
2.1 ELLIPSOIDAL HEIGHTS.....	7
2.2 GEOPOTENTIAL NUMBERS	10
2.3 DYNAMIC HEIGHTS.....	11
2.4 ORTHOMETRIC HEIGHTS	11
2.5 NORMAL HEIGHTS	13
2.6 COMPARISON OF DIFFERENT HEIGHT SYSTEMS	15
2.7 HEIGHT DETERMINATION.....	17
2.7.1 <i>Spirit-levelling</i>	17
2.7.2 <i>GNSS-levelling</i>	17
3 THEORY AND IMPLEMENTATION	19
3.1 GENERAL INFORMATION TO <i>ORTHOPLUS</i>	20
3.2 DATA SCREENING, GROSS ERROR DETECTION AND ELIMINATION.....	21
3.2.1 <i>Pre-adjustment data screening</i>	22
3.2.2 <i>Post-adjustment data screening</i>	23
3.2.2.1 Global test on the a-posteriori variance factor.....	26
3.2.2.2 Data snooping of individual residuals	29
3.2.2.3 τ -test.....	32
3.3 PARAMETRIC MODEL SURFACE FIT	34
3.3.1 <i>Combined height adjustment</i>	36
3.3.2 <i>Selection of parametric model</i>	38
3.3.2.1 Polynomial models up to 4 th order	38
3.3.2.2 Similarity-based transformation model and simplified forms	39
3.3.2.3 Orthogonal polynomials	39
3.3.2.4 Summary	42
3.3.3 <i>Gram-Schmidt orthonormalization</i>	42
3.3.4 <i>Assessing the parametric model performance</i>	44
3.3.4.1 Classical empirical approach.....	45
3.3.4.2 Assessing the goodness of fit	46
3.3.4.3 Cross-validation	47
3.3.4.4 Testing parameter significance.....	48
3.4 VARIANCE COMPONENT ESTIMATION (VCE).....	49
3.4.1 <i>Iterative minimum norm quadratic unbiased estimation (IMINQUE)</i>	54
3.4.2 <i>Iterative almost unbiased estimation (IAUE)</i>	56

4	DISCUSSION OF RESULTS.....	57
4.1	GPS-LEVELLING DATASET.....	57
4.2	EVALUATION OF BLUNDER DETECTION.....	58
4.3	ASSESSMENT OF PARAMETRIC MODEL PERFORMANCE.....	60
4.3.1	<i>Comparison between different RMS values.....</i>	<i>64</i>
4.3.2	<i>Goodness of fit.....</i>	<i>66</i>
4.3.3	<i>Parameter significance test and model stability.....</i>	<i>67</i>
4.3.4	<i>Conclusion of performance assessment.....</i>	<i>69</i>
4.4	RESULTS OF VARIANCE COMPONENT ESTIMATION.....	70
4.4.1	<i>Initial covariance matrices.....</i>	<i>70</i>
4.4.2	<i>IMINQUE.....</i>	<i>73</i>
4.4.3	<i>IAUE.....</i>	<i>75</i>
4.4.3.1	<i>IAUE with N separated in components.....</i>	<i>75</i>
4.4.3.2	<i>IAUE with summarized N.....</i>	<i>77</i>
4.5	REGIONAL ANALYSIS.....	81
4.5.1	<i>Results for the test network of Western Canada.....</i>	<i>82</i>
4.5.2	<i>Results for the test network of Central Canada.....</i>	<i>84</i>
4.5.3	<i>Results for the test network of Eastern Canada.....</i>	<i>86</i>
4.5.4	<i>Results for the test network of the Maritimes.....</i>	<i>88</i>
4.5.5	<i>Comparison and overview.....</i>	<i>90</i>
5	RADIAL BASIS FUNCTIONS AND GENETIC ALGORITHM.....	92
5.1	RADIAL BASIS FUNCTIONS.....	92
5.2	COMPUTATION OF THE SHAPE PARAMETER FOR A RBF KERNEL.....	95
5.3	EVALUATION OF THE SHAPE PARAMETER.....	96
5.3.1	<i>Cross-validation procedure (one-leave-out method).....</i>	<i>96</i>
5.3.2	<i>Evaluation at check points.....</i>	<i>97</i>
5.3.3	<i>Modified cross-validation procedure: single-step scheme.....</i>	<i>98</i>
5.3.4	<i>Combined evaluation.....</i>	<i>99</i>
5.4	OPTIMIZATION ALGORITHM FOR THE SHAPE PARAMETER.....	100
5.4.1	<i>Theoretical overview of the genetic algorithm (GA).....</i>	<i>101</i>
5.4.1.1	<i>Introduction.....</i>	<i>101</i>
5.4.1.2	<i>Terminology.....</i>	<i>102</i>
5.4.1.3	<i>Methodology.....</i>	<i>103</i>
5.4.2	<i>Practical implementation of the shape parameter optimization.....</i>	<i>108</i>
5.5	DISCUSSION OF RESULTS.....	111
5.5.1	<i>RBF without using GA.....</i>	<i>111</i>
5.5.2	<i>RBF with GA optimization.....</i>	<i>119</i>
6	CONCLUSIONS AND FUTURE WORK.....	125
	REFERENCES.....	131

1 Introduction

1.1 Motivation

One major goal of geodesy is the precise determination of the shape of the Earth and of its regions. The outcome does not only play a major role in all Earth sciences but also in other applications such as cartography, oceanography, civil engineering, exploitation, high-precision surveys and last but not least geographical information systems (GIS). At the same time, the precise measurement of physical heights, commonly known as a height above mean sea-level, is one of the most difficult objectives in geodesy as the knowledge of the “physical figure of the Earth”, i.e., the Earth’s gravity field, is additionally required. Therefore, it is much more demanding than the determination of two- or three-dimensional coordinates of points on the Earth’s surface which is only based on geometric techniques due to the measurements with respect to the reference ellipsoid. The combination of spirit-levelling and gravity measurements presents the traditional technique which determines physical heights of points with mm-precision in smaller regions. This technique, however, is very time consuming, expensive and makes the maintenance of a vertical control network difficult, especially in areas which are hard to access. Another disadvantage is the loss of precision over longer distances since each height system usually refers to a benchmark point close to the sea which is connected to a tide gauge station representing the mean-sea level. That means that heights within a national height system are usually comparable but that there are unavoidable discrepancies between heights of different systems, especially if they are divided by an ocean or if they are lying on different coasts.

In order to counteract these drawbacks and because of the growing importance of global navigation satellite systems (GNSS) in all kind of geodetic and surveying applications, the modern technique called GNSS-levelling could be considered as an alternative for practical height determination. It is based on the following fundamental equation (Hofmann-Wellenhof and Moritz, 2005):

$$h - H - N = 0, \tag{1.1}$$

where h is the ellipsoidal height with respect to a reference ellipsoid and is nowadays mostly obtained from GPS (global positioning system) measurements. In the sequel, the term ‘GNSS-levelling’ will be replaced by ‘GPS-levelling’ since all ellipsoidal heights used in this thesis are only provided by GPS observations. Furthermore, H in Eq. (1.1) is the orthometric height based on the geoid which is defined as the surface of equal gravitational potential of a hypothetical ocean at rest and is approximated by the mean sea-level (MSL). As the geometrical ellipsoidal height is based on the reference ellipsoid while the physical orthometric height is based on the geoid, the deviation between these two reference surfaces along the ellipsoidal normal has to be taken into account by introducing a third variable into Eq. (1.1), the so-called geoid height or geoid undulation N above the same adopted reference ellipsoid. The geoid heights can be computed from global geopotential models derived from satellite gravity missions such as CHAMP, GRACE and in future GOCE or from regional gravimetric geoid models. These geoid heights can take values up to ± 100 m due to inhomogeneous mass distribution within the Earth, for example. Thus, the physical height H can be either directly determined by spirit-levelling and gravity measurements as already mentioned or by converting the ellipsoidal height h into the orthometric height H by applying a slightly modified form of Eq. (1.1) with the geoid height N .

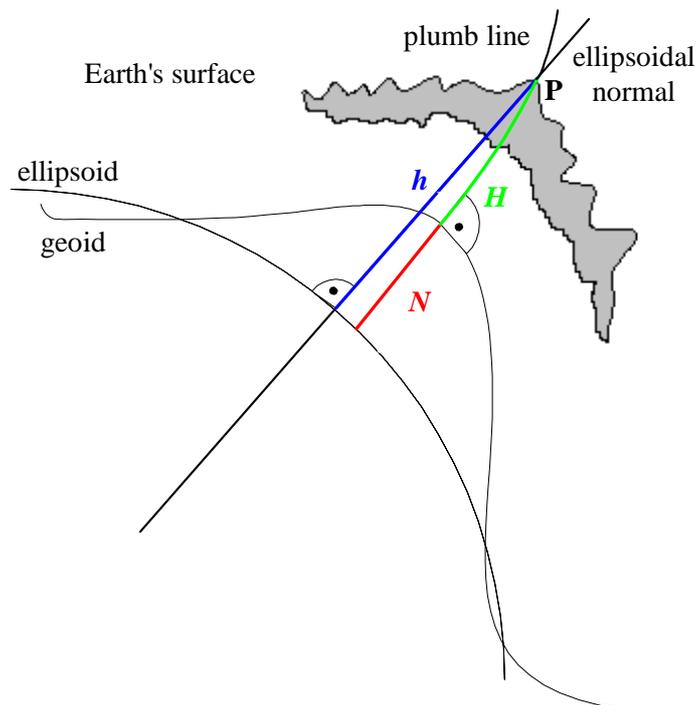


Figure 1.1: Relationship between ellipsoidal (h), orthometric (H) and geoid heights (N)

The geometrical relationship between the height triplet is visualized in Figure 1.1. It seems to be quite simple and implies that one of the three height types can be computed if the other two height types are

given. That means that GPS-levelling can be analogously used for local geoid height determination, for example. Furthermore, in areas with existing spirit-levelled heights, GPS-levelling will not replace them but rather serve as height control possibility.

Improving the accuracy of GPS measurements, on the one hand, and the accuracy of the geopotential models, on the other, will simultaneously lead to more accurate GPS-levelled heights. With the upcoming launch of the new satellite gravity mission GOCE (gravity field and steady-state ocean circulation explorer) in spring 2008, new progresses are expected with respect to an improved accuracy of the geoid heights and with more precise orthometric heights determined by GPS-levelling. The usage of GPS-levelling and the global unification of height systems which enables the direct comparison of world-wide sea-levels or of mountain ranges on one continent against those on others are explicitly stated as science objectives of the GOCE mission (ESA, 1999). The latter objective will be realized by using orthometric heights referenced to a common GOCE derived geoid. In order to enable the compliance of these science objectives, gravity anomalies and geoid heights will be derived from the measurements. They have to fulfill the following mission requirements:

- Earth's gravity field anomalies with an accuracy of better than 1 mGal
- Geoid with a radial accuracy better than 1-2 cm
- Gravity anomalies and geoid heights with a spatial resolution of 100 km or less

Even if the accuracy of GPS-levelling will still be well below the one of spirit-levelling, especially for short and medium distances, it is sufficient for lower order levelling networks virtually everywhere in the world. It is especially useful in areas without existing vertical control which are difficult to access such as cities with high traffic, mountainous terrain or remote areas. Since GPS-levelling can be seen as a by-product of GPS measurement campaigns, it is more economical than conventional techniques. For practical considerations, the application of Eq. (1.1) is, however, not as simple. It is important to know that the combination of the heterogeneous height triplet will practically never be exactly zero due to various factors such as datum inconsistencies, random errors and systematic effects in the derived heights, time variable effects as well as any assumptions and approximations made in the computations of the heights, among others (Fotopoulos, 2003). In the practical application, these discrepancies, especially the datum inconsistencies and systematic effects, are usually modelled by the incorporation of a parametric surface model, or also known as "corrector" surface model.

Accordingly, the main problem to be researched in this study is how the most appropriate model will look like, i.e., how it can be derived and additionally how the suitability can be assessed. Hereby, this research takes up and extends the work of Fotopoulos (2003) in which an optimal combination of heterogeneous height data is developed and analyzed by the incorporation of parametric models in a

combined height adjustment for the Swiss as well as the Western Canadian network. Fotopoulos is mainly interested in the modelling of systematic errors and datum inconsistencies, the separation of random errors and the estimation of variance components for each height type. The main objective of this study will lie in a refinement of the optimal combination of heterogeneous data consisting of ellipsoidal, orthometric and geoid heights of a vertical control network and the extension to the complete Canadian network.

1.2 Statement of problem - thesis objectives

Due to systematic effects and datum inconsistencies which are not negligible, the incorporation of parametric surface models in the common height adjustment is absolutely essential for practical purposes. Thus, selected parametric models will be applied in a combined least-squares height adjustment over a network of co-located GPS-levelling benchmarks in order to obtain the unknown model parameters. Since the choice and the selection of the most appropriate parametric model is rather complicated and arbitrary, validation procedures will be determined and implemented for a thorough assessment of the model performance. Before and after the adjustment, the dataset and the adjusted residuals will be screened for gross errors as they will influence the reliability of the assessment. Additionally and in order to account for any correlation among model parameters, every chosen model can be converted into an orthogonal base function by applying a Gram-Schmidt orthonormalization. Another very important issue of the common height adjustment is to test the suitability of the stochastic model for the observational noise since it contains only one variance component for all three height types. Thus, in order to improve the stochastic model, the random errors of each height type can be separated by computing different variance components. This objective is commonly realized by the application of a variance component estimation (VCE). All aforementioned tools will be implemented in a semi-automated software called *OrthoPlus*. The task of finding an appropriate parametric model depends very much on the type of height network data, such as data distribution, density and quality and is therefore hard to generalize. Accordingly, the discussion of the results provided by the software will concentrate on the evaluation of several pre-selected parametric models applied to the most current geoid model, GPS and levelling data in Canada and its regions.

Apart from the application of parametric models and for comparison purposes, this study will introduce another type of model in order to interpolate surface data, the radial basis function (RBF) interpolants. The RBF models have the advantage that they initially form an orthonormal basis with uncorrelated parameters, that they are not dependent on the order of the base function and that they take the distribution of data points into account. For the sake of a minimal interpolation error, the unknown

shape parameter in the interpolation kernel will be optimized by a genetic algorithm (GA). The comparison between the performance of an RBF and a parametric model applied to the Canadian network and its regions will be based on the assessment of the differences with respect to accuracy.

1.3 Thesis outline

The study consists mainly of three parts. The first one (chapter 3) deals with the application and assessment of parametric surface models within the combined height adjustment as well as with the implementation of the corresponding software *OrthoPlus*. Its numerical results will be discussed in the second part (chapter 4). Finally, the last part (chapter 5) represents an augmentation to the first part and introduces the interpolation of the height discrepancies by radial basis functions (RBF) together with a short discussion of the results. As this final part represents a totally new and independent approach, it is completely separated from the first two parts.

Before going into details, chapter 2 provides a short overview about different height systems together with a brief comparison. These include the geometric definition of ellipsoidal heights as well as the physical height definitions such as geopotential numbers, dynamic, orthometric and normal heights. The two concepts of spirit-levelling and GNSS-levelling are also explained.

Chapter 3 presents the theory and implementation of the GPS-levelling software *OrthoPlus*. After providing some general information about the software, the methodology and the actual implemented algorithms are described and supported by means of flowcharts. The user can choose between different pre-selected parametric surface models. Each of them belongs to one of the three pre-specified parametric model families which are polynomial models up to 4th order, similarity-based transformation models including its simplified forms and orthogonal polynomials. *OrthoPlus* contains several tools such as pre- and post-adjustment data screening, the combined height adjustment, the Gram-Schmidt orthonormalization and the selection of parametric models based on the assessment of the model performance. The assessment tool is divided into five different statistical and empirical tests applied to the adjustment results: the classical empirical approach, the computation of the goodness of fit, the investigation of the model stability, the cross-validation procedure and the parameter significance test. The last part of this chapter focuses on the description of two different iterative variance component estimation algorithms, the iterative minimum norm quadratic unbiased estimation (IMINQUE) and the iterative almost unbiased estimation (IAUE).

In chapter 4, the tools of the software *OrthoPlus* are tested and demonstrated for the pre-selected parametric models by applying them to the Canadian dataset containing the most current geoid model, GPS and levelling data. The most revealing numerical results for different models are analysed and compared to each other. The discussion of the results concentrates not only on the evaluation of blunder detection and of the VCE algorithms using these Canadian height data and the corresponding covariance matrices, but also tries to find an appropriate parametric model by assessing the performances of all implemented models. In order to realize a regional analysis, the Canadian dataset is divided into five regions and the performance assessment is repeated for every parametric model.

Finally, chapter 5 will introduce the theory of radial basis functions which presents a completely different approach for interpolating the height triplet. They contain a shape parameter in the interpolation kernel which can be defined and computed in various ways. The relationship between the choice of a shape parameter and the interpolation error of the RBF kernel will be investigated. Four alternative possibilities of evaluating the accuracy of the RBF including the cross-validation procedure are studied as well. As the interpolation error can be minimized by using an optimized shape parameter, the powerful optimization technique called genetic algorithm (GA) will be introduced and its application investigated. Afterwards, the performance of the RBF without and with GA optimized shape parameter will be analysed for the Canadian dataset and its regions. It will be also compared with the accuracy obtained by applying the conventional parametric surfaces implemented in *OrthoPlus* (chapter 4).

2 Background: Heights

Heights above the Earth's surface can be defined in several ways. In order to define a local height reference system, two different kinds of specifications are necessary (Marti and Schlatter, 2002): the *height of a reference point* and the *type of the height system*. The reference point is usually a tide gauge station with a defined height of zero. The heights of all other points are computed by adding the measured height differences to this reference height. The height system describes the equations which are used to compute the height differences either from the original spirit-levelling measurements or the geopotential numbers which will be explained later in the study.

Above all, it is important that the user understands the basics, equations and differences between the height systems and in which cases and ways each of them needs to be applied. Therefore, this section will provide some background information in order to give a better understanding of the next chapters of this thesis. It will include a brief overview, some explanations and a short comparison of the modern height definitions. Several methods used to determine the various heights will be also mentioned. For further information about this topic and full formulations and derivations see also Jekeli (2000), Fotopoulos (2003) and Hofmann-Wellenhof and Moritz (2005).

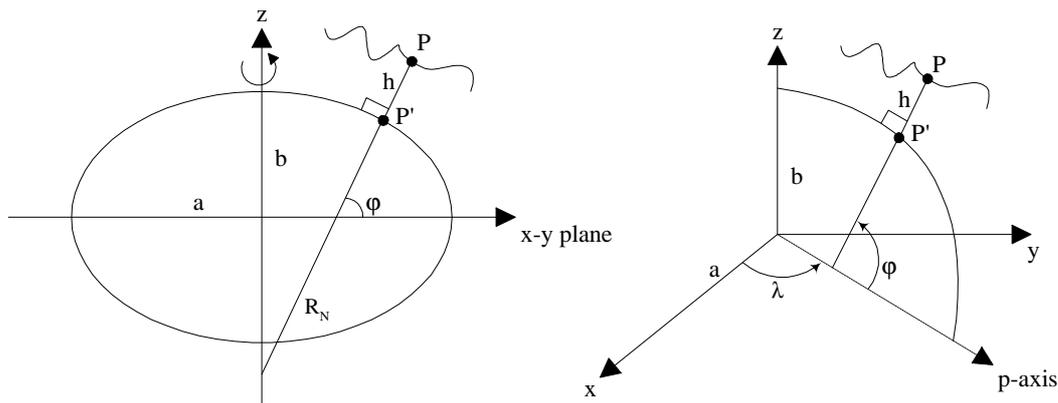
2.1 Ellipsoidal heights

The most common way to specify mathematical point coordinates on the surface of the Earth is in geodetic latitude φ , geodetic longitude λ and ellipsoidal height h . These coordinates refer to a reference rotational ellipsoid which is a mathematically defined surface and chosen to approximate the geoid, the true figure of the Earth. The ellipsoid's centre is assumed to be at the Earth's centre of mass and its semi-minor axis coincides with the Earth's reference pole. The type of the ellipsoidal reference surface and its relation to the geoid surface and the local vertical datum varies from region to region. For GPS users, the typical reference ellipsoid is the GRS 1980 (Geodetic Reference System 1980) which is defined by the four parameters given in the following Table 2.1 (Moritz, 1980).

Table 2.1: Defining parameters of the GRS 1980 (Hofmann-Wellenhof and Moritz, 2005)

parameter	value	description
a	6378137 m	semi-major axis of the ellipsoid
GM	$3986005 \cdot 10^8 \text{ m}^3 \text{ s}^{-2}$	geocentric gravitational constant of the Earth times mass (including the atmosphere)
J_2	$108263 \cdot 10^{-8}$	dynamical form factor of the Earth (excluding the permanent tidal deformation)
ω	$7292115 \cdot 10^{-11} \text{ rad s}^{-1}$	angular velocity of the Earth

The ellipsoidal height is defined as the distance along the perpendicular from a point P on the surface of the Earth to its projection P' on the reference ellipsoid.

**Figure 2.1:** Reference ellipsoid and geodetic coordinates (Fotopoulos, 2003)

Another way to specify the position of P is by Cartesian coordinates (x, y, z) which are mainly provided by GPS measurements since the 3-D coordinates of satellite-signal receiver and satellite can be determined within the same reference frame.

To transform the geodetic coordinates $(\varphi, \lambda, h) \rightarrow (x, y, z)$, the following closed form equations have to be applied (Torge, 1991):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (R_N + h) \cos \varphi \cos \lambda \\ (R_N + h) \cos \varphi \sin \lambda \\ [(1 - e^2)R_N + h] \sin \varphi \end{bmatrix}, \quad (2.1)$$

with the curvature in the prime vertical:

$$R_N = \frac{a}{\sqrt{1 - e^2 \sin^2 \varphi}}, \quad (2.2)$$

where e^2 is the first eccentricity squared and a is the semi major axis of the ellipsoid.

The reverse transformation from Cartesian to geodetic coordinates $(x, y, z) \rightarrow (\varphi, \lambda, h)$, mainly used for GPS processing software packages, is more difficult since the equations are not closed form but iterative:

$$\begin{aligned} \tan \lambda &= \frac{y_P}{x_P}, \\ h &= \frac{p}{\cos \varphi} - R_N, \\ \varphi &= \arctan \left(\frac{z_P}{p} \left(1 - e^2 \frac{R_N}{R_N + h} \right)^{-1} \right), \end{aligned} \quad (2.3)$$

where x_P , y_P and z_P are the coordinates of the point of interest P and $p = \sqrt{x_P^2 + y_P^2}$. For this iterative transformation, initial values for h and φ have to be estimated and a convergence criterion for h or R_N has to be defined which is dependent on the application (Seeber, 1993). Alternatively, approximation formulas can be used with a sufficient accuracy in the derived geodetic coordinates.

Apart from the most common satellite positioning system GPS, there are nowadays also a few other technologies from which ellipsoidal heights can be obtained. These include very long baseline interferometry (VLBI), satellite laser ranging (SLR), satellite altimetry (mainly over oceans), and other navigation based systems such as DORIS, GLONASS and GALILEO in the future (see IAG Comm.1). It should also be noted that the ellipsoidal height is a geometrical height since it refers to a reference ellipsoid. All other height systems which will be introduced in the next sections are physical heights as they take the gravity field of the Earth into account.

2.2 Geopotential numbers

Geopotential numbers are defined as the difference between the potential at the geoid with $W = W_0$ and the potential at points on the surface of the Earth (Hofmann-Wellenhof and Moritz, 2005). They can be computed by the integration of the potential differences along a levelling line, starting from a reference point O on the geoid to a point P on the surface:

$$C_P = W_0 - W_P = \int_O^P g dn \approx \sum_O^P g \delta n, \quad (2.4)$$

where C_P is the geopotential number, g is the gravity, δn is the vertical levelling increment between equipotential surfaces along a path which is denoted as dn in an integral and $W_0 - W_P$ is the potential difference between points O and P.

In applications, a sum is used instead of the integral. The physical unit is [m^2/s^2] or [$\text{kGal}\cdot\text{m}$].

The levelling line can be chosen randomly since the geopotential number C_P only depends on the potential of point P which is the same for all points on the corresponding equipotential surface, i.e., the gravity field is conservative. However, if we only consider the height differences Δn_{OP} from the spirit-levelling measurements without taking any gravity measurements into account, it can be recognized that the height differences depend on the path of integration:

$$\Delta n_{OP} = \int_O^P dn \approx \sum_O^P \delta n. \quad (2.5)$$

The advantage of the geopotential number is that it can be determined by a combination of spirit-levelling and gravity measurements without making any hypotheses. Assumptions about the density distributions within the Earth are not necessary. However, since a geometrical interpretation is not possible and the unit is not in meters, the geopotential number is only used for scientific purposes.

In order to convert a geopotential number C into meters, it can be divided by a location dependent gravity value g [m/s^2]. According to the choice of g , different physical height systems can be defined: for instance the dynamic height H^{dyn} , the orthometric height H or the normal height H^N . All of them will be described in the next sections. The main focus, however, will be placed on describing the orthometric heights as they will be used for the numerical computations in this thesis.

2.3 Dynamic heights

The dynamic heights will compensate for the unity problem by dividing the geopotential number C by the constant normal gravity γ_0 . This yields the following height definition in the unit of meters:

$$H^{\text{dyn}} = \frac{C}{\gamma_0}, \quad (2.6)$$

with $\gamma_0 = \gamma_{45^\circ} = 9.806\,199\,203 \text{ ms}^{-2} = 980.619\,920\,3 \text{ Gal}$ as a function of an arbitrary standard latitude. Here, the given gravity value refers to a latitude of 45° on the rotational ellipsoid GRS 1980.

It should be recognized that this height system describes the flowing direction of the water (Marti and Schlatter, 2002). Water will not flow between points of the same height.

In order to convert the measured height differences Δn_{AB} between two surface points A and B directly into a difference of dynamic heights $\Delta H_{AB}^{\text{dyn}}$, a so-called dynamic correction DC_{AB} has to be added:

$$\Delta H_{AB}^{\text{dyn}} = \Delta n_{AB} + DC_{AB} = \int_A^B \frac{g - \gamma_0}{\gamma_0} dn. \quad (2.7)$$

Like the geopotential numbers, the dynamic heights are practically not in use since this height system has been derived from the geopotential numbers by a transformation of units but has no geometrical meaning and at the same time very large dynamic corrections.

2.4 Orthometric heights

Orthometric heights are defined as the heights above the geoid. They represent the length of the curved plumb-line segment through a point P on the Earth's surface and through a point P_0 on the geoid. These heights can be interpreted as the heights above the mean sea level if the geoid is defined through the mean sea level.

They can be computed by dividing the geopotential number by a mean gravity:

$$H = \frac{C_P}{g}, \quad (2.8)$$

where $C_P = W_0 - W_P$ is the geopotential number of P and $\bar{g} = \frac{1}{H} \int_{P_0}^P g dH$ the mean gravity along the plumb-line between P and P_0 .

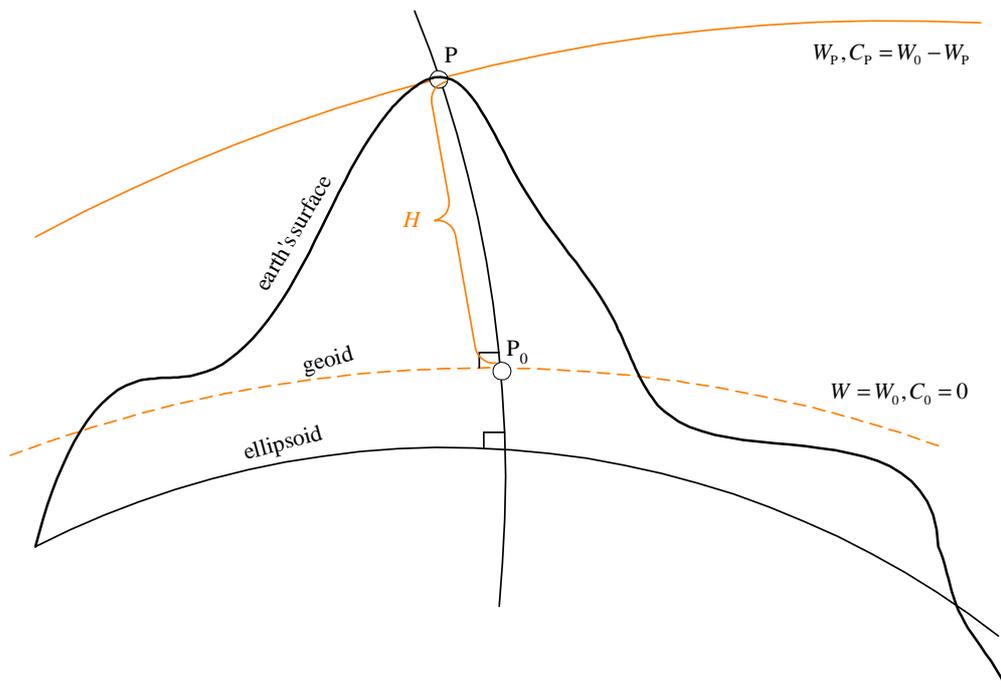


Figure 2.2: Definition of the orthometric height

Yet, the orthometric heights cannot be determined exactly since the mean gravity cannot be measured within the Earth. The determination of the mean gravity is consequently based on a hypothesis made about the mass density within the Earth's crust.

There are some basic approaches for the approximations of the gravity within the crust. One of the easier ones is the Prey-reduction which provides a model based on the gravity g measured at the surface point P. With this approach, the orthometric height H results in the Helmert height (Hofmann-Wellenhof and Moritz, 2005; Jekeli, 2000):

$$H = \frac{C}{g + 0.0424 \frac{\text{mGal}}{\text{m}} H}, \quad (2.9)$$

with C in $\text{kGal} \cdot \text{m}$, g in Gal and H in km .

Analogue to the dynamic heights, the measured height difference Δn_{AB} between two surface points A and B can be converted to a difference in orthometric height ΔH_{AB} by adding an orthometric correction OC_{AB} :

$$\Delta H_{AB} = \Delta n_{AB} + OC_{AB} = \sum_A^B \frac{g - \gamma_0}{\gamma_0} \delta n + \frac{\bar{g}_A - \gamma_0}{\gamma_0} H_A - \frac{\bar{g}_B - \gamma_0}{\gamma_0} H_B, \quad (2.10)$$

where \bar{g}_A and \bar{g}_B are the mean values of gravity along the plumb lines of A and B, γ_0 is the normal gravity at a latitude of 45° , H_A and H_B are the orthometric heights of A and B and δn is the levelling increment.

Another disadvantage of this height system is that water might flow between points of the same orthometric height. The reason is that the mean values of gravity \bar{g}_A and \bar{g}_B are usually different at two points A and B with the same geopotential numbers $C_A = C_B$. For small areas, these differences are negligible (Schneid and Meichle, 2006). On the other hand and as shown by Hofmann-Wellenhof and Moritz (2005), the accuracy of determining orthometric heights is very high.

2.5 Normal heights

Another physical height system with the unity of length is the normal height. This height type has been introduced by M.S. Molodenskij. As opposed to the orthometric heights, these heights can be determined without making any hypotheses. The reason is that instead of the mean actual gravity \bar{g} , the mean normal gravity $\bar{\gamma}$ is used which can be exactly computed for each chosen normal gravity field. At the moment, the used model is the Global Reference System 1980 (GRS80) which has been adopted as the official reference system by the IUGG (International Union of Geodesy and Geophysics) and which forms the fundamental basis of the WGS84 (World Geodetic System 1984). The defining parameters of the GRS80 are listed in Table 2.1. With these parameters, normal heights denoted by H_p^N can be obtained as follows:

$$H_p^N = \frac{C_p}{\bar{\gamma}}, \quad (2.11)$$

$$\text{where } C_p = U_0 - U_p = W_0 - W_p. \quad (2.12)$$

The latter formula means that H_p^N specifies a point P' which lies on the plumb-line through the surface point P . For this point, the normal potential difference with respect to the reference ellipsoid $U = U_0$ equals the geopotential number C_p of the point P which is the actual potential difference. These kinds of points P' for which Eq. (2.12) is fulfilled, lie on a surface called the telluroid. Its form is similar to the physical surface of the Earth. The ellipsoidal height of a point P' on a telluroid above the reference ellipsoid is the same as the normal height H_p^N of the corresponding surface point P . The distance between the telluroid and the surface of the Earth is called the height anomaly ζ .

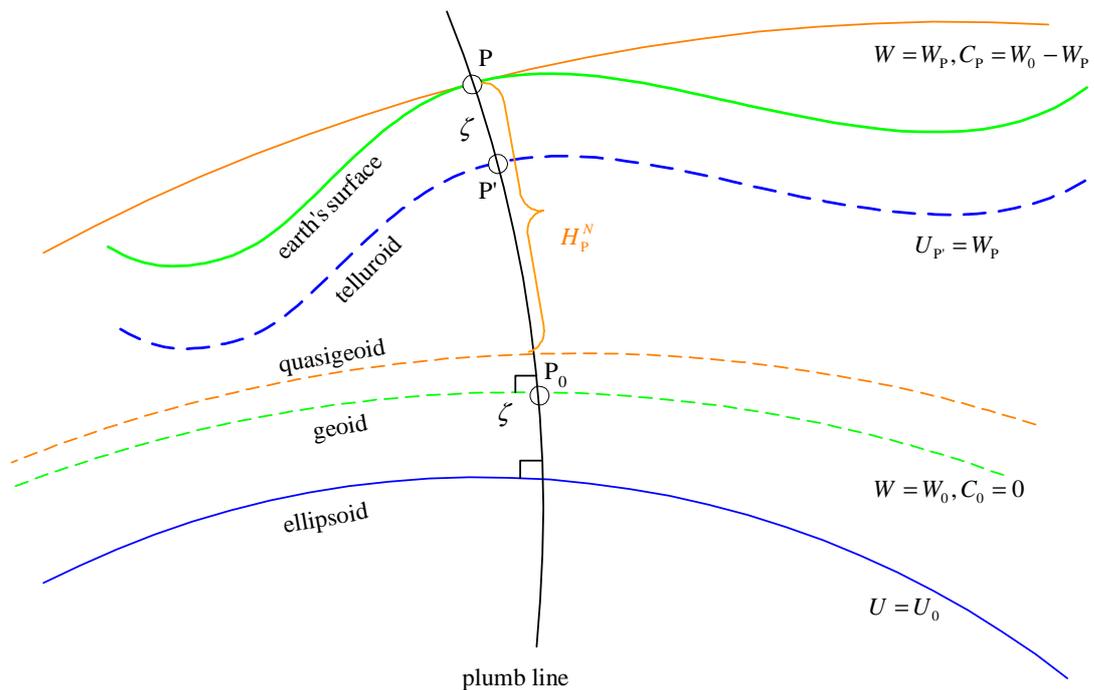


Figure 2.3: Definition of the normal height

On the other hand, if we are going from any surface point along the plumb-line with the length of the normal height, we obtain a surface called the quasigeoid which is close to the geoid and coincides with the geoid at sea. The disadvantage herein is that the reference surface has no physical meaning and that this time, the height anomalies ζ provide the distance between the quasigeoid and the reference ellipsoid. The normal heights include the same problem as the orthometric heights, i.e., the condition for a still water surface between points of the same height can be only complied for small areas.

An approximation of the mean normal gravity $\bar{\gamma}$ along the plumb-line from the surface point to the quasigeoid is computed via (Hofmann-Wellenhof and Moritz, 2005):

$$\bar{\gamma} \approx \gamma_{\text{ell}} \left[1 - (1 + f + m - 2f \sin^2 \varphi) \frac{H^N}{a} + \frac{(H^N)^2}{a^2} \right], \quad (2.13)$$

$$\text{where } m = \frac{\omega^2 a^2 b}{GM}. \quad (2.14)$$

Since the dependency of $\bar{\gamma}$ on the normal height H^N is not very strong, an iterative solution is not a problem. The ellipsoidal normal gravity γ_{ell} of the point only depends on the latitude φ and can be obtained by the normal gravity field of Somigliana (Hofmann-Wellenhof and Moritz, 2005), for example:

$$\gamma_{\text{ell}} = \gamma_e \frac{1 + k \sin^2 \varphi}{\sqrt{1 + e^2 \sin^2 \varphi}}, \quad (2.15)$$

$$\text{where } k = \frac{b\gamma_p - a\gamma_e}{a\gamma_p}. \quad (2.16)$$

a , b , ω , GM , f , k , m and e^2 are parameters of the reference ellipsoid GRS80, φ is the geodetic latitude of the point and γ_e and γ_p are the normal gravities on the equator and on the pole.

Here again, as for the dynamic and orthometric heights, a normal correction NC_{AB} can be added to a measured height difference Δn_{AB} between two points A and B in order to get a normal height difference ΔH_{AB}^N :

$$\Delta H_{AB}^N = \Delta n_{AB} + NC_{AB} = \sum_A^B \frac{g - \gamma_0}{\gamma_0} \delta n + \frac{\bar{\gamma}_A - \gamma_0}{\gamma_0} H_A^N - \frac{\bar{\gamma}_B - \gamma_0}{\gamma_0} H_B^N. \quad (2.17)$$

2.6 Comparison of different height systems

With different types of height systems available, the question is which system would suit best the purposes of a national height system. In most cases, the ellipsoidal heights are not suitable since their definition is merely based on a geometric basis, i.e., they are defined as the perpendicular to the surface of the reference ellipsoid. The fact that the ellipsoid is not parallel to the different equipotential surfaces

results in a misclosure of up to 10 cm/km for spirit-levelling measurements (Schneid and Meichle, 2006), which is not desired in height measurement.

The dynamic heights cannot be used in a practical height system. The reason is that the geopotential numbers are divided by a constant gravity value. That means that the polar convergence of an equipotential surface is not taken into account.

Therefore, the choice of a national reference system is restricted to one of the remaining physical height systems, the orthometric or the normal heights. However, it's not easy to say which of them might be more suitable. Both of them are based on the geopotential numbers. Thus, the heights can be determined independently from the levelling line. Their only disadvantage is that water flows between two points of the same orthometric or normal height since these points will lie on surfaces with different potentials.

The important difference between the two height systems is the gravity value which is used to convert the geopotential number into a height in units of length. While the mean actual gravity \bar{g} is based on models of density along the plumb line, the mean normal gravity $\bar{\gamma}$ can be computed explicitly via the defining parameters of the reference ellipsoid. That is why the effort to compute orthometric heights is much bigger than for the normal heights unless Helmert's simple formula in Eq. (2.9) is used.

Table 2.2 summarizes the characteristics of the orthometric and normal heights.

Table 2.2: Summary of the main characteristics of the orthometric and normal heights (Marti and Schlatter, 2002)

characteristics	orthometric height H	normal height H^N
rigorous system with dependency on the potential	yes	yes
system without a hypothesis (about mass distribution)	no	yes
easy computation	no	yes
physical meaning of the reference surface	yes	no
geometrical relation to ellipsoidal heights	yes	yes

The ellipsoidal height can be either computed as the sum of the normal height H^N and the height anomaly ζ or the sum of the orthometric height H and the geoid undulation N , which is the difference between the geoid and the reference ellipsoid.

The crucial argument for using orthometric rather than normal heights in Canada is the issue of using GPS for height determination (Marti and Schlatter, 2002). The conversion of ellipsoidal heights in orthometric heights is easier and more precise, especially in areas with steep and fast changing slopes as e.g. in the Rocky Mountains. The reason is that the geoid model is relatively smooth compared to the quasigeoid.

2.7 Height determination

2.7.1 Spirit-levelling

As already mentioned above, there are several possibilities to determine the elevations of points on or near the Earth's surface. The most conventional technique is called spirit-levelling. An instrument called a spirit level points exactly in the direction of the local horizontal and is positioned in the middle between two points. It measures certain graduations from each of the vertical rods set up at these points. The difference in the rod readings will provide the height difference. By including a vertical control point and combining the levelling results along the levelling lines with gravity measurements, the orthometric heights can be derived directly. They will be given with respect to a vertical geodetic datum. The Canadian Geodetic Vertical Datum of 1928 (CGVD 28) is determined by spirit-levelling but does not represent today's required national accuracy (Véronneau et al., 2006). Another problem is that this kind of levelling is too costly, time consuming and laborious for any kind of maintenance or expansion of the vertical network.

2.7.2 GNSS-levelling

An alternative for the realization of a long-term regional vertical datum for Canada is applying GNSS-levelling (global navigation satellite systems). This relatively modern procedure of determining orthometric heights with respect to a local vertical datum is the main topic of this report and will be referred to as GPS-levelling as only ellipsoidal heights derived from GPS measurements are used in this work. The theoretical basic equation is given by (Hofmann-Wellenhof and Moritz, 2005):

$$h - H - N = 0. \quad (2.18)$$

This equation shows the fundamental linear relationship between the ellipsoidal height h (above the ellipsoid), the orthometric height H (above the geoid) as a combination of spirit-levelling and gravity measurements, and the geoid height (or geoid undulation) N which is the distance between the geoid

and the reference ellipsoid obtained from a regional gravimetric model or a global geopotential model (Figure 1.1). Measuring two of these quantities makes it possible to compute the third one (Hofmann-Wellenhof and Moritz, 2005).

Theoretically, the orthometric height can be obtained immediately if a reliable geoid model as well as an ellipsoidal height are available:

$$H = h - N \tag{2.19}$$

This implies that the combination of classical spirit-levelling together with gravity measurements, i.e., conventional levelling methods, are no longer needed and can be replaced by GPS. However, there are some issues inherent in the relationship of the triplet of height data which will be explained and must be taken into account in the sequel (e.g. datum inconsistencies, systematic errors, data accuracy). For increasing accuracy requirements, it is important to compensate for this problem by the incorporation of a parametric “corrector” surface model which can absorb the datum inconsistencies and any systematic distortions.

3 Theory and Implementation

One of the main objectives of this report was to implement and add some modules to a semi-automated program in Matlab called *OrthoPlus* (© Fotopoulos and Sideris). It facilitates every user to find an optimal combination of orthometric, geoid and ellipsoidal heights by the incorporation of a pre-selected parametric model. This best combination which should fit as good as possible to the user's expectations and to the corresponding dataset, can then be used for GPS-levelling.

The program is mainly based on a combined height adjustment with a selected parametric surface model. A performance assessment of this chosen model is also included as well as the computation of a stochastic model by means of variance component estimation. Thus, the implementation of a subsequent regional analysis will be relatively straight forward. However, it must be mentioned that *OrthoPlus* is a semi-automated program. The term semi-automated means that using the program also requires some user intervention. The modules implemented in the program *OrthoPlus* are summarized in the flowchart in Figure 3.1.

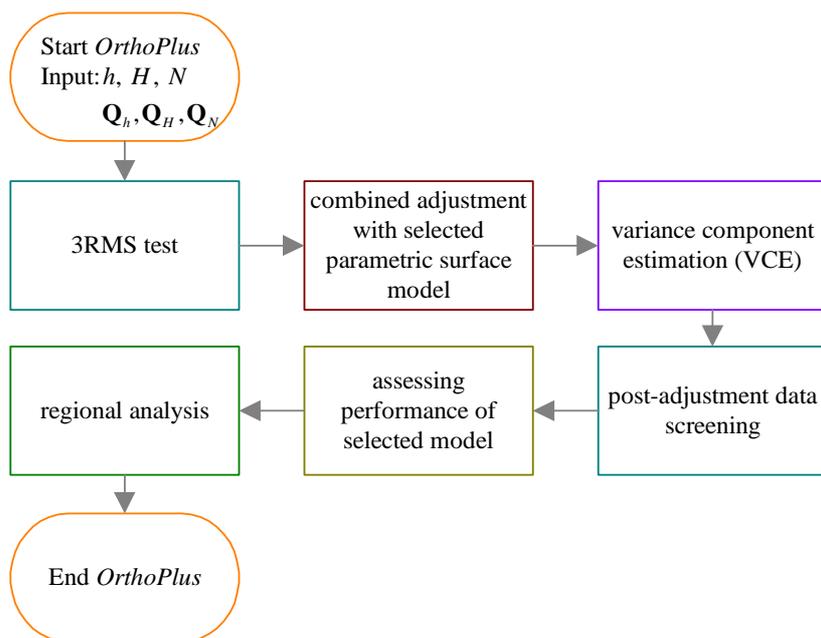


Figure 3.1: Summary of modules in *OrthoPlus*

The next sections will explain these different modules step-by-step. This will include the description of the theoretical background as well as the practical realization in the program. By means of flowcharts, the user should get an overview and an idea of the possible interactions with *OrthoPlus*.

3.1 General information to *OrthoPlus*

OrthoPlus is written in Matlab, Version 7.0.0.19920 (R14). The file name of the main program is `OrthoPlus.m` which will call the various functions listed in Table 3.1 depending on the choice of the user.

Table 3.1: Overview and descriptions of functions called in `OrthoPlus.m`

name of function	description
<code>fplot_bms.m</code>	plots the GPS-levelling benchmark network (φ, λ)
<code>fgeod2cart.m</code>	converts geodetic to Cartesian coordinates for the GRS80
<code>fgramschmidt.m</code>	applies Gram-Schmidt orthonormalization
<code>fglobaltest_chi2.m</code>	detects blunders via a global test (two-tail χ^2)
<code>fglobaltest_chi1.m</code>	detects blunders via a global test (one-tail upper bound χ^2)
<code>fglobaltest_F.m</code>	detects blunders via a global test (F-distribution)
<code>fdatasnooping.m</code>	detects and identifies blunders via data snooping for individual residuals
<code>ftautest.m</code>	detects and identifies blunders via the τ -test (unknown covariance matrix)
<code>fplot_surface.m</code>	plots the corrector surface model using the estimated parameters x
<code>fcrossvalid.m</code>	applies cross validation algorithm

The two files `fplot_bms.m` and `fplot_surface.m` use functions from the additional Matlab mapping toolbox called “*M_Map*” which is available as a free download in the internet (<http://www.eos.ubc.ca/~rich/map.html>).

The input data are read into the program as *.txt files with the following structure:

BMS-ID	lon[°]	lat[°]	h[m]	H[m]	N[m]	h-H-N[m]
1	-119.6250	49.3226	541.8619	559.0322	-16.646	-0.5243
2	-122.1678	52.2369	1095.6595	1110.9024	-14.773	-0.4699

whereas the first line has just been added for demonstration purposes and must be deleted in practice. The first column specifies the benchmark identification numbers, the next columns define the

corresponding longitude in degrees ($lon[^\circ]$), the latitude in degrees ($lat[^\circ]$), the ellipsoidal height ($h[m]$), the orthometric height ($H[m]$), the geoid height ($N[m]$) and the height misclosure ($h-H-N[m]$). If the *.dat files of the covariance matrices of the height triplets are given, they can be read in as well. Otherwise, the user can choose to let *OrthoPlus* define these covariance matrices as identity matrices. After setting the correct path, the user can run and open the file `OrthoPlus.m` in the Matlab command window by entering the command `OrthoPlus`. All further steps will be self-explanatory.

3.2 Data screening, gross error detection and elimination

In order to get a reliable assessment of the results from the program, it is important to make sure that the dataset itself does not contain any gross errors, also known as blunders. The following sections will introduce an implementation of a general blunder detection within *OrthoPlus*. Although the Canadian dataset is free of blunders as it was already screened beforehand, other networks could contain height data blunders. Blunders in data affect the subsequent processes like the variance component estimation, for example, as they propagate into the final estimates and provide unrealistic results. Least squares techniques are very sensitive to gross errors. Hence, it is important to test for blunders first and remove them before the estimation if possible.

The tested dataset in our case is represented by the resulting height misclosure vector implemented by the basic equation for every measured height triplet of n benchmark points (see also Eq. 2.18):

$$l_i = \begin{bmatrix} \vdots \\ h_i - H_i - N_i \\ \vdots \end{bmatrix}. \quad (3.1)$$

While random errors can be minimized through any adjustment procedures, blunders across the entire dataset will be smoothed out by a least squares minimization. Especially large blunders contaminate the neighbouring observations. The reason is that the least squares method assumes a normal distribution and blunders are considered to be totally unwanted, as for example mistakes in the data or in the measurements made by an observer or even due to mistakes in the functional model (Gao and Sideris, 2005). In case of spirit-levelling, for instance, a gross error could be caused by the observer accidentally switching the figures when reading off the levelling rod. Mostly, these types of errors affect individual observations as opposed to the unavoidable measurement noise which has an influence on all observations. Thus, to avoid blunders to be inherent in the dataset, there are two possibilities.

First of all, observational procedures have to be done properly, repeated and even with slightly different techniques for instance. The second step is to come up with algorithms to detect blunders, to identify their source and to finally eliminate them from the mathematical models before the determination of a solution. The latter step is called data screening and can be separated in two different parts which are implemented in *OrthoPlus* and will be explained in the following two sections: the pre-adjustment data screening and the post-adjustment data screening (Kuang, 1996). It must be mentioned that the correct detection and removal of blunders of moderate magnitude cannot be guaranteed even with modern statistical techniques (Mikhail, 1976). This semi-automated program is implemented to find blunders in the dataset, but actually puts the responsibility for the removal on the user. Experiences and intuition of the investigator are expected. He should find a trade-off between accepting and refusing remaining blunders for the sake of redundancy.

3.2.1 Pre-adjustment data screening

The purpose of the pre-adjustment data screening is to detect and remove any possible gross errors of large magnitude in the given dataset prior to any network adjustment as they will be smoothed out. This kind of blunder detection has been realized by a routine 3-RMS test (Kuang, 1996), which is very commonly used in literature and designed to get rid of obvious outliers. The root mean square value (RMS) of all observations can be computed as follows:

$$RMS_l = \sqrt{\mu_l^2 + \sigma_l^2} = \sqrt{\frac{1}{N} \sum_i l_i^2}, \quad (3.2)$$

where μ_l is the mean value and σ_l is the standard deviation of the observation vector \mathbf{l} .

For every height misclosure l_i , the following condition will be checked:

$$|l_i| > 3 \cdot RMS_l. \quad (3.3)$$

If the statement is true, the observation l_i will be flagged as a blunder. The user has the choice to either keep all observations and to continue with the program or to eliminate these blunders manually from the original dataset together with the corresponding rows and columns of the cofactor matrices and to restart the program afterwards (Figure 3.2).

Analogously, no blunders can be validated if the contrary of the statement in Eq. (3.3) is true. That is the case when the absolute value of an observation l_i lies within a limit of three times the RMS value of all observations with a probability of 97.3% (Welsch et al., 2000).

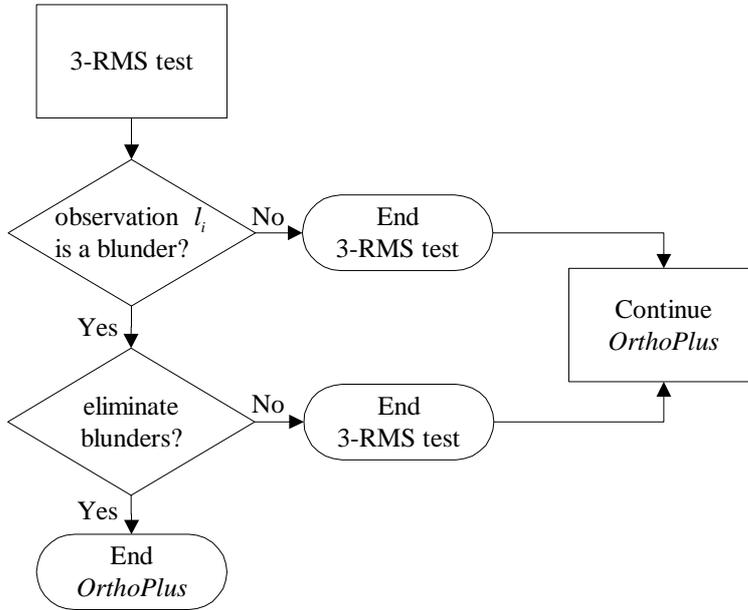


Figure 3.2: Flowchart for 3-RMS test

3.2.2 Post-adjustment data screening

In order to detect and eliminate gross errors of small (or marginal) magnitude, post-adjustment data screening has been applied after the least-squares network adjustment based on statistical testing on the residuals. Residuals are estimated corrections which have to be applied to the survey data. The residual vector \mathbf{v} with a length of n , which is the number of observations, can be computed via:

$$\mathbf{v}_{n \times 1} = -(\mathbf{C}_l \mathbf{R}) \mathbf{l}, \quad (3.4)$$

where \mathbf{C}_l is the $n \times n$ covariance matrix for all observations, \mathbf{l} the $n \times 1$ vector of observations and \mathbf{R} a full matrix of size $n \times n$ which provides geometrical information of the network design (Fotopoulos, 2003):

$$\mathbf{R} = \mathbf{P}(\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}). \quad (3.5)$$

In this equation, \mathbf{P} is the block diagonal weight matrix of size $n \times n$ and \mathbf{I} is a $n \times n$ unit matrix. The $n \times u$ design matrix \mathbf{A} depends on the parametric model where u is the number of unknown parameters. The derivation of these matrices will be explained in more detail in the sequel of this study (section 3.4.1).

Any gross error inherent in the observation vector will cause an outlier. An outlier is a residual which exceeds a boundary value of a statistical test rule. The influence of a blunder in the observations, denoted by $\nabla \mathbf{I}$, on the residuals can be described in three different ways (Gao and Sideris, 2005):

1. impact of the blunder in a specific observation on corresponding residual:

$$\nabla v_i = -(\mathbf{C}_l \mathbf{R})_{ii} \nabla l_i \quad (3.6)$$

2. impact of blunder in a specific observation on more than one residual

$$\nabla v_j = -(\mathbf{C}_l \mathbf{R})_{ji} \nabla l_i \quad j = 1, 2, \dots, n \quad (3.7)$$

3. impact of any blunder in all observations on a specific residual

$$\nabla v_i = -[(\mathbf{C}_l \mathbf{R}) \nabla \mathbf{I}]_i = -\sum_{j=1}^n [(\mathbf{C}_l \mathbf{R})_{ij}] \nabla l_j. \quad (3.8)$$

This shows that even if an outlier can be detected, it is difficult to find the corresponding erroneous observation. The other difficulty is that only small parts of the gross errors show up in the residuals, the rest will be absorbed during the determination of the unknown parameters in the adjustment procedure. Therefore, we must distinguish between the detection of outliers and the actual localization of gross errors or blunders. These two steps are based on a statistical testing of the following general null hypothesis (Kuang, 1996):

$$H_0 : \mathbf{I} \in N(\mathbf{A}\mathbf{x}, \sigma_0^2 \mathbf{P}^{-1}), \quad (3.9)$$

which implies the basic assumption that the observations \mathbf{I} are normally distributed with the expected value of $\mathbf{A}\mathbf{x}$ and the covariance matrix $\mathbf{C}_l = \sigma_0^2 \mathbf{P}^{-1}$ according to the model in the combined height adjustment described in section 3.4.1.

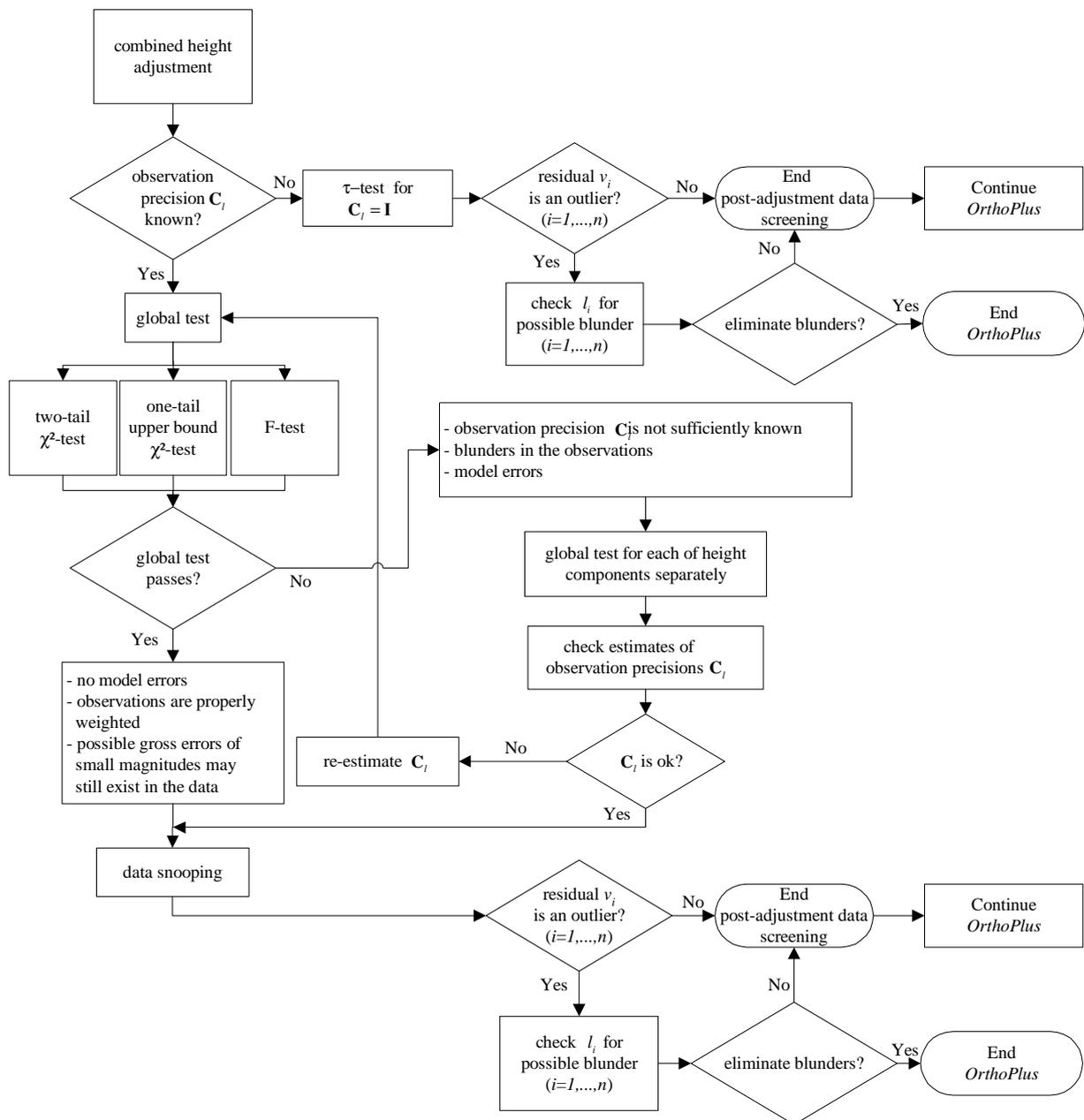


Figure 3.3: Flowchart for post-adjustment data screening (Kuang, 1996)

In *OrthoPlus*, there are mainly two different approaches to test the null hypothesis H_0 . In both approaches, the residuals estimated in the combined height adjustment are used. If the variance of the observations is known, the statistical testing on blunders is implemented in two steps: first the detection of outliers by a global test and then the localization of gross errors by data snooping. If the observational precision is unknown and thus, the user had chosen identity matrices as covariance matrices, then the global test and the data snooping procedures will be replaced by the τ -test.

The following three sections will give an overview of these main modules of post-adjustment data screening in *OrthoPlus*. Together with the flowchart in Figure 3.3, the user should get an idea of how possible blunders in the dataset can be detected, located and then potentially eliminated with the help of *OrthoPlus*.

3.2.2.1 Global test on the a-posteriori variance factor

In the case of a known covariance matrix \mathbf{C}_l , the global test applied after the network adjustment decides whether there are any blunders in the measurements or not. It tests the compatibility of the estimated a-posteriori variance factor $\hat{\sigma}_0^2$ with the a-priori selected variance factor $\sigma_0^2 = 1$.

The a-posteriori variance factor is computed by (Niemeier, 2002):

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T \mathbf{P} \mathbf{v}}{r}, \quad (3.10)$$

where \mathbf{v} is the $n \times 1$ vector of estimated residuals, \mathbf{P} is the block diagonal weight matrix of size $n \times n$ and $r = n - u$ is the redundancy, which is the difference in the number of observations and unknowns. \mathbf{P} is determined by small modifications of the following formula:

$$\mathbf{C}_l = \sigma_0^2 \mathbf{Q} = \mathbf{Q} = \mathbf{P}^{-1}. \quad (3.11)$$

The user can choose between three different kinds of global tests:

- Two-tail χ^2 -global test
- One-tail upper bound χ^2 -global test
- Global F-test.

The following sections will give an overview of how the null and alternative hypotheses are tested, the corresponding test statistic and the test decision for each test.

Two-tail χ^2 -global test. The two-tail χ^2 -global test (`fglobaltest_chi2.m`) assumes that the test statistic y , denoted by (Kuang, 1996):

$$y = \frac{r \cdot \hat{\sigma}_0^2}{\sigma_0^2}, \quad (3.12)$$

follows the chi-square distribution $\chi^2(r)$ with the degrees of freedom r under the null hypothesis H_0 :

$$y|H_0 \in \chi^2(r). \quad (3.13)$$

This assumption implies that the expected value of the estimated a-posteriori variance factor equals the a-priori variance factor. The test decision is based on a two-tailed statistical testing with the significance level α , which is usually 5% but can be manually chosen by the user. If the following condition is complied, the null hypothesis can be accepted which means that no blunders have been detected in the entire dataset:

$$\chi_{\alpha/2}^2(r) \leq y \leq \chi_{1-\alpha/2}^2(r). \quad (3.14)$$

$\chi_{\alpha/2}^2(r)$ and $\chi_{1-\alpha/2}^2(r)$ are computed from the standard statistical tables for a significance level α and degrees of freedom r (see Niemeier, 2002). If, on the other hand, H_0 is rejected, the same test will be repeated for each height component, i.e., for the orthometric, the geoid and the ellipsoidal height separately, in order to find out which of them could be responsible for the rejection of H_0 .

Rejecting the null hypothesis means at the same time that the alternative hypotheses have to be tested which are defined as follows (Kuang, 1996):

H_{a1} : incorrect observational weighting scheme

H_{a2} : gross errors exist in observation data (data snooping of individual residuals).

The test decision for the first alternative hypothesis H_{a1} can be noted as:

$$|v_{\max}| < RMS_l. \quad (3.15)$$

If this condition is verified which means that the magnitudes of the residuals with a maximum value of v_{\max} are small compared to the RMS of the observations, denoted by RMS_l , and if $\hat{\sigma}_0^2$ is not compatible with σ_0^2 (rejection of the null hypothesis), it can be assumed that the a-priori given precision of the observations with $\sigma_0^2 = 1$ is not correct. The consequence is that the covariance matrix of the observations C_l has to be modified by:

$$\tilde{\mathbf{C}}_l = \hat{\sigma}_0^2 \mathbf{C}_l. \quad (3.16)$$

Afterwards, the global test will be repeated with a newly computed test statistic y using $\tilde{\mathbf{C}}_l$.

While H_{a1} will only be tested if the null hypothesis is rejected, H_{a2} will be tested in any case since it is quite possible that a significant error may stay hidden in the total sum of the squares of the residuals when performing the global test. This procedure which is called data snooping, however, might even detect this specific error and will be explained in section 3.2.2.2.

One-tail upper bound χ^2 -global test This kind of χ^2 -global test is very similar to the two-tail χ^2 -test and can be implemented in exactly the same way (`fglobaltest_chi2.m`). It is more frequently applied in practice. The only difference, however, lies in the one-tail test decision. The null hypothesis will be accepted with a significance level of α if (Mikhail, 1976):

$$y \leq \chi_\alpha^2(r). \quad (3.17)$$

In this case, no outliers have been detected.

Global F-test Another possibility of a global test is the F-test (`fglobaltest_F.m`). Its test statistic is defined as (Mikhail, 1976):

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}, \quad (3.18)$$

$$\text{where } \begin{cases} \hat{\sigma}_1^2 = \hat{\sigma}_0^2 \text{ and } \hat{\sigma}_2^2 = \sigma_0^2 \text{ with } r_1 = r \text{ and } r_2 = n \text{ if } \hat{\sigma}_0^2 > \sigma_0^2 \text{ or} \\ \hat{\sigma}_1^2 = \sigma_0^2 \text{ and } \hat{\sigma}_2^2 = \hat{\sigma}_0^2 \text{ with } r_1 = n \text{ and } r_2 = r \text{ if } \hat{\sigma}_0^2 < \sigma_0^2. \end{cases} \quad (3.19)$$

In other words, the larger one of the two variance factors σ_0^2 (a-priori) and $\hat{\sigma}_0^2$ (a-posteriori) is assigned as numerator of the test statistic, the other one as denominator. The parameters r_1 and r_2 will be correspondingly assigned to either the degrees of freedom r or the numbers of observations n . The test statistic F follows the F-distribution with the parameters r_1 and r_2 under the null hypothesis H_0 and with a significance level of α , chosen by the user:

$$F | H_0 \in F_{\alpha, r_1, r_2}. \quad (3.20)$$

H_0 will be accepted if F is smaller than the test threshold F_{α, r_1, r_2} which can be extracted from the standard statistical tables (see Niemeier, 2002):

$$F \leq F_{\alpha, r_1, r_2}. \quad (3.21)$$

Outliers can be verified if the contrary is true, i.e.,

$$F > F_{\alpha, r_1, r_2}. \quad (3.22)$$

In this case, the next steps are the repetition of the F-test for each height component separately and after that the testing of the first alternative hypothesis which are explained in detail for the two-tail χ^2 -global test. The F-test is often used in situations where the best available variances are also estimates with their own statistical fluctuations (Mikhail, 1976).

The testing of the second alternative hypothesis H_{a2} will be applied even if the null hypothesis has been accepted in order to detect outliers of individual residuals (Kuang 1996). This is described in the next section. So far, the only purpose of the data screening was to find out whether there are any outliers or not. As opposed to this blunder detection, the actual blunder identification is much more complicated since not only the numbers but also the locations of blunders are unknown.

3.2.2.2 Data snooping of individual residuals

Data snooping is a combined process of outlier detection, gross error localization and elimination (`fdatasnooping.m`). Since it is not easy to tell whether blunders nor how many of them and in which observations they are inherent, it will be assumed that only one blunder is present at a time. Therefore, at first, every residual will be tested for outliers individually as opposed to the global test which looks for outliers in the whole dataset.

For this purpose, we need to derive the covariance matrix for the adjusted residuals \mathbf{C}_v , which is the sum of the covariance matrices for the residuals of each height type, i.e., the ellipsoidal height h , the orthometric height H and the geoid height N (Fotopoulos, 2003):

$$\mathbf{C}_v = \mathbf{C}_{v_h} + \mathbf{C}_{v_H} + \mathbf{C}_{v_N} . \quad (3.23)$$

The covariance matrices for the residuals of each separate height type are computed as follows:

$$\mathbf{C}_{v_k} = \mathbf{C}_k \mathbf{RQ}(\mathbf{C}_k \mathbf{R})^T \quad k \in [h, H, N] \quad (3.24)$$

where \mathbf{R} has been already derived in Eq. (3.5) and \mathbf{C}_k are the covariance matrices of each separate height type. \mathbf{Q} is the positive definite cofactor matrix for the entire height triplet (see also section 3.4).

The criteria to assess if the i -th observation contains a gross error or not becomes (Gao and Sideris, 2005):

$$w_i | H_0 \in N(0,1) , \quad (3.25)$$

which means that under the null hypothesis H_0 , the test statistic w_i is normally distributed with zero mean and unit variance.

The corresponding test statistic is (Gao and Sideris, 2005):

$$w_i = \frac{\mathbf{e}_i^T \mathbf{Pv}}{\sigma_0 \sqrt{\mathbf{e}_i^T \mathbf{P} \mathbf{C}_v \mathbf{P} \mathbf{e}_i}} , \quad (3.26)$$

where \mathbf{P} is the block diagonal weight matrix with size $n \times n$ and \mathbf{e} the influence vector with $n \times 1$ in which all elements apart from the i -th are zero:

$$\mathbf{e}_i = [0, 0, \dots, 0, 1, 0, \dots, 0]^T . \quad (3.27)$$

Assuming that the observations are uncorrelated, the test statistic of data snooping can be simplified and we get a standardized residual:

$$w_i = \frac{v_i}{\sigma_0 \sqrt{(C_v)_{ii}}} . \quad (3.28)$$

The decision is based on a two-tailed test which accepts the null hypothesis H_0 , if:

$$N_{\alpha/2r}(0,1) \leq w_i \leq N_{1-\alpha/2r}(0,1) \quad (3.29)$$

This condition is examined for only one standardized residual w_i at a time. However, for the implementation of a complete data snooping algorithm, the whole observation vector has to be screened for blunders by a subsequent repetition of the test above by taking the index i to be successively $1, \dots, n$. Thus, in order to preserve the total significance level α for this multiple test, Niemeier (2002) suggests using a significance level of $\alpha_0 = \alpha/n$ for single tests of every individual residual. The corresponding test thresholds $N_{\alpha/2r}(0,1)$ and $N_{1-\alpha/2r}(0,1)$ follow the Gaussian distribution and can be found in statistical handbooks. If the above condition in Eq. (3.29) is not fulfilled, i.e., H_0 is rejected, a residual v_i can be identified as an outlier.

For the localization and later on the elimination of a blunder, it is very important to check if the excessive, and therefore flagged, residual v_i has been actually caused by a gross error ∇l_i in its corresponding observation l_i . The reason why the localization of the detected blunders could be difficult is that the residuals might be strongly correlated due to the geometry of the network and the number and magnitudes of the gross errors inherent in the dataset. In this case, the observation l_i should not be eliminated and can be tested as follows (Kuang, 1996):

$$r_i \leq \max \{ |r_{ji}| (j \neq i, j = 1, \dots, n) \}, \quad (3.30)$$

where r_i and r_{ji} refer to the diagonal and off-diagonal elements of the covariance matrix for residuals \mathbf{C}_v . The smaller the r_i value which is actually the i -th local redundancy number as a function of the actual adjustment model with $0 \leq r_i \leq 1$, the more difficult for gross errors to be identified in the corresponding observation.

For the post-adjustment data screening, *OrthoPlus* will, for example, provide the following output table:

observation	test statistic (w)	difference (w -0)	gross error?
164	-7.041643	7.041643	difficult to determine
343	+4.423378	4.423378	yes
349	+5.086000	5.086000	yes
351	+4.506073	4.506073	yes
361	+6.532196	6.532196	yes
362	+5.086472	5.086472	yes
416	-5.017148	5.017148	yes
419	+4.949534	4.949534	yes

If you like to remove the gross errors, please delete only the observation with the maximum difference (-> observation 361) from your original dataset and run the program again!

The table contains the observations with flagged outliers, their test statistics and the differences from the mean of the acceptance region which is zero for the symmetric normal distribution. The last column tells the user whether it is difficult to determine if the outlier v_i has actually been caused by a gross error in the observation l_i . Thus, the observation 164 in the example above should not be eliminated. Finally, *OrthoPlus* determines the observation with the maximum test statistic out of all these remaining flagged observations with an outlier (Kuang, 1996). The user has the possibility to stop and then re-run the program after eliminating this specific observation from the observation vector \mathbf{I} as well as the corresponding rows and columns in the covariance matrices. It is highly recommended to only delete one observation at a time. This can imply that the user has to run the program several times in order to isolate all of the gross errors by “trial and error” (Kuang, 1996) and by own responsibility, conscience and comprehension of the problem and the specific network.

3.2.2.3 τ -test

The global test and the data snooping algorithm will be replaced by the τ -test (`ftautest.m`) if the observational precision is unknown. That will automatically happen if the user has chosen the option in *OrthoPlus* to set the covariance matrices as well as the weighting matrices as identity matrices after selecting the parametric model and the algorithm for the variance component estimation.

The τ -test is very similar to the data snooping procedure. The only difference in the derivation of the test statistic τ_i , proposed by Pope (1976), is that the a-posteriori estimated variance factor $\hat{\sigma}_0^2$ is used instead of the a-priori variance factor σ_0^2 :

$$\tau_i = \frac{v_i}{\hat{\sigma}_0 \sqrt{(C_v)_{ii}}}, \quad (3.31)$$

where v_i are the residuals estimated in the combined height adjustment and $(C_v)_{ii}$ are the corresponding diagonal elements of the covariance matrix C_v for the adjusted residuals which has been derived in the data snooping section in Eq. (3.23) and (3.24).

This test statistic follows the τ -distribution under the null hypothesis H_0 with degrees of freedom r (Kuang, 1996):

$$\tau_i | H_0 \in \tau(r). \quad (3.32)$$

The corresponding test threshold is based on the maximum and also called in-context τ -test (Kuang, 1996), i.e., the significance level is reduced to $\alpha_0 = \alpha/n$ where α is the significance level chosen by the user and n is the number of GPS-levelling benchmark points. According to Niemeier (2002), the significance level of this multiple test is, however, defined as $\alpha_0 = \alpha/r$ where r are the degrees of freedom. In contrast to the other distribution types used so far, the values of the τ -distribution cannot be directly derived from a standard statistical table but need to be computed by using the relation to the student's distribution $t_{\alpha_0}(r-1)$ with a significance level of α_0 and degrees of freedom $r-1$ (see Kuang, 1996):

$$\tau_{\alpha_0/2}(r) = \tau_{\alpha/(2r)}(r) = \frac{\sqrt{r} \cdot t_{\alpha_0/2}(r-1)}{\sqrt{r-1 + t_{\alpha_0/2}^2(r-1)}}. \quad (3.33)$$

After computing the test statistic τ_i for every observation and the test threshold $\tau_{\alpha/(2r)}(r)$, the following condition will be checked successively for all residuals v_i :

$$|\tau_i| > \tau_{\alpha/(2r)}. \quad (3.34)$$

If the condition is fulfilled, the corresponding observation l_i will be flagged as an outlier and listed in an output table, shown in section 3.2.2.2. It should be noted that especially blunders with small magnitudes are likely to stay undetected by applying the τ -test. The reason is that the test statistic is in

inverse proportion to the a-posteriori estimated variance factor $\hat{\sigma}_0^2$, i.e., with increasing magnitudes of the blunders and thus increasing $\hat{\sigma}_0^2$, the test statistic decreases. Therefore, a small blunder will not be flagged as an outlier. This aspect, on the other hand, can also be an advantage in the case of a small dataset and for the sake of a higher redundancy.

After the detection of an outlier, the actual localization of a blunder is implemented in the same way as in the data snooping technique (see Eq. 3.30). Again, the user can decide to either keep all observations and continue the program or to stop *OrthoPlus* in order to eliminate the erroneous observation with the maximum test statistic together with its components in the covariance matrices. This whole process of data screening consisting of either the global test and the data snooping or the τ -test should guarantee that all further results from the assessment of the parametric models as well as the variance component estimation will not be influenced by any observations with gross errors.

3.3 Parametric model surface fit

As already mentioned, the linear relationship between the height triplet h , H and N seems to be quite simple:

$$h - H - N = 0. \quad (3.35)$$

In practice however, some other factors have to be taken into consideration when combining the heterogeneous heights. Otherwise, this will lead to poorly estimated height values and to an inaccurate assessment of these resulting heights. Such factors can be (Fotopoulos, 2003):

- random errors as a result of deriving h , H and N ,
- datum inconsistencies between the different height types because of slightly different reference surfaces,
- systematic effects and distortions in the height data,
- assumptions and theoretical approximations made in computation of height data and
- temporal deviations of reference station monuments.

In order to compensate for these possible discrepancies, the incorporation of a parametric “corrector” surface model is essential in practice. The unknown model parameters can be estimated from a combined least-squares adjustment of ellipsoidal, orthometric and geoid height data. There are various

options to define a parametric surface model. In the beginning, *OrthoPlus* will ask the user to decide for one out of nine different model types, shown in the following list:

- A. first-order polynomial
- B. classic four-parameter
- C. classic five-parameter
- D. second-order polynomial
- E. differential similarity
- F. third-order polynomial
- G. fourth-order polynomial
- H. Legendre polynomial
- I. Fourier series

Among others, the program will provide the user in the sequel with results from various empirical and statistical tests which have been developed to assess the performance of the chosen parametric model. In order to contrast and to find the appropriate model for the specific dataset, the user has the possibility to re-run *OrthoPlus* several times, each time with another parametric model, and compare the results afterwards. The following sections will present the methodology and algorithm of the combined height adjustment, explain the background of the different parametric models and describe the methods to assess the model performance. Figure 3.4 summarizes these major steps and modules in a flowchart.

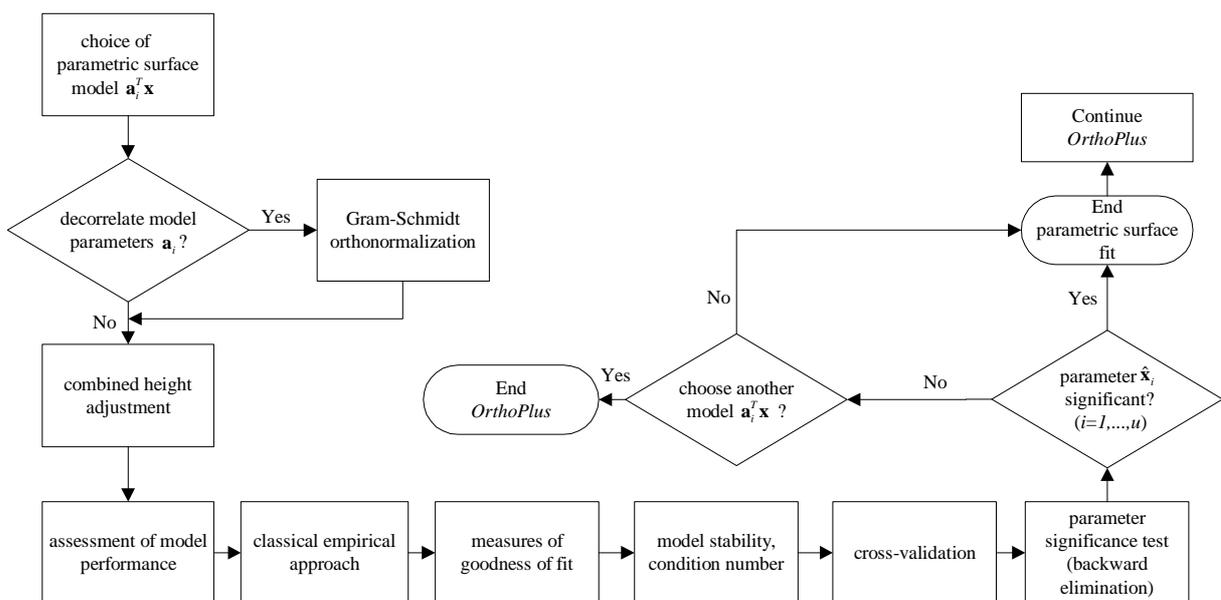


Figure 3.4: Flowchart of major steps of parametric surface fit (Fotopoulos and Sideris, 2005)

3.3.1 Combined height adjustment

The purpose of this combined 1D least-squares adjustment of GPS, levelling and geoid height data is to solve for the coefficients of a parametric surface model. This model will be included in the linear relationship of the height triplet h , H and N for dealing with systematic errors and datum inconsistencies. The input dataset for *OrthoPlus* (Figure 3.1) is a network of points with known ellipsoidal heights from GPS measurements, orthometric heights from levelling and gravity data and geoid heights from the gravimetric geoid model. *OrthoPlus* will provide the user with a map of the test area including these control station sites (`fp1ot_bms.m`) which will be called GPS-levelling benchmarks in the sequel. The observations, i.e., the height misclosures for each control point in the network, are derived by using this triplet of height info in the basic equation Eq. (3.35). This leads to the following observation equation for each GPS-levelling benchmark (Fotopoulos, 2003):

$$l_i = f_i + v_i^h - v_i^H - v_i^N, \quad (3.36)$$

where $v_i^{(\cdot)}$ are the zero-mean random errors in h , H and N and \mathbf{l} is the height misclosure vector representing the observations:

$$l_i = h_i - H_i - N_i. \quad (3.37)$$

The vector f_i is defined as a total (combined) correction term of the specific chosen parametric model for systematic errors and datum inconsistencies in the following deterministic parametric form:

$$f_i = \mathbf{a}_i^T \mathbf{x}, \quad (3.38)$$

where \mathbf{a}_i^T are the known base functions for the parametric model and \mathbf{x} is the $u \times 1$ vector of unknown parameters which will be estimated in the adjustment procedure. The number of unknowns depends on the selection of the parametric surface model.

The observation equation (3.36) can be reformulated to obtain the general linear functional model used for the combined height adjustment:

$$\mathbf{l} = \mathbf{Ax} + \mathbf{Bv}, \quad E\{\mathbf{v}\} = 0. \quad (3.39)$$

The explanation for the used variables is listed below:

I: $n \times 1$ misclosure vector of height triplet at each GPS-levelling benchmark

A: $n \times u$ design matrix dependent on choice of parametric model type

x: $u \times 1$ vector of unknown parameters corresponding to selected model

B: block-structured matrix consisting of $u \times u$ unit matrices **I**:

$$\mathbf{B} = [\mathbf{I} \quad -\mathbf{I} \quad -\mathbf{I}]$$

v: random errors

$$\mathbf{v} = [\mathbf{v}_h^T \quad \mathbf{v}_H^T \quad \mathbf{v}_N^T]^T$$

$E\{\mathbf{v}\} = 0$: zero mean of **v** defined by mathematical expectation operator E .

The corresponding stochastic model is given by (Fotopoulos, 2003)

$$\mathbf{C}_l = \sigma_0^2 \mathbf{Q} = \mathbf{Q} = \mathbf{P}^{-1} = \mathbf{C}_h + \mathbf{C}_H + \mathbf{C}_N, \quad (3.40)$$

where \mathbf{C}_l is the general $n \times n$ covariance matrix obtained by the sum of the covariance matrices \mathbf{C}_h , \mathbf{C}_H and \mathbf{C}_N for each height type which might either be known or equated with an identity matrix, according to the decision of the user. \mathbf{Q} is the cofactor matrix and σ_0^2 is the a-priori selected variance factor, in *OrthoPlus* fixed to 1. \mathbf{P} is the $n \times n$ block diagonal weight matrix determined to apply the least-squares minimization principle of

$$\mathbf{v}^T \mathbf{P} \mathbf{v} = \text{minimum}, \quad (3.41)$$

which makes it possible to estimate the unknown parameters of the parametric model by (Niemeier, 2002)

$$\hat{\mathbf{x}} = [\mathbf{A}^T \mathbf{P} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{P} \mathbf{l}, \quad (3.42)$$

with the cofactor matrix of the unknowns

$$\mathbf{C}_{\hat{\mathbf{x}}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1}. \quad (3.43)$$

The adjustment can be also used to solve for the combined adjusted residuals (Niemeier, 2002):

$$\mathbf{v} = \mathbf{P}^{-1}\mathbf{B}^T\mathbf{k} , \quad (3.44)$$

$$\text{with } \mathbf{k} = (\mathbf{B}\mathbf{P}^{-1}\mathbf{B}^T)(\mathbf{I} + \mathbf{A}\hat{\mathbf{x}}) , \quad (3.45)$$

where all parameters of Eq. (3.44) and (3.45) have been already defined above.

The stochastic model and the separation of residuals according to the height types will be further discussed in the section dealing with the variance component estimation (see 3.4).

3.3.2 Selection of parametric model

The results of the combined height adjustment will be directly related to the choice of the parametric surface model. As already mentioned, the modelling options are various. At the same time, the selection of the appropriate analytical model, i.e., the type and extent, is quite arbitrary, both in choosing the model type and in assessing its performance. The question for a best model also varies from region to region depending on the data distribution, density and quality. Fotopoulos (2003) proposed to limit the choice of parametric surfaces to two pre-specified families:

- polynomial models up to 4th order and
- similarity-based transformation model and its more simplified forms.

For the sake of comparison, a third type has been added within the scope of this thesis:

- two different kinds of orthogonal polynomials with 4th order (Legendre and Fourier)

The reason for choosing this third type of parametric surface model is to make sure that all correlations among the model parameters are cancelled out since they might distort the results. Section 3.3.3 will introduce another possibility to derive a model with uncorrelated parameters: the Gram-Schmidt orthonormalization. This procedure, however, is just a re-formulation of the selected model mapped to an empirical base which is orthonormal.

3.3.2.1 Polynomial models up to 4th order

One option to model the parametric surface, denoted by $\mathbf{a}_i^T\mathbf{x}$, is to apply a polynomial (of various orders) (Lancaster and Salkauskas, 1986). Its general form can be described by the following multiple regression equation (MRE):

$$\mathbf{a}_i^T \mathbf{x} = \sum_{m=0}^M \sum_{n=0}^N (\varphi_i - \bar{\varphi})^n (\lambda_i - \bar{\lambda})^m x_q, \quad (3.46)$$

where x_q represents the q unknown parameters and $\bar{\varphi}, \bar{\lambda}$ the mean values of the latitude φ_i and longitude λ_i of all GPS-levelling benchmarks.

In order to test this model type, the user of *OrthoPlus* should choose one of the models A (1st order), D (2nd order), F (3rd order) or G (4th order) from the menu listed in the beginning of section 3.3.

3.3.2.2 Similarity-based transformation model and simplified forms

These other types of models are based on the general 7-parameter similarity datum shift transformation or also referred to as differential similarity model (model E) in *OrthoPlus*. Simplification leads to the classic 4-parameter model (model B), implemented as (Hofmann-Wellenhof and Moritz, 2005):

$$\mathbf{a}_i^T \mathbf{x} = x_1 + x_2 \cos \varphi_i \cos \lambda_i + x_3 \cos \varphi_i \sin \lambda_i + x_4 \sin \varphi_i \quad (3.47)$$

By adding a fifth parameter, denoted by $x_5 \sin^2 \varphi_i$, the extended classic 5-parameter model (model C) can be obtained.

The third and most complicated form of a differential similarity transformation model (Model E) is defined as (Fotopoulos, 2003):

$$\begin{aligned} \mathbf{a}_i^T \mathbf{x} = & x_1 \cos \varphi_i \cos \lambda_i + x_2 \cos \varphi_i \sin \lambda_i + x_3 \sin \varphi_i + x_4 \left(\frac{\sin \varphi_i \cos \varphi_i \sin \lambda_i}{W} \right) + \\ & x_5 \left(\frac{\sin \varphi_i \cos \varphi_i \cos \lambda_i}{W} \right) + x_6 \left(\frac{1 - f^2 \sin^2 \varphi_i}{W} \right) + x_7 \left(\frac{\sin^2 \varphi_i}{W} \right) \end{aligned} \quad (3.48)$$

with $W = \sqrt{1 - e^2 \sin^2 \varphi_i}$ and where e^2 is the eccentricity and f the flattening of the reference ellipsoid.

3.3.2.3 Orthogonal polynomials

To make sure that the results of the combined height adjustment and the assessment of the model performance are not influenced by correlations among the model parameters (Fotopoulos, 2003), Lancaster et al. (1986) and Crain (1970) encourage the reader to use orthogonal polynomials for surface

interpolation, especially in the case of datasets with a large number of points and for polynomials of order greater than five. The reason is to avoid large and ill-conditioned systems of equations which will cause computational problems and inaccuracies for the inversion of large matrices. Therefore, *OrthoPlus* provides the possibility to choose one of the following orthogonal polynomial models of 4th order:

- Model H: fully normalized spherical harmonics (Legendre polynomials)
- Model I: Fourier polynomials in two variables (polar coordinates)

Fully normalized spherical harmonics (Legendre polynomials) of 4th degree The fully normalized spherical harmonics of 4th degree k and up to the 4th order m are a function of geodetic coordinates φ_i and λ_i with the colatitude $\nu_i=90^\circ-\varphi_i$ at the i -th GPS-levelling benchmark point. The basis function is defined as follows (Hofmann-Wellenhof and Moritz, 2005):

$$\mathbf{a}_i^T \mathbf{x} = \sum_{k=0}^4 \sum_{m=0}^k \left[x_{km}^a \bar{R}_{km}(\nu_i, \lambda_i) + x_{km}^b \bar{S}_{km}(\nu_i, \lambda_i) \right], \quad (3.49)$$

where x_{km}^a and x_{km}^b are the unknown parameters estimated in the combined height adjustment and where the known base functions are defined as

$$\begin{aligned} \bar{R}_{km}(\nu_i, \lambda_i) &= \bar{P}_{km}(t_i) \cos m \lambda_i \quad \text{and} \\ \bar{S}_{km}(\nu_i, \lambda_i) &= \bar{P}_{km}(t_i) \sin m \lambda_i. \end{aligned} \quad (3.50)$$

$\bar{P}_{km}(t_i)$ are the normalized Legendre polynomials up to the order $k=4$ with the substitution $t_i = \cos(\nu_i)$. They can be easily obtained by applying the Matlab function `legendre.m`. The corresponding design matrix is implemented as follows:

$$\mathbf{A} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{P}_{00}(t_i) & \bar{P}_{10}(t_i) & \bar{P}_{11}(t_i) \cos(1 \cdot \lambda_i) & \bar{P}_{11}(t_i) \sin(1 \cdot \lambda_i) & \cdots & \bar{P}_{44}(t_i) \sin(4 \cdot \lambda_i) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (3.51)$$

Those sine-terms of the base functions which are zero for $m=0$ with $\bar{S}_{k0}(\nu_i, \lambda_i) = \bar{P}_{k0}(t_i) \sin(0 \cdot \lambda_i) = 0$ are not included in the design matrix \mathbf{A} in order to avoid rank-deficiency. Therefore, the number of unknown parameters is $u = (k+1)^2 = 25$ for $k=4$.

3.3.2.4 Summary

The following table will give an overview of the base functions used in the different parametric surface models. The numbers in brackets behind the model names show the number of parameters for each model. These base functions also define the structure of the design matrix, i.e., they will form the columns of **A**.

Table 3.2: Summary of parametric surface models

model	base functions
A. 1 st order polynomial (3)	$1, \delta\varphi, \delta\lambda$
B. classic 4-parameter (4)	$1, \cos \varphi \cos \lambda, \cos \varphi \sin \lambda, \sin \varphi$
C. classic 5-parameter (5)	$1, \cos \varphi \cos \lambda, \cos \varphi \sin \lambda, \sin \varphi, \sin^2 \varphi$
D. 2 nd order polynomial (6)	$1, \delta\varphi, \delta\lambda, \delta\varphi\delta\lambda, \delta\varphi^2, \delta\lambda^2$
E. differential similarity (7)	$\cos \varphi \cos \lambda, \cos \varphi \sin \lambda, \sin \varphi, \frac{\sin \varphi \cos \varphi \sin \lambda}{W},$ $\frac{\sin \varphi \cos \varphi \cos \lambda}{W}, \frac{1 - f^2 \sin^2 \varphi}{W}, \frac{\sin^2 \varphi}{W}; W = \sqrt{1 - e^2 \sin^2 \varphi}$
F. 3 rd order polynomial (10)	$1, \delta\varphi, \delta\lambda, \delta\varphi\delta\lambda, \delta\varphi^2, \delta\lambda^2, \delta\varphi^2\delta\lambda, \delta\varphi\delta\lambda^2, \delta\varphi^3, \delta\lambda^3$
G. 4 th order polynomial (15)	$1, \delta\varphi, \delta\lambda, \delta\varphi\delta\lambda, \delta\varphi^2, \delta\lambda^2, \delta\varphi^2\delta\lambda, \delta\varphi\delta\lambda^2, \delta\varphi^3, \delta\lambda^3, \delta\varphi^2\delta\lambda^2,$ $\delta\varphi^3\delta\lambda, \delta\varphi\delta\lambda^3, \delta\varphi^4, \delta\lambda^4$
H. Legendre polynomials (25)	$\bar{P}_{00}(t), \bar{P}_{10}(t), \bar{P}_{11}(t) \cos(1 \cdot \lambda), \bar{P}_{11}(t) \sin(1 \cdot \lambda), \dots, \bar{P}_{44}(t) \sin(4 \cdot \lambda)$
I. Fourier series (9)	$\frac{1}{2} \hat{u}_0, \hat{u}_1, \hat{v}_1, \hat{u}_2, \hat{v}_2, \hat{u}_3, \hat{v}_3, \hat{u}_4, \hat{v}_4$

3.3.3 Gram-Schmidt orthonormalization

After selecting the parametric surface model which automatically defines the form and structure of the design matrix **A** used in the combined height adjustment, *OrthoPlus* asks if the user wants to keep this original design matrix and continue with the program or if the Gram-Schmidt orthonormalization should be applied. This procedure, implemented by the function `fgramschmidt.m`, decorrelates the parameters of the parametric surface model. This means that the arbitrary basis of the selected model, represented by the column vectors of the original design matrix, will be transformed and normed so that the basis becomes orthonormal. An orthonormal basis is defined in the way that firstly the scalar product of any pair of basis vectors is zero and secondly each basis vector is of unit length (Carroll and

Green, 1997). The advantage of using the Gram-Schmidt orthonormalization procedure is that we get a new set of uncorrelated parameters by means of a relatively simple method. These re-formulated parameters are easier to determine independently from each other since any possible correlations have been cancelled out.

The disadvantage compared to the direct use of orthogonal polynomials is that the new uncorrelated model is just created by a re-formulation which does not allow an analytical interpretation. It is an empirical set of base functions. Since the base function and the unknown coefficients are computed for particular data only, it can not be applied as general base function in order to compute new points, for example.

The algorithm itself can be implemented relatively straightforward (Carroll and Green, 1997). The first step was to select the first column \mathbf{a}_1 of the design matrix as the first reference vector \mathbf{b}_1 of the new basis and keep it fixed:

$$\mathbf{b}_1^T = \mathbf{a}_1^T . \quad (3.56)$$

In the sequel, one basis vector after the other will be derived by the following equation so that the resulting sets are mutually orthogonal:

$$\mathbf{b}_k^T = \mathbf{a}_k^T - \left[\frac{\mathbf{a}_k^T \mathbf{b}_{k-1}}{\mathbf{b}_{k-1}^T \mathbf{b}_{k-1}} \right] \mathbf{b}_{k-1}^T - \dots - \left[\frac{\mathbf{a}_k^T \mathbf{b}_1}{\mathbf{b}_1^T \mathbf{b}_1} \right] \mathbf{b}_1^T \quad k = 1, \dots, u \quad (3.57)$$

where $\mathbf{b}_1, \dots, \mathbf{b}_u$ are the u vectors of the new orthogonal basis and $\mathbf{a}_1, \dots, \mathbf{a}_u$ are the u basis vectors of the original parametric surface model, defined by the column vectors of the design matrix \mathbf{A} . In order to get an orthonormal basis $\mathbf{b}_1^*, \dots, \mathbf{b}_u^*$, the orthogonal basis has to be normalized by dividing each vector component \mathbf{b}_k^T by its length:

$$\mathbf{b}_k^{*T} = \frac{\mathbf{b}_k^T}{\|\mathbf{b}_k^T\|} \quad k = 1, \dots, u . \quad (3.58)$$

Thus, the vectors $\mathbf{b}_1^*, \dots, \mathbf{b}_u^*$ form the new design matrix which will be used in the combined height adjustment and which guarantees that the parameters of the selected model are uncorrelated.

After applying either the original or the orthonormal model in the adjustment, it is very important to assess its performance as well as to evaluate the significance of the estimated parameters for a general surface. Therefore, the next section will provide details for the implementation of statistical and empirical tests which have been implemented in *OrthoPlus* to test the results of the combined least-squares adjustment of ellipsoidal, orthometric and geoid heights.

3.3.4 Assessing the parametric model performance

A very important module to help the user to come to a decision for an appropriate parametric surface model is the assessment of the model performance. The user can test the dataset for a particular region with various models in *OrthoPlus*. The program output presents the results of the four following tests:

- classical empirical approach
- assessing the goodness of fit
- cross-validation
- testing parameter significance

Another component provided by *OrthoPlus* is the condition number which gives information about the numerical stability of the selected model and which is the ratio of the maximum λ_{\max} and minimum λ_{\min} eigenvalues:

$$\text{cond}(\mathbf{A}^T \mathbf{A}) = \frac{\lambda_{\max}}{\lambda_{\min}}. \quad (3.59)$$

OrthoPlus will also provide the user with a visualization of each selected parametric model (`fplot_surface.m`) in the form of a 3D-plot of the model over the test area (Figure 3.5) and a plot of the projected model into a map (Figure 3.6). Additionally, the positions of the benchmark points are shown in both figures.

From results of these tests, it cannot be guaranteed to derive the obviously best model. The user has to take charge for choosing the model which meets his requirements and expectations as good as possible. The purpose of the implemented tests is just to simplify the user's decision and consider the consequences of the model selection from all sides.

The next sections will give an overview and a description of these tests.

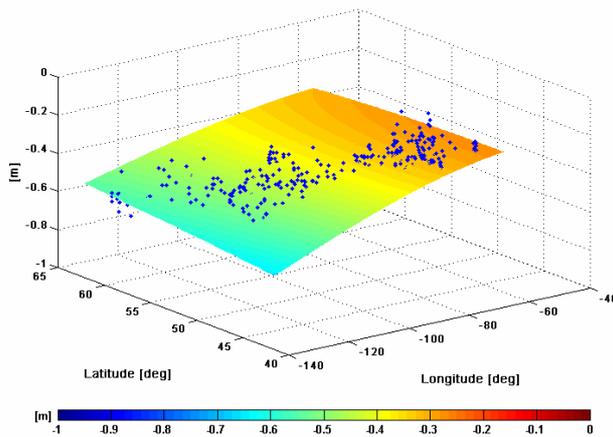


Figure 3.5: 3D parametric model

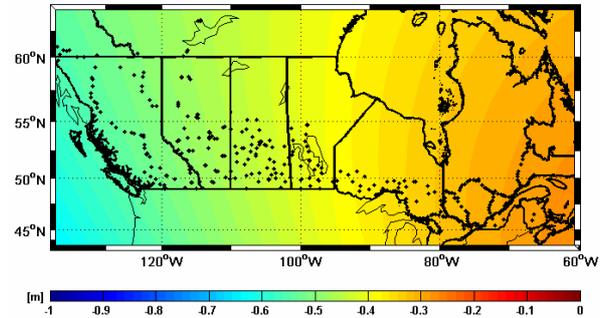


Figure 3.6: Projection of model in area map

3.3.4.1 Classical empirical approach

The reason for incorporating a parametric surface model is to compensate for discrepancies between the ellipsoidal, orthometric and geoid height. That implies that after applying the combined height adjustment with a specific model $\mathbf{a}_i^T \mathbf{x}$, the original height misclosures $h_i - H_i - N_i$ should be reduced to residuals v_i oscillating around a zero mean for all benchmarks of the network:

$$v_i = h_i - H_i - N_i - \mathbf{a}_i^T \hat{\mathbf{x}}. \quad (3.60)$$

OrthoPlus provides a plot (Figure 3.7) of the height misclosures for each benchmark point before and after the fit in which this described shift of the curves can be clearly recognized.

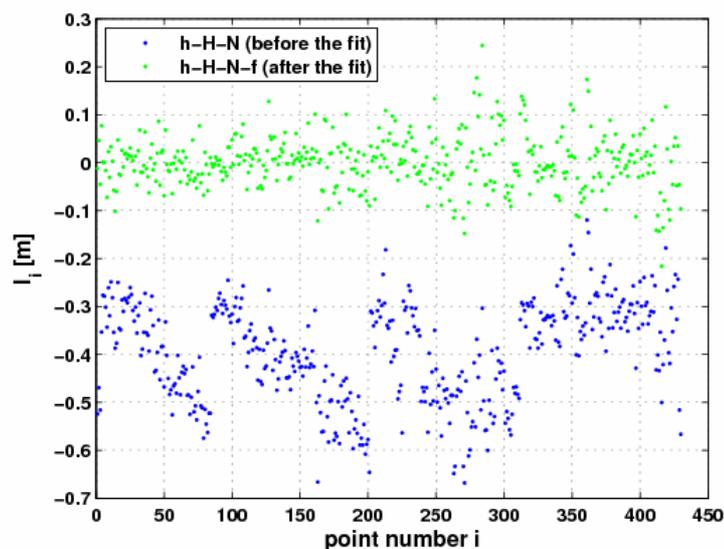


Figure 3.7: Example of height misclosures before and after parametric model fit

It is assumed that the set of residuals is smallest for the ‘best’ fitting parametric model with respect to its precision. However, the precision of the model only indicates how well the data sets fit each other. Therefore, this selected model should not immediately be assigned as the most appropriate parametric model in view of accuracy assessment or prediction capability as it is known that the lowest RMS usually corresponds to the highest order model.

3.3.4.2 Assessing the goodness of fit

The goodness of parametric model fit for a discrete set of points is expressed in terms of the coefficient of determination (Sen and Srivastava, 1990):

$$R^2 = 1 - \frac{\sum_{i=1}^n v_i^2}{\sum_{i=1}^n (l_i - \bar{l}_i)^2}, \quad (3.61)$$

where $\sum_{i=1}^n v_i^2$ is the sum of the squared residuals adjusted for each of the n stations by Eq. (3.60) in the fit and $\sum_{i=1}^n (l_i - \bar{l}_i)^2$ is the sum of the squared differences between the original height misclosures $l_i = h_i - H_i - N_i$ and their mean value \bar{l}_i . This statistical measure R^2 takes values between 0 and 1, with values closer to 1 implying a better fit since in this case the sum of the squared residuals is approaching 0.

In order to account for the number of parameters u in the model and for the sample size n , the coefficient of determination R^2 can be modified as follows (Sen and Srivastava, 1990):

$$R_\alpha^2 = 1 - \frac{\left[\sum_{i=1}^n v_i^2 \right] / (n-u)}{\sum_{i=1}^n (l_i - \bar{l}_i)^2 / (n-1)} = 1 - \frac{(n-1)}{(n-u)} (1 - R^2). \quad (3.62)$$

This new measure is called adjusted coefficient of determination. It can be interpreted in the same manner as the coefficient of determination, with $R_\alpha^2 = 1$ indicating a perfect parametric model fit. Such a modification is especially necessary if the sample size is small and the number of parameters large,

i.e., if the data redundancy (or degrees of freedom) $r = n - u$ is small. In this case, the R^2 value will increase and thus, be mistakably interpreted as a high quality of fit, even if this is not correct. Here again, the user should be aware that both measures of fit should be only interpreted together with the results of the other tests described in this section. At least, the method of assessing the goodness of fit can be used for determining inappropriate models.

3.3.4.3 Cross-validation

Another empirical test implemented in *OrthoPlus* besides the classical empirical approach is the cross-validation (`fcrossvalid.m`) which has the advantage that it will provide a more realistic measure of the accuracy of the chosen parametric model. This method applies the combined least-squares adjustment with a chosen subset of vertical control points in order to compute the model parameters. Later on, the difference between the estimated and the known height misclosure at the points not inherent in the subset will be used to assess the accuracy of the model. Since the objective of a least-squares adjustment is to use as many observations as possible and at the same time, the value resulting from the cross-validation will be dependent on the accuracy of the subset of the points, the special leave-one-out cross-validation is applied in *OrthoPlus*. The process can be summarized in five steps (Moore and Lee, 1994; Fotopoulos and Sideris, 2005):

- (i) Select a subset of $n-1$ GPS-levelling benchmarks, i.e., leave one point P_i out.
- (ii) Apply combined height adjustment for this subset without P_i to compute the parameters $\hat{\mathbf{x}}$ of the selected model.
- (iii) Use selected model with estimated parameters $\hat{\mathbf{x}}$ to predict a new height misclosure \hat{l}_i at point P_i which is pretended to be unknown.
- (iv) Compare predicted value \hat{l}_i with known height misclosure l_i at P_i : $\Delta_i = \hat{l}_i - l_i$.
- (v) Repeat (i) – (iv) for each GPS-levelling benchmark in the network and compute root mean square of Δ_i by $RMS_{\Delta} = \sqrt{\mu_{\Delta}^2 + \sigma_{\Delta}^2}$ (see Eq. 3.2).

In the sequel, this RMS_{Δ} can be interpreted as a value to evaluate the accuracy of the selected parametric surface model with respect to its performance in predicting new points.

3.3.4.4 Testing parameter significance

In fact, the more parameters are included in a parametric surface model, the better it will fit the dataset. However, it must be also considered how well this model can predict future data. Besides, any insignificant parameters in the model can lead to wrong assumptions when assessing the model performance. The reason is that they will bias the other terms but only if they are correlated. That is known as over-parameterization. In order to avoid this effect, as well as to keep the model simple and at the same time the computational efficiency high, the significance of the parameters in the selected model can be tested. However, this test implies that nested models have been used. In a loose sense, we can say that one model A is nested within another model B if it is a special case of the model B and can be derived just by deleting some terms of model B. That is indeed the case for the model families described in section 3.3.2 and used in *OrthoPlus*.

In general, there are three different possibilities to test, identify and eliminate unnecessary terms to derive a more simplified parametric model:

- the backward elimination,
- the forward selection and
- the stepwise procedure.

Since *OrthoPlus* is a semi-automated program, its only objective is to test the terms of the selected model and thereupon to identify and indicate any insignificant parameters. With this knowledge, the user can decide if he wants to re-run the program by using another model with less parameters. That means that *OrthoPlus* will not automatically eliminate insignificant parameters. For this purpose, the backward elimination worked out to be the best procedure. The usage in the program is described in more detail in the sequel. The reader is referred to Sen et al. (1990), Koch (1999) or Fotopoulos and Sideris (2005) for details regarding the other parameter significance tests.

Backward elimination At the start, *OrthoPlus* asks the user to choose a model. With respect to the backward elimination procedure, this selected model should have as many parameters as possible since it will be assumed to be the most extended form of the model. Within this selected highest order model, each of the u parameters $\hat{\mathbf{x}}_i$ will be tested individually if it is significant or not. The corresponding null hypothesis H_0 in comparison to the alternative hypothesis H_a looks as follows:

$$H_0 : \hat{\mathbf{x}}_i = 0 \longleftrightarrow H_a : \hat{\mathbf{x}}_i \neq 0. \quad (3.63)$$

The test statistic F can be computed via (Dermanis and Rossikopoulos, 1991):

$$F = \frac{\hat{\mathbf{x}}_i^T \mathbf{N}_{ii} \hat{\mathbf{x}}_i}{k \hat{\sigma}_0^2}, \quad (3.64)$$

where $\hat{\mathbf{x}}_i$ is the tested parameter which has been estimated in the combined height adjustment, k is the number of parameters tested which is one in our case, $\hat{\sigma}_0^2$ is the a-posteriori variance factor and \mathbf{N}_{ii} is the corresponding diagonal element in the normal matrix with

$$\mathbf{N} = \mathbf{A}^T \mathbf{P} \mathbf{A}. \quad (3.65)$$

The test statistic F follows the F-distribution under the null hypothesis H_0 . The tested parameter $\hat{\mathbf{x}}_i$ will be flagged as insignificant if the null hypothesis is accepted, i.e., if the following condition is fulfilled:

$$F \leq F_{\alpha, k, f} \quad (3.66)$$

where $F_{\alpha, k, f}$ is the test threshold of the acceptance region with $k=1$ and the degrees of freedom $f = n - u$. It can be extracted from the standard statistical tables (see Niemeier, 2002). The significance level α can be chosen by the user of *OrthoPlus*.

This testing procedure will be repeated step by step for all model parameters. At the end all insignificant parameters of the selected model will be listed and the user can decide to either accept the model and continue or re-run the program with another parametric surface model.

3.4 Variance component estimation (VCE)

Adjustment procedures of geodetic data are based on an observation model which consists mainly of two mathematical models. That is, on the one hand, the functional model which defines the deterministic mathematical relationship between unknown parameters and observations together with the proper modelling of systematic effects. These types of errors which are for example caused by using defective instruments, affect whole sets of observations. The general linear functional model used for the combined height adjustment has been introduced by Eq. (3.39) and explained in detail in section 3.3.1:

$$\mathbf{l} = \mathbf{Ax} + \mathbf{Bv}, \quad E\{\mathbf{v}\} = 0. \quad (3.67)$$

In the case of a combined height adjustment, the observation vector \mathbf{l} is defined as the height misclosure at each GPS-levelling benchmark (Eq. 3.37) and \mathbf{v} is a vector of zero-mean random errors. The other component of every least-squares adjustment is to define a realistic stochastic model in the form of a covariance matrix which describes the errors of a functional model, being random in nature. These errors can be for example due to observational noise which is the uncertainty of the measurements. The stochastic model in the combined height adjustment is given by (see also Eq. 3.40):

$$E\{\mathbf{vv}^T\} = \mathbf{C}_l = \sigma_0^2 \mathbf{Q} = \mathbf{Q} = \mathbf{P}^{-1}, \quad (3.68)$$

where \mathbf{C}_l is the general $n \times n$ variance-covariance matrix or shortly covariance matrix. \mathbf{P} is the $n \times n$ block diagonal weight matrix and an inverse of the symmetric positive definite cofactor matrix \mathbf{Q} . σ_0^2 is the a-priori selected variance factor which is often set to one.

The corresponding a-posteriori variance factor $\hat{\sigma}_0^2$ can be computed via Eq. (3.10) by applying the adjusted residual vector \mathbf{v} and will be used to modify the prior covariance matrix by a simple scaling:

$$\tilde{\mathbf{C}}_l = \hat{\sigma}_0^2 \mathbf{Q}. \quad (3.69)$$

However, a stochastic model implemented before an adjustment is often not precise or even incorrect and would therefore have an impact on the estimation results. That does not necessarily mean that the parameter solution will be directly influenced by an incorrect stochastic model but the a-posteriori variance factor will be (Gao and Sideris, 2005). The classic stochastic model described above (Eq. 3.68), for example, only allows to adjust for one common variance factor of the covariance matrix. If there are observations of different types, as it is in our case by using heterogeneous heights (ellipsoidal, orthometric and geoid height), one variance factor is definitely not enough to improve the covariance matrix. Thus, it is very important to apply the well known statistical tool of variance component estimation (or shortly VCE). In general, the VCE makes it possible to estimate one variance component for each observation type (Niemeier, 2002) in order to obtain a more precise stochastic model. In this thesis, the VCE is mainly used for refinement of the covariance matrices of the three different height types in order to assess the error budget of all data used in the combined height adjustment. There are several different techniques to implement a variance component estimation. The

program *OrthoPlus* provides the possibility to choose between four different possibilities from the following menu:

- IMINQUE (iterative minimum norm quadratic unbiased estimation),
- IAUE (iterative almost unbiased estimation) with N separated in components,
- IAUE with summarized N or
- no VCE algorithm.

The IAUE method can either be applied by defining the geoid undulation N as a variable under the assumption that it has only one common variance component (IAUE with summarized N) or by separating it in another three components where each of them has their own covariance matrix (IAUE with N separated in components). The IMINQUE algorithm also uses this splitting up of N . The user has also the possibility to run the combined height adjustment without the VCE algorithm. Furthermore, he gets the choice between using the original fully-populated covariance matrices or only their diagonal forms.

The decision for implementing the IMINQUE as well as the IAUE method in *OrthoPlus* was made according to the recommendation in Fotopoulos (2003) of the most appropriate VCE methods applied to the combined height adjustment. The criteria have been computational load, balanced versus unbalanced data, ease of implementation, flexibility for modifications, unbiasedness and non-negative variance factors.

A variable which will be needed in the VCE algorithms is the adjusted residual vector \mathbf{v} which is modified in the following way:

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_h^T & \mathbf{v}_H^T & \mathbf{v}_N^T \end{bmatrix}^T, \quad (3.70)$$

where \mathbf{v}_h , \mathbf{v}_H and \mathbf{v}_N are the $n \times 1$ vectors of random errors for each of the height types. These three sub-vectors are correlated among the same height type, but uncorrelated with each other. Therefore, the estimated adjusted residuals can be computed separately for each height type, generally denoted by (Gao and Sideris, 2005):

$$\mathbf{v}_k = -(\mathbf{Q}_k \mathbf{R}) \mathbf{l}, \quad k \in [h, H, N], \quad (3.71)$$

where \mathbf{Q}_k describes the corresponding cofactor matrix which is either known beforehand or defined as an identity matrix. \mathbf{l} is the observation vector and \mathbf{R} is a symmetric matrix which can be derived by Eq. (3.5).

The only disadvantage when applying the VCE algorithms to the combined height adjustment is the fact that the covariance matrix of the observations \mathbf{C}_l has to be known in order to compute the \mathbf{R} matrix, i.e., \mathbf{C}_l is derived by means of initial estimates of the unknown variance components. However, the deficiency of good initial estimates would be exactly the reason to induce the implementation of a VCE. Thus, the MINQUE or AUE method are only locally best estimators which provide variance components dependent on the choice of the a-priori variance factors. Iterative approaches (IMINQUE or IAUE), as they are implemented in *OrthoPlus*, can compensate for this problem and are able to provide a globally best estimator which is less dependent on the initial estimates for the unknown variance components. This is especially important if the initial estimates are unreliable or uncertain (Fotopoulos, 2003).

Figure 3.8 gives an overview of the iteration loop as it is used for both VCE methods. The vector containing the unknown variance components $\hat{\sigma}_h^2$, $\hat{\sigma}_H^2$ and $\hat{\sigma}_N^2$ is denoted as $\hat{\boldsymbol{\theta}}$. The a-priori variance components $\hat{\theta}_k^\circ = \sigma_k^2$ have been initially specified for each height type $k = h, H, N$. These approximations are used in the first run of the VCE in order to derive a first variance estimate $\hat{\theta}_k^\alpha$ for each height component which is applied to the corresponding a-priori cofactor matrix as scale factor. In this way, the a-priori CV matrix can be modified and updated as well. In the next iteration $\alpha = \alpha + 1$, the VCE is repeated by defining the resulting variance estimates from iteration α as the a-priori variances and a new set of variance estimates $\hat{\theta}_k^{\alpha+1}$ can be derived. The convergence criterion developed in *OrthoPlus* stops the iteration procedure as soon as the variance estimates approach unity (Koch, 1999):

$$\hat{\boldsymbol{\theta}}^\alpha = \begin{bmatrix} \hat{\theta}_h^\alpha \\ \hat{\theta}_H^\alpha \\ \hat{\theta}_N^\alpha \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \quad (3.72)$$

This criterion can be realized by testing if the difference between the variance estimates from two successive iterations is smaller than a certain value, e.g. 0.0001:

$$|\hat{\theta}^\alpha - \hat{\theta}^{\alpha-1}| < 0.0001. \quad (3.73)$$

The user can choose this threshold in *OrthoPlus*. The condition in Eq. (3.73) will be tested for all height components to make sure that the criterion in (3.72) is fulfilled since the convergence rate of each of them is different. Therefore, the computational efficiency of the entire algorithm might be affected negatively by only one unreliable a-priori component. *OrthoPlus* provides a visualization of this convergence behaviour by a plot showing the variance estimates for each iteration (see example in Figure 3.9). Eventually, the final variance components $\hat{\sigma}_k^2 = \hat{\theta}_k$ for each of the three height types is defined as the product of the variance estimates $\hat{\theta}_k^\alpha$ which were estimated within n iterations of a VCE procedure (Niemeier, 2002):

$$\hat{\sigma}_k^2 = \hat{\theta}_k = \prod_{\alpha=0}^n \hat{\theta}_k^\alpha, \quad k \in [h, H, N]. \quad (3.74)$$

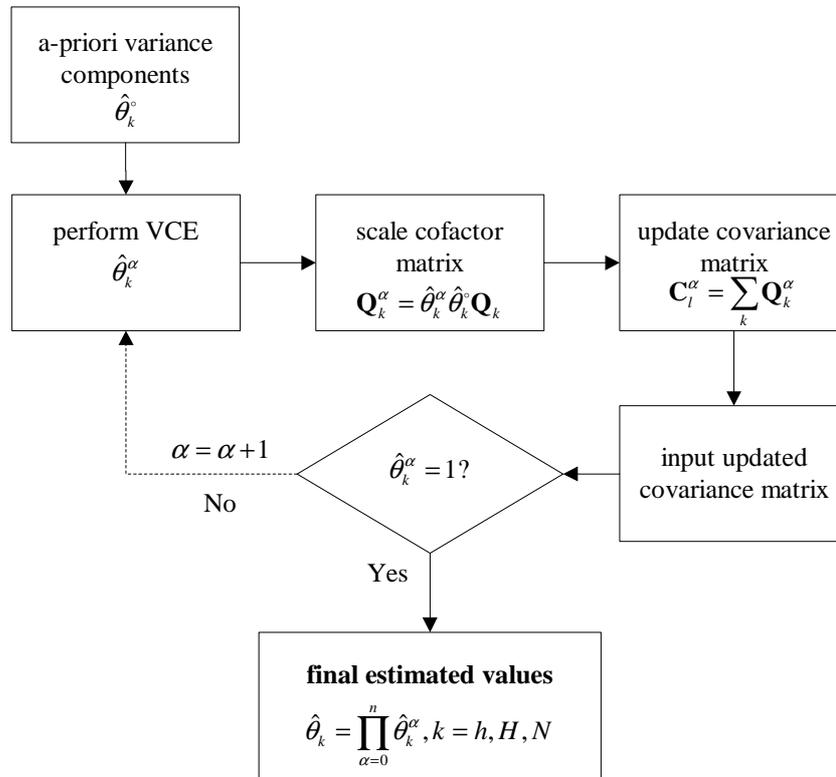


Figure 3.8: Iterative VCE computational scheme (Fotopoulos, 2003)

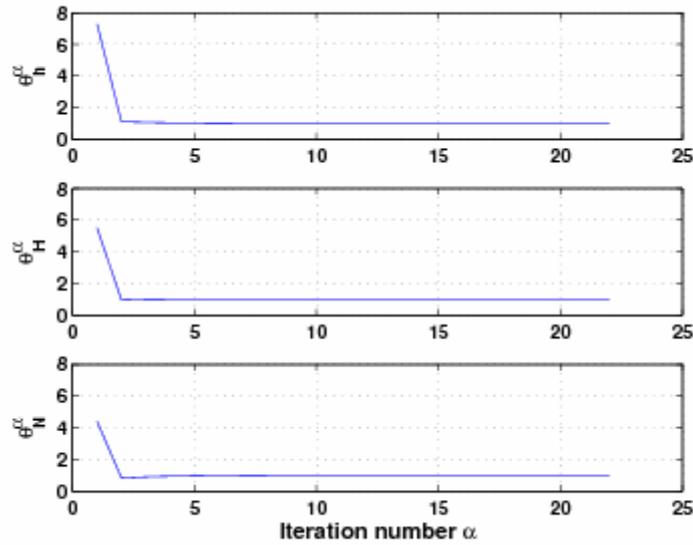


Figure 3.9: Convergence behaviour of variance estimates

The next two sections will give an application-oriented description of both VCE algorithms. For more detail, the reader is referred to Fotopoulos (2003) and Koch (1999).

3.4.1 Iterative minimum norm quadratic unbiased estimation (IMINQUE)

As already mentioned, one of the objectives of the VCE applied to the combined height adjustment, is to compute a variance component for each height type. In contrast to the general stochastic model from above with only one observation type (Eq. 3.68), the cofactor matrix is not explicitly given when applying the combined height adjustment, but can be derived by summing up the known positive definite cofactor matrices for the ellipsoidal, orthometric and geoid height:

$$\mathbf{Q} = \mathbf{Q}_h + \mathbf{Q}_H + \mathbf{Q}_N, \quad (3.75)$$

and the covariance matrix is defined as (Niemeier, 2002):

$$\mathbf{C}_l = \sigma_h^2 \mathbf{Q}_h + \sigma_H^2 \mathbf{Q}_H + \sigma_N^2 \mathbf{Q}_N, \quad (3.76)$$

where σ_h^2 , σ_H^2 and σ_N^2 are the unknown individual variance components which will be solved for in the VCE. Applying the general minimum norm quadratic unbiased estimation (MINQUE) obtains a solution of these variance components by solving for the following system (Fotopoulos, 2003):

$$\begin{bmatrix} tr(\mathbf{RQ}_h\mathbf{RQ}_h) & tr(\mathbf{RQ}_h\mathbf{RQ}_H) & tr(\mathbf{RQ}_h\mathbf{RQ}_N) \\ tr(\mathbf{RQ}_H\mathbf{RQ}_h) & tr(\mathbf{RQ}_H\mathbf{RQ}_H) & tr(\mathbf{RQ}_H\mathbf{RQ}_N) \\ tr(\mathbf{RQ}_N\mathbf{RQ}_h) & tr(\mathbf{RQ}_N\mathbf{RQ}_H) & tr(\mathbf{RQ}_N\mathbf{RQ}_N) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_h^2 \\ \hat{\sigma}_H^2 \\ \hat{\sigma}_N^2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_h^T \mathbf{Q}_h \mathbf{v}_h \\ \mathbf{v}_H^T \mathbf{Q}_H \mathbf{v}_H \\ \mathbf{v}_N^T \mathbf{Q}_N \mathbf{v}_N \end{bmatrix}, \quad (3.77)$$

where $tr(\cdot)$ is the trace operator. Eq. (3.77) can be summarized in the well-known form (Koch, 1999)

$$\mathbf{S}\hat{\boldsymbol{\theta}} = \mathbf{q}. \quad (3.78)$$

The vector $\hat{\boldsymbol{\theta}}$ contains the unknown variance components $\hat{\sigma}_h^2$, $\hat{\sigma}_H^2$ and $\hat{\sigma}_N^2$. If the system in Eq. (3.77) or (3.78) is solved, the three properties of MINQUE, which are minimum norm, translation invariance and unbiasedness are guaranteed. However, to ensure the VCE algorithm to be a globally best estimator, the corresponding iterative approach, the iterative minimum norm quadratic unbiased estimation (IMINQUE), has been implemented in *OrthoPlus*. That means that the VCE procedures will be repeated, as described above, till the convergence criterion in Eq. (3.72) is satisfied. This leads to the final estimated variance component values computed in Eq. (3.74).

It should be noted that using the IMINQUE method might imply some problems. In some cases, the property of unbiased values cannot be verified (Fotopoulos, 2003). Furthermore, it is possible that this algorithm does not achieve any convergent solution during the iteration since the variances do not satisfy the criteria of non-negative variance values (i.e., $\hat{\boldsymbol{\theta}} \in \mathfrak{R}_+^3$). This problem appeared when testing the IMINQUE algorithm on the Canadian dataset and was solved by the application of another type of VCE, the IAUE, which will be explained in the next section.

Remark on an alternate formulation of N Instead of computing different variance components for the ellipsoidal, the orthometric and the geoid height, Huang et al. (2005) proposes to separate the errors for the geoid height N into further components for the satellite gravity model, terrestrial gravity data, and the combined gravity model, denoted by σ_{SG}^2 , σ_{TG}^2 and σ_{CG}^2 . In this case, the covariance matrix of the observations is defined as:

$$\mathbf{C}_l = \sigma_h^2 \mathbf{Q}_h + \sigma_H^2 \mathbf{Q}_H + \sigma_{SG}^2 \mathbf{Q}_{SG} + \sigma_{TG}^2 \mathbf{Q}_{TG} + \sigma_{CG}^2 \mathbf{Q}_{CG}, \quad (3.79)$$

where \mathbf{Q}_{SG} , \mathbf{Q}_{TG} and \mathbf{Q}_{CG} are the corresponding cofactor matrices derived from the geoid errors for the Canadian Gravimetric Geoid 2005 (CGG05) model.

With appropriate expansions of the Eq. (3.70) - (3.78), the VCE algorithm has been applied in the same way as described above for this modified dataset with N separated in three different components. In order to achieve a more realistic estimate of the geoid model error, *OrthoPlus* provides the user with three individual variance components for the geoid height additionally to those of the ellipsoidal and orthometric height. The convergence criterion to stop the IMINQUE algorithm is then defined by:

$$\hat{\boldsymbol{\theta}}^\alpha = \begin{bmatrix} \hat{\theta}_h^\alpha \\ \hat{\theta}_H^\alpha \\ \hat{\theta}_{SG}^\alpha \\ \hat{\theta}_{TG}^\alpha \\ \hat{\theta}_{CG}^\alpha \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \quad (3.80)$$

3.4.2 Iterative almost unbiased estimation (IAUE)

The other type of VCE algorithm implemented in *OrthoPlus* is the IAUE which is the acronym for iterative almost unbiased estimation. The solution of the IAUE also guarantees invariance with respect to translation of the unknown parameters, but the results are only ‘almost’ unbiased.

The IAUE compensates for the main drawback of the IMINQUE algorithm which is the possibility of getting negative variance values from the estimation due to either low redundancy or an incorrect stochastic model (Koch, 1999). The compensation is, however, mainly based on a constraint which means that a convergent and computationally simpler solution will be available for sure, but any problems inherent in the dataset or the modelling might not be recognized (Fotopoulos, 2003).

Since this algorithm only works for semi-positive definite cofactor matrices \mathbf{Q}_k , *OrthoPlus* checks if the corresponding eigenvalues are positive. The IAUE is implemented by computing the current iteratively-derived variance estimates $\hat{\theta}_k^\alpha$ as a function of the previous ones $\hat{\theta}_k^{\alpha-1}$ (Fotopoulos, 2003):

$$\hat{\theta}_k^\alpha = \frac{\hat{\theta}_k^{\alpha-1} \mathbf{1}^T \mathbf{R} \mathbf{Q}_k \mathbf{R}^T \mathbf{1}}{\text{tr}(\mathbf{R} \mathbf{Q}_k)}, \quad k \in [h, H, N] \text{ or } k \in [h, H, N_{SG}, N_{TG}, N_{CG}], \quad (3.81)$$

where all parameters have been already explained in section 3.4.1. Other than that, all steps for the VCE procedure can be adopted from the IMINQUE algorithm. Apart from the variance components for the ellipsoidal and orthometric height which are estimated always, the user can choose if only one general variance component for the geoid height is estimated or if the geoid height error is separated in three components (satellite gravity model, terrestrial gravity data, combined gravity model) with different variances, mentioned above.

4 Discussion of results

The objective of the next sections is to demonstrate and test the program *OrthoPlus* with the most current geoid model, GPS and levelling data in Canada. The most important results will be shown and analyzed.

4.1 GPS-levelling dataset

The GPS-levelling test network consists of 430 co-located benchmark points (Figure 4.1) distributed throughout Canada in an area with a latitude from 43.8°N to 62.6°N and a longitude from 60.8°W to 135.1°W. The height misclosure $h - H - N$ for each of these points is determined by the height triplet of ellipsoidal, orthometric and geoid height.

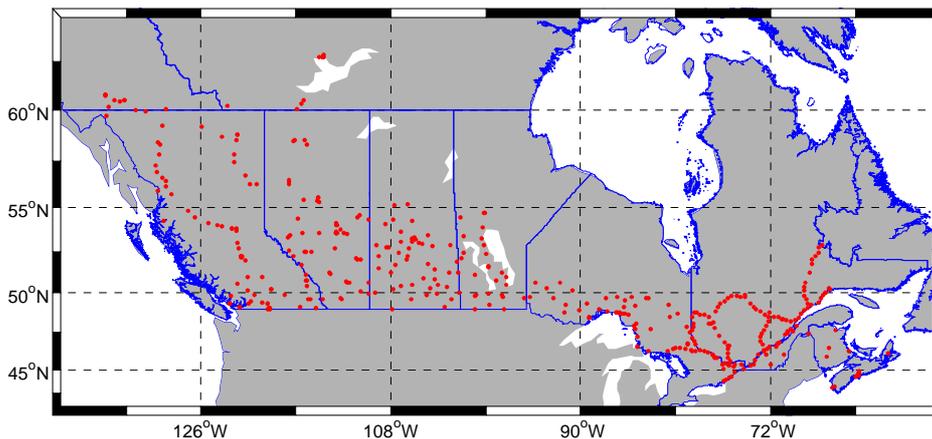


Figure 4.1: GPS-levelling benchmark network of Canada

The ellipsoidal heights are given with respect to the reference ellipsoid GRS80 and the reference frame ITRF97. They have been derived from a least-squares adjustment with Canadian GPS surveys after 1994. In order to compute orthometric heights, the gravity values at each benchmark point are

interpolated from local measurements and then corrected for the variable terrain by converting these interpolated gravity values to mean values along the plumb line. Together with geopotential numbers based on geodetic levelling observations after 1981, the orthometric heights are finally constrained to the single fixed benchmark located along the St-Lawrence River in Rimouski, Québec which is the same as for the NAVD88 (North American Vertical Datum of 1988). Finally, the latest high-resolution geoid model for North America, called CGG05, has been used to compute the geoid height, the third height component given for each benchmark point. The CGG05 has a geographical grid spacing of 2 minutes of arc and is also given in the ITRF reference frame with respect to the GRS80 ellipsoid.

Without introducing a parametric model, the average value of the height misclosure is approximately -40 cm ranging from a minimum of -66.8 cm to a maximum of -12.0 cm with an overall RMS of 40.9 cm and a standard deviation of 10.3 cm. The negative values indicate that the zero-height point of the levelling network is lower than the CGG05 geoid model.

The fully-populated cofactor matrices for every height type are available. Further information will be given within the scope of the analysis of the results from the variance component estimation (section 4.4).

4.2 Evaluation of blunder detection

In order to detect gross errors of large magnitude prior to any network adjustment in the Canadian dataset, the procedure of pre-adjustment data screening has been executed. In this dataset, no blunders were detected by a routine 3-RMS test with a confidence level of 97.3%. Even a 2-RMS (95.5%) test proved that the absolute values of all observations lie within a limit of only two times the RMS value of all observations with a probability of 95.5%. Thus, the pre-adjustment data screening does not find any gross errors which is not surprising since this national a-priori dataset is supposed to contain very good observations. Nevertheless, the post-adjustment data screening is still necessary in order to test the residuals after the least-squares network adjustment for any gross errors of small (or marginal) magnitude or any model errors. If the covariance matrices of the observations are also known, the global test additionally checks these observation precisions for their sufficiency. All three variations of the global test, the two-tail and one-tail upper bound χ^2 -test as well as the F-test, have different test statistics and thresholds, but show the same final results. In all cases, the null hypothesis is rejected and the repetition of the test for each height component shows that not a special but all three height types are responsible for the rejection. Therefore, the first alternative hypothesis was tested and discovers that the a-priori observation precision has not been weighted properly. Repeating the global test with a new covariance matrix, the null hypothesis passed. Thus, in order to make sure that gross errors with a small

magnitude also can be detected, and to check every single observation for an outlier, the procedure of data snooping was carried out.

In the following Table 4.1, the parts in black show the number of outliers in percent detected by a combination of global F-test and data snooping in the case the a-priori covariance matrix is known. If not, the τ -test has been applied and the corresponding results are shown in blue. As significance level, a value of 5% was chosen. The results are based on a run of *OrthoPlus* without using the variance component estimation.

Table 4.1: Comparison between global test & data snooping and τ -test

model	number of outliers in %		outlier with maximum test statistic			
			P ₁	P ₂	P ₃	
A. 1 st order polynomial	2.3	(1.6)	(✓)			✓ Global test and data snooping
B. classic 4-parameter	2.1	(1.9)	(✓)			
C. classic 5-parameter	2.3	(1.9)	(✓)			
D. 2 nd order polynomial	1.9	(1.6)	(✓)	✓		(✓) τ -test
E. differential similarity	1.9	(1.6)	(✓)	✓		
F. 3 rd order polynomial	1.6	(1.9)		✓	(✓)	
G. 4 th order polynomial	1.6	(2.6)		✓	(✓)	

Apart from the third and fourth-order polynomial, the τ -test detects less outliers with either 1.6% or 1.9%, i.e., 7 or 8 outliers out of 430 observations, as opposed to the 1.9 – 2.3% of outliers flagged by the data snooping procedure. For the last two models F and G, it's the other way round since the τ -test finds more outliers. For the 4th order polynomial, the percentage of outliers is even 2.6%. It seems, however, that the low percentage of 1.6% for the data snooping algorithm is too optimistic since higher order models lead to smaller residuals and therefore the test statistic also decreases.

Those benchmark points which lead to an outlier with a maximum test statistic will be flagged as blunders. Within the Canadian dataset, the global and the τ -test indicate blunders in three different points P₁, P₂ and P₃ in a first run of the post-adjustment data screening. The results for all parametric models are summarized in Table 4.1 and the position of the points is shown in Figure 4.2.

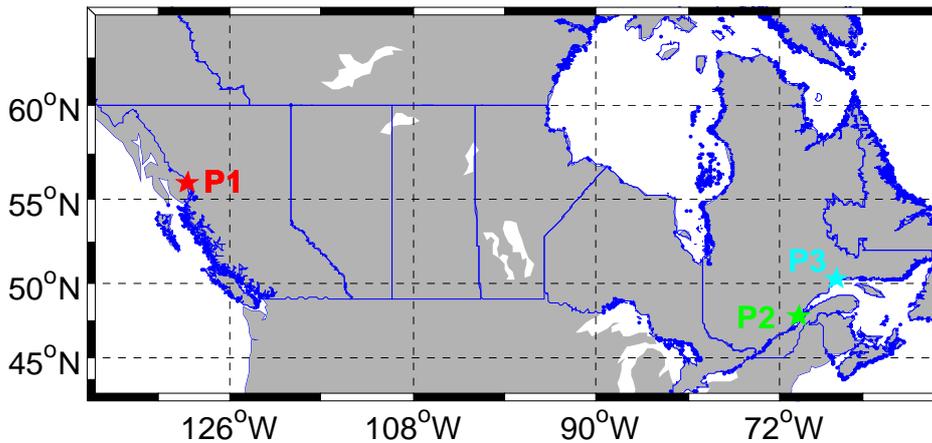


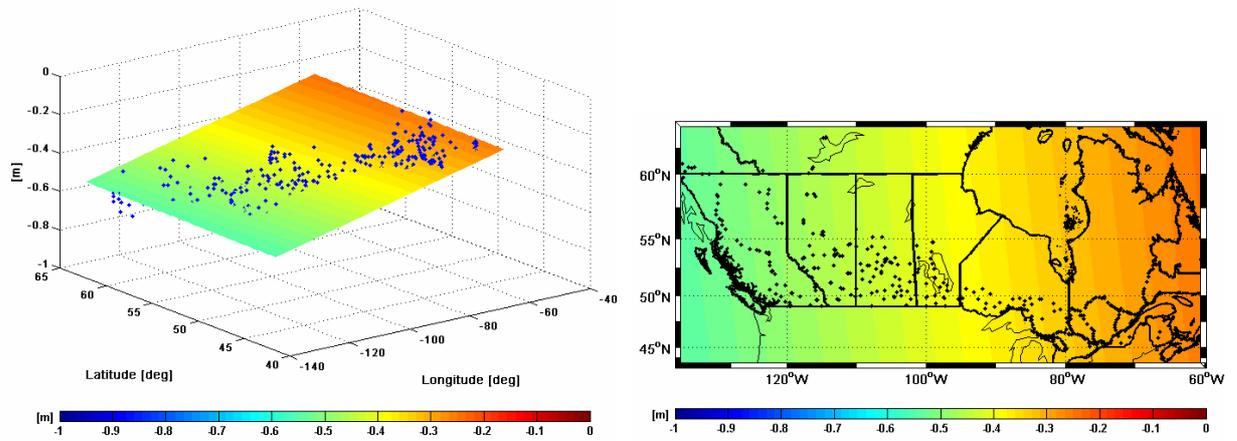
Figure 4.2: Possible blunders detected in the Canadian dataset

With the global test and data snooping, the point P_3 is flagged as a blunder for the lower-order models A – C and point P_2 for the higher-order models D – G. The τ -test, by contrast, interprets a gross error inherent in point P_1 for the models A – E. For the last two models, the τ -test also indicates a blunder in point P_2 . The reason for both tests giving the same results could be that in the higher order models F and G, more blunders are smoothed out and the a-priori covariance matrix has less influence on the test results.

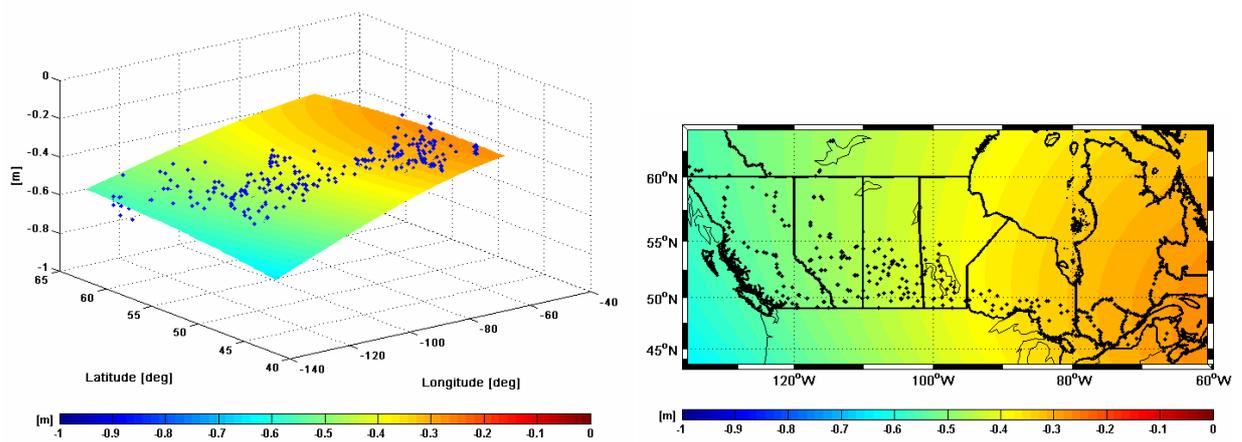
In Figure 4.2, it is obvious that all three points P_1 , P_2 and P_3 are situated in regions next to the coast. The fact that especially these points are flagged could be related to problems with the parametric surface. Since this surface is less supported by benchmark points on the edges, it has more discrepancies in these areas. Another reason why P_1 , P_2 and P_3 are assumed to contain blunders is that observations made on these points are influenced by the relatively difficult geophysical conditions in costal areas.

4.3 Assessment of parametric model performance

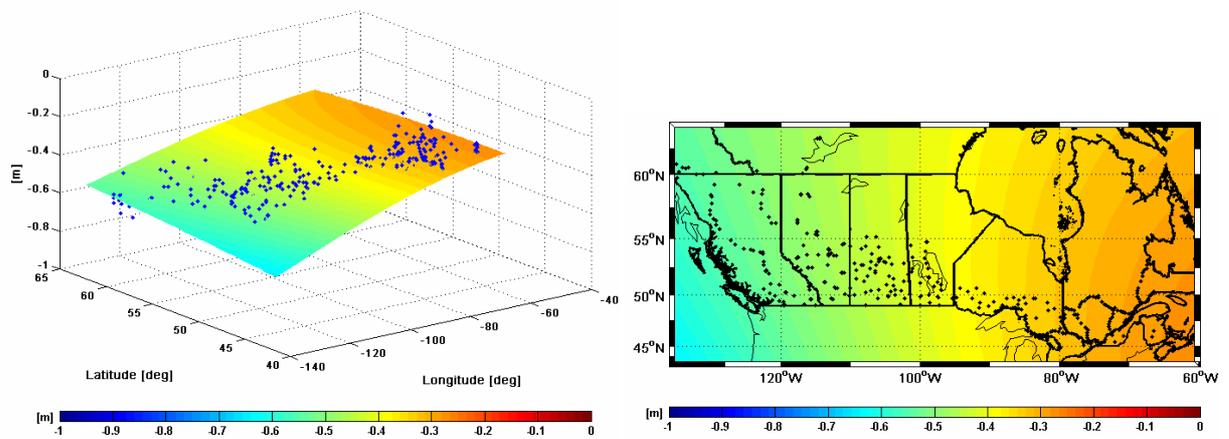
The main objective of *OrthoPlus* is to find a most appropriate parametric surface model for a certain region. Therefore, the program has been executed for every single model. The next Figure 4.3 shows the original models on the left side and the parametric surface fit, i.e., the projection of the model into the Canadian map, on the right side, together with the benchmark points.

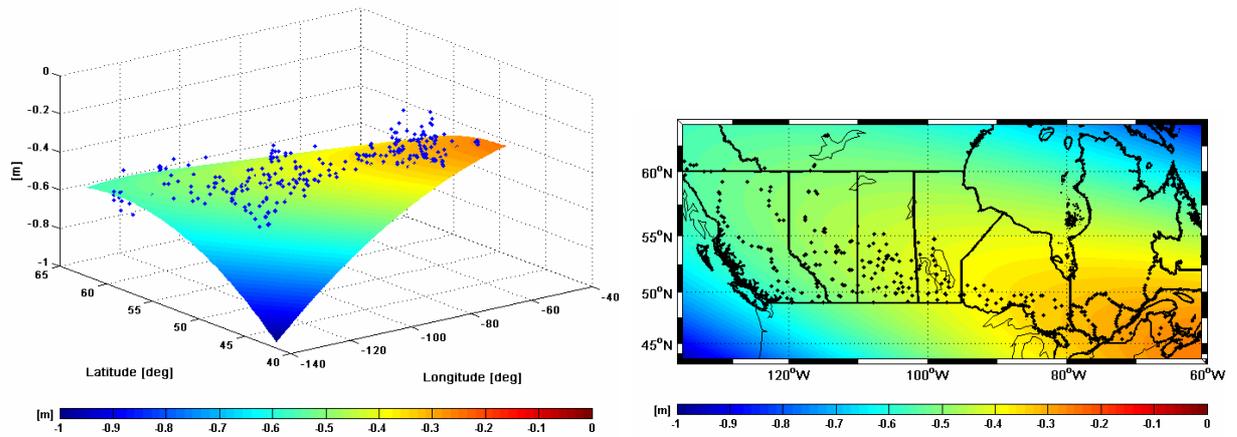
- A. 1st order polynomial

- B. classic 4-parameter

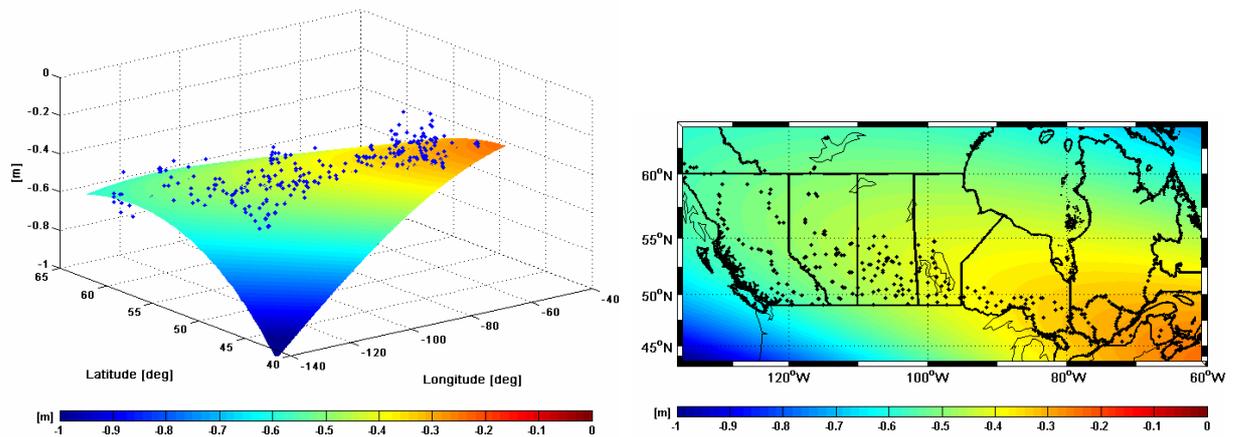
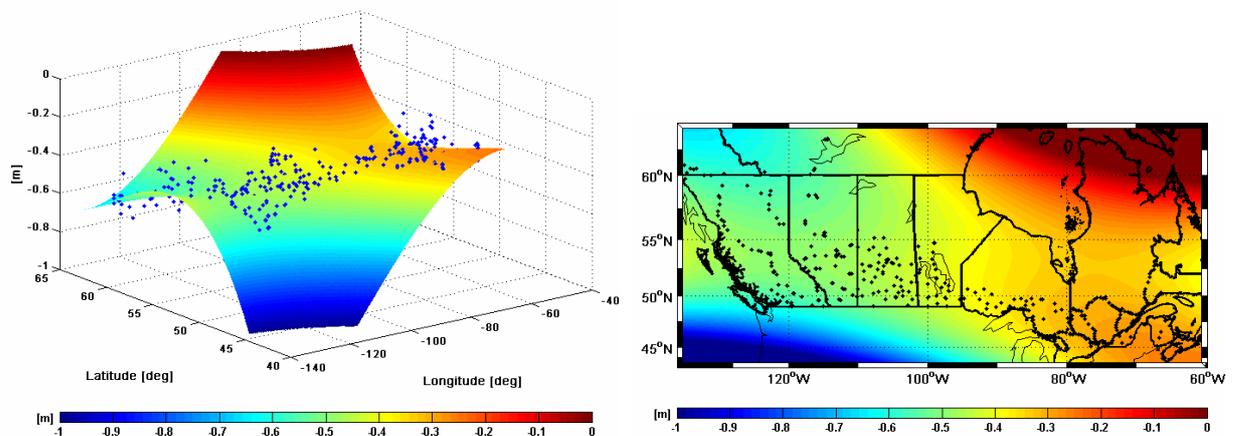


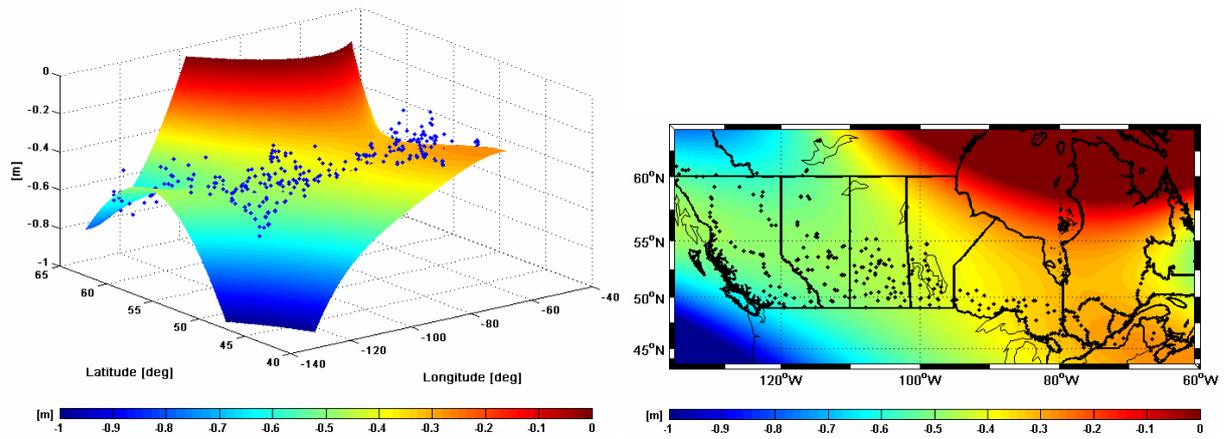
- C. classic 5-parameter



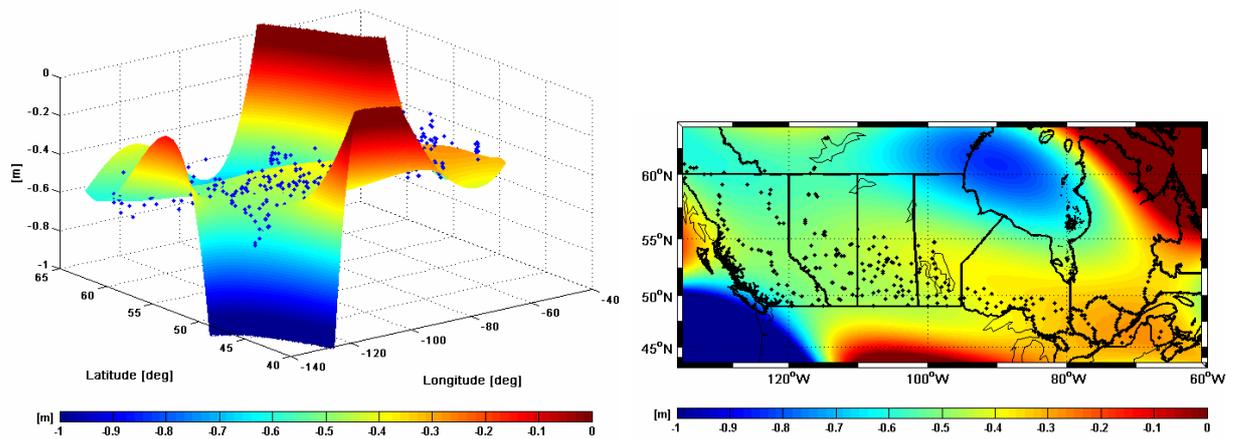
- D. 2nd order polynomial

- E. differential similarity (7-parameter)

- F. 3rd order polynomial

- G. 4th order polynomial

- H. Legendre polynomials



- I. Fourier series

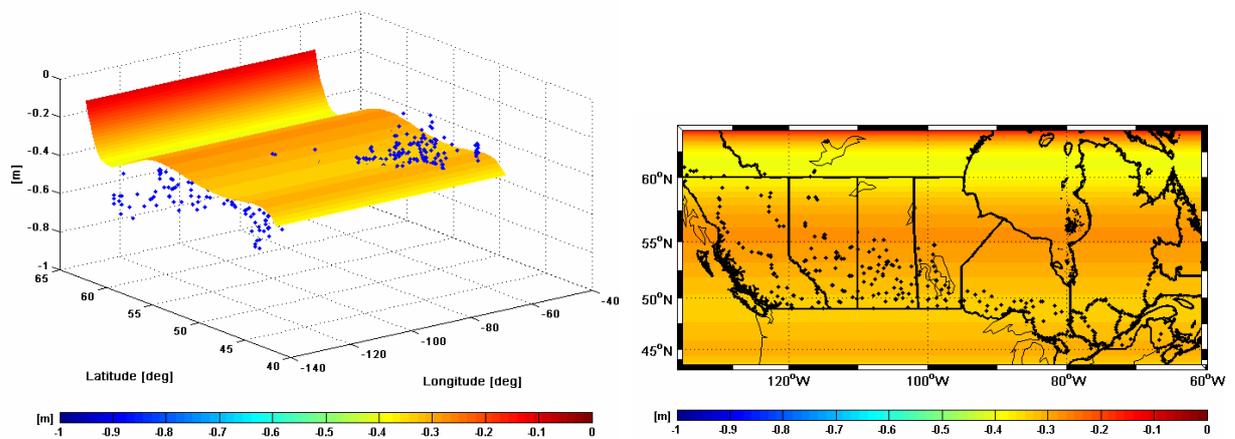


Figure 4.3: left side: models A – I, right side: parametric surface fit of models A – I on grid

At first sight, the user will not be able to find out which model leads to the best parametric model fit. The figure rather tells the user that the orthogonal models, the Legendre polynomials and the Fourier series respectively, in the form as they are chosen in this work are not suitable to be incorporated in the combination of the height triplet h , H and N within the Canadian dataset. The reason is mainly that the selected orthogonal models up to degree and order 4 only provide global support. Consequently, the values of the computed models become up to -14 meters on the south-west edge of the Canadian dataset and the model of the Fourier series partly follows the wave forms. An indication for that problem already appears during the execution of *OrthoPlus*. The normal matrices in the adjustment become singular. Accordingly, the height discrepancies are not adjusted but modelled. For more local areas, such as the Canadian network, the choice of higher degrees and orders can compensate for shorter wavelengths in the network. It should be noted, however, that the incorporation of orthogonal polynomials was only an initial attempt in this work and needs further investigation in the future. That is why all further steps of analyzing the different types of surfaces and the corresponding variance component estimation will only cover the models A – G.

It is obvious that the selection of the appropriate analytical model, i.e., the type and extent, is quite arbitrary. Thus, in order to get some numerical results and assess the performance of the parametric models, it is necessary to apply some hypothetical tests as they are described in section 3.3.4. The numerical results will be analyzed for the Canadian dataset in the following sections.

4.3.1 Comparison between different RMS values

The user of *OrthoPlus* has the possibility to evaluate two different RMS values. One of them is computed by the classical empirical approach and is rather a measure of the precision of the observations. It indicates how well the data sets of the different heights fit each other. The other RMS is computed by the cross-validation procedure and can be interpreted, in fact, as the accuracy with respect to the model's performance in predicting new points (sections 3.3.4.1 and 3.3.4.3). These two RMS values derived after the parametric surface fit with around 5 cm will be compared to each other (Table 4.2 and Figure 4.4). They are certainly at a significantly lower level than the RMS of the misclosure vector before the fit with almost 41 cm. For the Canadian dataset, the classical RMS is only slightly decreasing from 5.82 cm for the first-order polynomial model to 5.28 cm for the third-order polynomial and then again increasing up to 5.36 cm for the fourth-order polynomial. The decrease is expected since more parameters will absorb more differences between the known height misclosure and the estimated "corrector" model and therefore, the lowest RMS usually corresponds to the highest order model. Therefore, the increase between the models F and G, however, is quite unusual. Moreover, the

difference from the minimum to the maximum value is 0.54 cm. In this case, the third order model would be the model with the best precision since it has the smallest set of adjusted residuals.

Table 4.2: RMS values for classical empirical approach and cross-validation procedure

model	number of terms	RMS [cm]	
		classical empirical approach (precision)	cross-validation (accuracy)
A. 1 st order polynomial	3	5.82	5.25
B. classic 4-parameter	4	5.59	4.96
C. classic 5-parameter	5	5.59	4.95
D. 2 nd order polynomial	6	5.38	4.89
E. differential similarity	7	5.33	4.88
F. 3 rd order polynomial	10	5.28	4.84
G. 4 th order polynomial	15	5.36	4.87

The RMS values of the empirical cross-validation are smaller than those of the classical approach for all models with 5.25 cm to 4.87 cm. Here again, the value for the fourth-order polynomial attracts attention since it is slightly higher with 4.87 cm than the 4.84 cm from the third-order polynomial. In this case, the difference between lowest and highest value is only 0.41 cm. That means that this RMS which evaluates the model accuracy at independent control points, is more stable and does not indicate significant differences between the performances of the different models. This can be also seen in Figure 4.4.

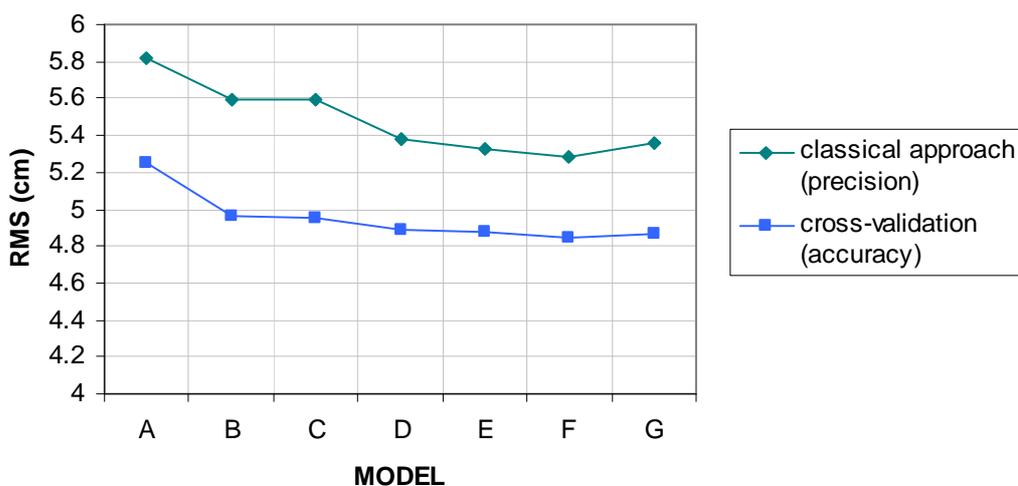


Figure 4.4: RMS values for classical empirical approach and cross-validation procedure

Since these results are not obvious enough to choose a model with the best parametric fit, the problem needs further investigation.

4.3.2 Goodness of fit

Another measure which contributes to the assessment of the model performance is the goodness of parametric model fit for a discrete set of points. It is expressed in terms of the coefficient of determination R^2 or of the adjusted coefficient of determination R_α^2 if the degrees of freedom should be taken into account (section 3.3.4.2). Usually, these statistical measures take values between 0 and 1.

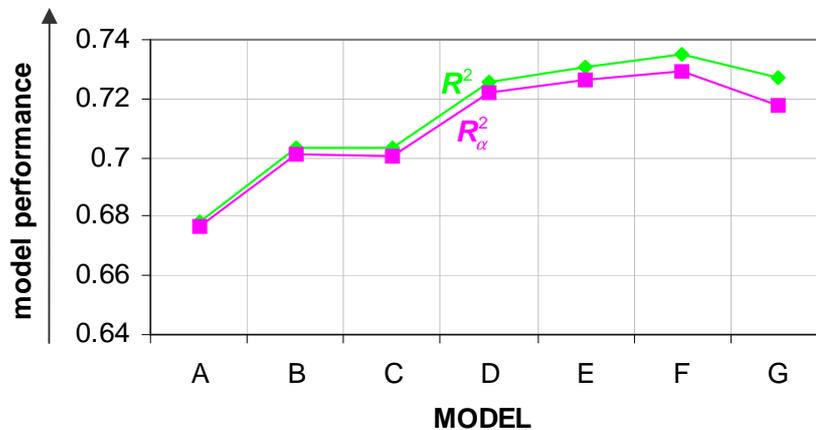


Figure 4.5: Statistical measures of goodness of fit

Figure 4.5 shows the measures of goodness of fit for all models. The gap between the coefficient of determination R^2 and the adjusted coefficient of determination R_α^2 increases with increasing number of parameters since the latter value is corrected for the degrees of freedom. If both measures are approaching the value 1, it means that the parametric fit improves. The difference between the performance of the models is relatively small, for R^2 ranging from 0.68 to 0.73 and for R_α^2 from 0.68 to 0.72. The third-order polynomial model F shows the highest values for both measures and has therefore the best goodness of fit. It should be also noted that the most significant jump is between the classic 5-parameter model C and the second-order polynomial model D which makes the latter model more suitable.

4.3.3 Parameter significance test and model stability

Furthermore, it is also important to make sure that, on the one hand, all parameters of the model are significant and, on the other hand, the model is stable. For the sake of simplicity, computational efficiency and avoidance of over-parameterization, the significance of each model parameter was tested by backward elimination via a hypothesis test, starting with the most extended model. Table 4.3 and Figure 4.6 clearly indicate that the Gram-Schmidt orthonormalization introduced in section 3.3.3 has an influence on the determination of insignificant model parameters.

Table 4.3: Number of insignificant parameters with and without Gram-Schmidt orthonormalization

model	number of terms	number of insignificant terms	
		without Gram-Schmidt	with Gram-Schmidt
A. 1 st order polynomial	3	1	0
B. classic 4-parameter	4	0	1
C. classic 5-parameter	5	0	2
D. 2 nd order polynomial	6	1	1
E. differential similarity	7	0	2
F. 3 rd order polynomial	10	0	4
G. 4 th order polynomial	15	2	7

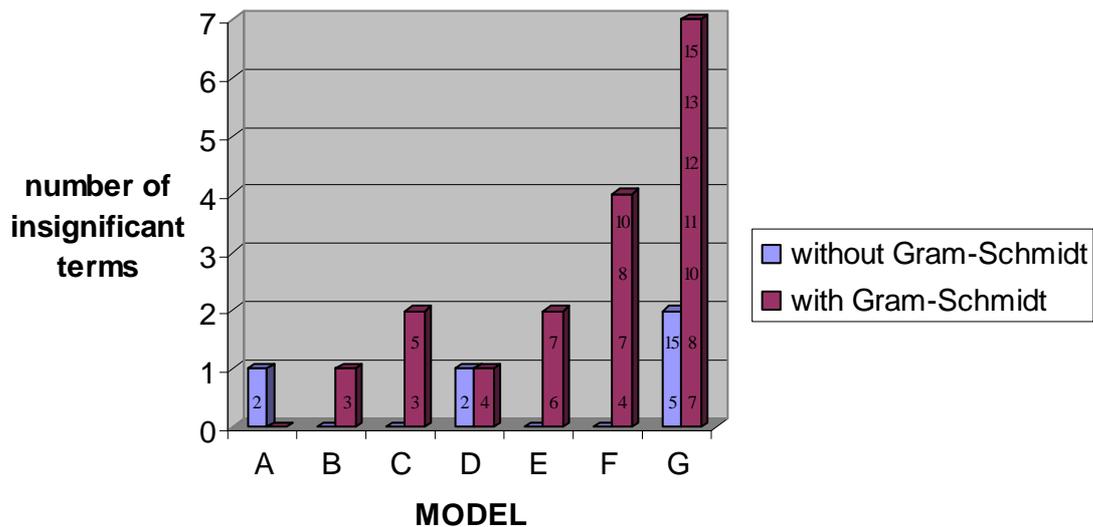


Figure 4.6: Specification of insignificant terms

Whereas all terms of the original selected models A, B, E and F seem to be significant, the parameter significance test shows another result after applying the Gram-Schmidt orthonormalization prior to the combined height adjustment. For the orthonormal form of the classic 5-parameter model, for example, only three of the total five parameters were deemed significant for a 95% confidence level. Compared to the 13 significant parameters of the original fourth-order polynomial, the orthonormalized model only has eight significant parameters, i.e., almost half of the terms are insignificant. Figure 4.6 visualizes the differences between the results of the parameter significant tests by either using the orthonormalization or not. In addition, the figures on the front of the balks of the histogram specify the parameters which are insignificant. This means, for example, that two parameters are insignificant in the classic 5-parameter model C, the third term which is $\cos\varphi\cos\lambda$ and the fifth term which is $\sin^2\varphi$. Figure 4.7 shows that the elimination of the two insignificant parameters hardly changes the appearance of the computed surface fit. The plot of the differences between both surfaces lies on centimeter-level. Therefore, the model without the third and fifth parameter can be used as well.

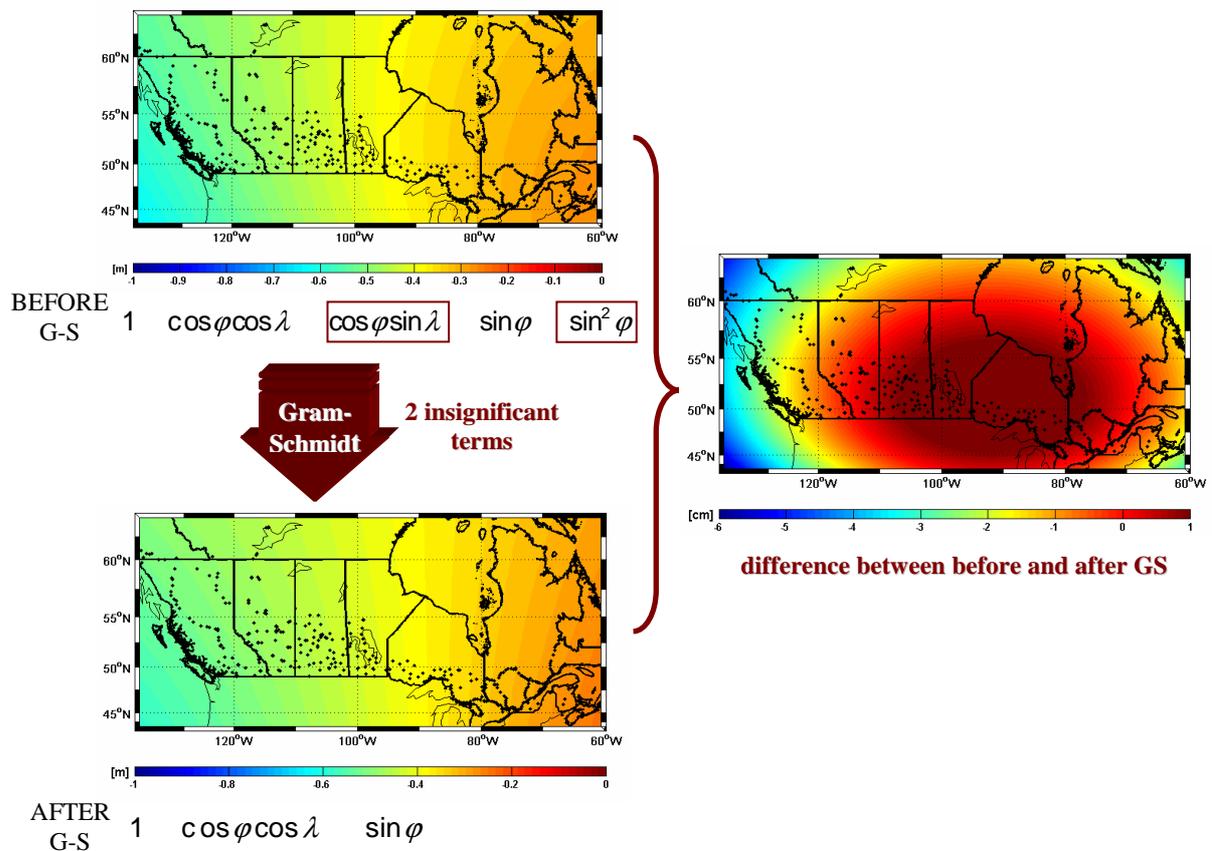


Figure 4.7: Parametric surface fit before and after Gram-Schmidt orthonormalization (classic 5-parameter model)

Another effect of the Gram-Schmidt orthonormalization is that the condition number of the orthonormalized model becomes one. The lower the condition number, the better is the numerical stability of the model. This implies that the orthonormalized models have the best possible numerical stability compared to the original models. Figure 4.8 shows the condition number for every single model A – G with increasing number of terms on a logarithmic scale. The trend of the curve is increasing from a condition number of 499 for the first-order polynomial up to 1.8×10^{10} for the fourth order polynomial. This shows that, in general, the numerical stability decreases with increasing number of terms in the model. It is, however, evident that the classic 5-parameter model with five terms is less stable than the second order polynomial with six terms and the differential similarity model with seven parameters less stable than the third-order polynomial with ten terms. Thus, one might conclude that the polynomial order models are more stable than the similarity-based transformation models.

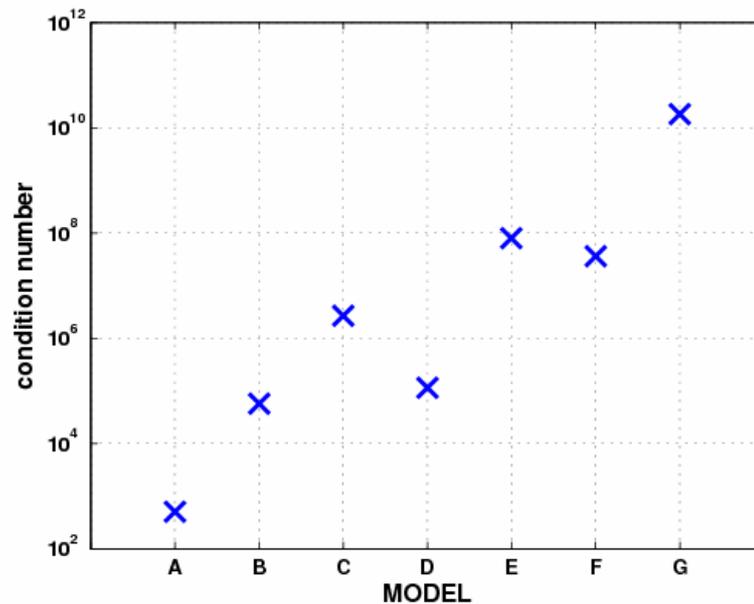


Figure 4.8: Condition number of parametric models

4.3.4 Conclusion of performance assessment

The selection of the appropriate analytical model, i.e., the type and extent, is quite arbitrary, both in choosing the model type and in assessing its performance. This can be verified by the results from the sections 4.3.1 – 4.3.3, which are not obviously indicating a ‘best’ model for the Canadian network. The 5-parameter model C and the second-order polynomial model D, however, seem to have slightly better results than the other models. The decision would probably have to be made between those two models with five and six parameters. The second-order polynomial has slightly lower RMS values with a

precision of 5.38 cm compared to 5.59 cm of the classic 5-parameter model. With respect to the accuracy, however, where the differences between both models lie within a submillimeter-level with nearly identical values of 4.89 cm for model D as opposed to 4.95 cm for model C (Figure 4.4), both models can be assumed to be appropriate. The adjusted coefficient of determination of model D is slightly better with 0.722 compared to 0.700 of model C (Figure 4.5). Furthermore, model C has a higher numerical stability (Figure 4.8). The only disadvantage of the second-order polynomial is that there is always one parameter which is insignificant, with and without Gram-Schmidt orthonormalization. This would likely lead the user to choose the 5-parameter model or rather an orthonormalized model reduced by the two insignificant parameters as shown in Figure 4.7. This uncorrelated model, however, is just created by a re-formulation. As it is only an empirical set of base functions, it does not allow an analytical interpretation (see section 3.3.3).

The reason for the arbitrariness might be that the benchmark points are spread over the whole Canadian network. A suggestion for getting more obvious results is to divide the original dataset of Canada in smaller regions. Results of a regional analysis are shown in section 4.5.

4.4 Results of variance component estimation

The objective of the variance component estimation is to test the statistical model of the combined height adjustment. This section will analyze the various VCE algorithms used in *OrthoPlus* with respect to their iteration numbers as well as the results of the variance components for every height component.

4.4.1 Initial covariance matrices

This section describes the a-priori covariance (CV) matrices for the ellipsoidal, orthometric and geoid heights which were used in the estimation of the variance components for the Canadian dataset. The CV matrices will be interpreted with the help of the following two characteristics:

- the average standard deviation:

$$\sigma_{\text{ave}} = \sqrt{\text{tr}(\mathbf{Q}_k / n)}, \quad k \in [h, H, N], \quad (4.1)$$

with \mathbf{Q}_k representing the cofactor matrices for each height type \mathbf{Q}_h , \mathbf{Q}_H and \mathbf{Q}_N with $n = 430$ which is the number of benchmark points. It has to be mentioned, however, that this average

standard deviation is too optimistic as only the diagonal elements of the matrices are used whereas the correlations are ignored.

- the condition number:

$$\text{cond}(\mathbf{Q}_k) = \frac{\lambda_{\max}}{\lambda_{\min}}, \quad k \in [h, H, N], \quad (4.2)$$

where λ_{\max} and λ_{\min} are the maximum and minimum eigenvalues of the positive definite cofactor matrices \mathbf{Q}_k .

The results of the two characteristics for all three types of covariance matrices are listed in the Table 4.4 below:

Table 4.4: Characteristics of the initial covariance matrices

characteristic	GPS height	orthometric height	geoid height
condition number	3.45×10^5	8.55×10^6	4.24×10^8
average standard deviation (cm)	1.60	5.91	2.48

The initial CV matrix of the GPS height has the lowest average standard deviation with 1.6 cm. It should be kept in mind, however, that this value is an indication of the overoptimistic a-priori accuracy. A visualization of this matrix is shown in Figure 4.9. The plot mainly reveals peaks along the diagonal with relatively small correlations as opposed to Figure 4.10 which depicts the initial fully-populated CV matrix for the orthometric heights. With the average standard deviation of 5.91 cm, the levelling error component seems to have the highest contribution to the overall accuracy of the height misclosure vector $h-H-N$. This fact and the relatively high correlations between the GPS-levelling benchmark points are expected especially in the Canadian dataset. The reason is that the conventional spirit-levelling technique usually suffers from problems if determining orthometric heights over very long distances between the benchmark points in a very large network. Furthermore, the benchmarks are not only distributed inhomogeneous over a very large area, in most parts of Canada the coverage is even quite sparse due to rough landscapes. The average standard deviation of the a-priori fully-populated CV matrix for geoid heights (Figure 4.11) is approximately 2.5 cm. It should be recognized that the

corresponding condition number is relatively high with a value of 4.24×10^8 . That means that the risk of a numerical instability increases when inverting the CV matrix for geoid heights compared to an inversion of the CV matrices for the GPS and orthometric heights.

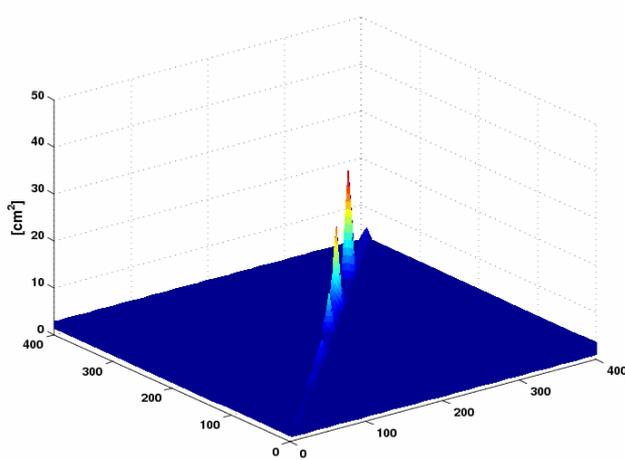


Figure 4.9: Initial covariance matrix for GPS heights

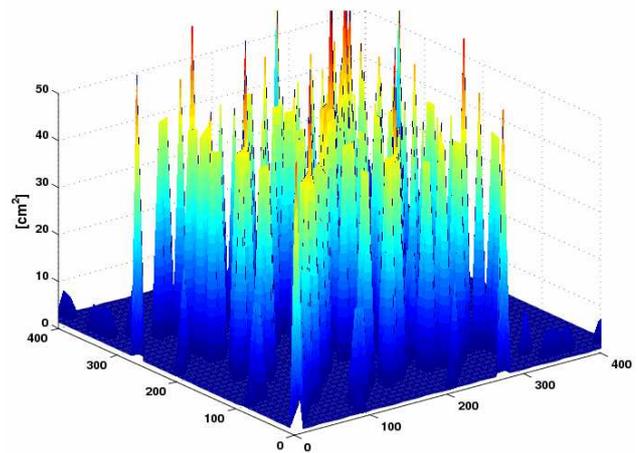


Figure 4.10: Initial covariance matrix for orthometric heights

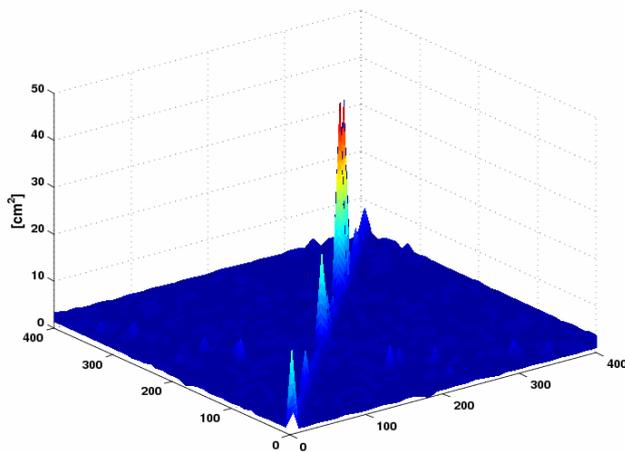


Figure 4.11: Initial covariance matrix for geoid heights

In general, the objective of the implementation of VCE algorithms is the correction of these initial covariance matrices by scaling them with newly estimated variance components for each height type. The following two sections will summarize the results which were attained when applying the IMINQUE and IAUE algorithms (section 3.4) to the Canadian dataset.

4.4.2 IMINQUE

As already expected and mentioned in the description of the iterative minimum norm quadratic unbiased estimation (section 3.4.1), using the IMINQUE algorithm for the Canadian dataset will not lead to any convergent solution. The variance component for the geoid heights is again separated into further components for the satellite gravity model (SG), the combined gravity model (CG) and the terrestrial gravity data (TG). The convergence criterion implies that all variance estimates approach unity. This is not the case for any of the parametric models.

As an example, Figure 4.12 shows the variance estimates $\hat{\theta}_k^\alpha$ from the IMINQUE algorithm for ellipsoidal, orthometric and geoid heights. They are plotted with respect to increasing iteration numbers α for the classic 4-parameter model. The convergence criterion has been chosen with a threshold of 0.05 as follows:

$$\left| \hat{\theta}_k^\alpha - 1 \right| \leq 0.05 \quad \text{for } k \in [h, H, N_{SG}, N_{CG}, N_{TG}]$$

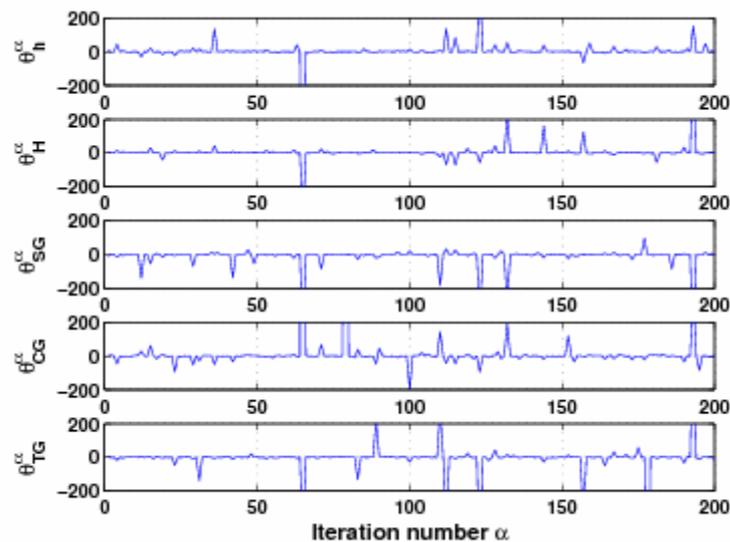


Figure 4.12: Variance estimates of IMINQUE algorithm (classic 4-parameter model)

After 200 iterations, the algorithm has been stopped since no convergence could be achieved. The variance estimates show sudden peaks of large magnitudes, up to almost 5384 for the combined gravity model at the 79th iteration. The reason for the non-convergent behaviour is that the variances do not satisfy the criteria of non-negative variance estimates as it can be seen in Figure 4.12. Even with another threshold, the solution will not be convergent.

Theoretically, the final variance components are computed by multiplying the variance estimates from all iterations (Eq. 3.74). Table 4.5 shows a comparison of the final variance components for the classic 4-parameter model in the third column and for the second-order polynomial model in the fourth column. The algorithm had to be aborted after 200 iterations as no convergent solution could be achieved, whether for one of the two models from this example, nor for any other parametric model. Therefore, it can be assumed that the IMINQUE algorithm will not lead to a reliable result for the computed variance components in the Canadian dataset, i.e., they are not qualified for an improvement of the stochastic model. This fact is also proved by Table 4.5. On the one hand, some of the variance components become negative which conflicts with their definition and on the other hand, the values seem to be totally random with big variations for different parametric models. While, for example, the variance component for the satellite gravity model is 36.4 for the classic 4-parameter model, it is negative with -967.4 for the second-order polynomial model.

Table 4.5: Final variance components for IMINQUE after 200 iterations

height component	final variance components		
		classic 4-parameter model	2 nd order polynomial model
ellipsoidal height	$\hat{\sigma}_h^2$	10.8	4.5
orthometric height	$\hat{\sigma}_H^2$	10.1	17.0
satellite gravity model	$\hat{\sigma}_{SG}^2$	36.4	-967.4
combined gravity model	$\hat{\sigma}_{CG}^2$	21.3	300.8
terrestrial gravity model	$\hat{\sigma}_{TG}^2$	-8.7	-5.3

Negative variance components might be an indicator for low redundancy in the dataset or an incorrect stochastic model and can be at least helpful in order to gain some information regarding the problem set up (Fotopoulos, 2003). But as the IMINQUE algorithm cannot ensure non-negative variance values, it is not sufficient for enhancing initial covariance matrices within the Canadian dataset. This is why this special kind of VCE procedure will not be further investigated in this thesis.

4.4.3 IAUE

In contrast to the IMINQUE algorithm, the iterative almost unbiased estimation, in short IAUE (section 3.4.2), guarantees positive final variance components. Another advantage of this constrained algorithm is that its computations are simpler. This leads to a convergence which is approximately 50% faster than with the rigorous IMINQUE approach. The user of *OrthoPlus* can choose between two options for applying the IAUE algorithm:

- IAUE with N separated in components:

For the sake of direct comparison, the variance component for the geoid height $\hat{\sigma}_N^2$ is set up in the same way as for the IMINQUE algorithm (4.4.2), i.e., it is split up in three further components. These are variance components for the satellite gravity model $\hat{\sigma}_{SG}^2$, the terrestrial gravity data $\hat{\sigma}_{TG}^2$ and the combined gravity model $\hat{\sigma}_{CG}^2$. Section 4.4.3.1 will give a short overview of the results obtained by applying this kind of VCE to the classic 4-parameter model and to the second-order polynomial model.

- IAUE with summarized N :

This type of VCE will provide the user with only one component for the geoid height. The resulting variance components for the ellipsoidal, the orthometric and geoid height will be presented in section 4.4.3.2 and will be applied as scale factors in order to enhance the initial covariance matrices (section 4.4.1). Moreover, a detailed analysis will reveal how the results are affected by using various parametric models as well as diagonal-only instead of fully-populated CV matrices.

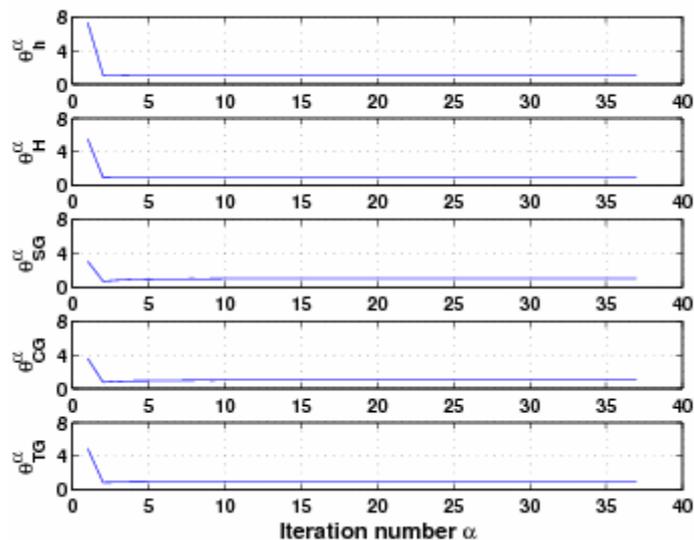
4.4.3.1 IAUE with N separated in components

As the IAUE algorithm constrains the resulting variance components to positive values, it is ensured that this iterative approach will provide a convergent solution as soon as the convergence criterion is fulfilled, i.e., if all variance estimates approach unity (Eq. 3.80). As already explained in section 3.4, this criterion can be realized by testing if the difference between the variance estimates from two successive iterations is smaller than a certain threshold. For the analysis, this threshold was chosen as 0.0001 (Eq. 3.73). With higher values for the threshold, the convergence is reached even faster and with less iterations. In this case, however, the estimated final variance components which are computed by multiplication of the variance estimates from all iterations, will be less reliable.

Table 4.6: Final variance components for IAUE with N separated in components

height component	final variance components		
		classic 4-parameter model	2 nd order polynomial model
ellipsoidal height	$\hat{\sigma}_h^2$	9.05	9.06
orthometric height	$\hat{\sigma}_H^2$	5.55	5.50
satellite gravity model	$\hat{\sigma}_{SG}^2$	1.40	1.19
combined gravity model	$\hat{\sigma}_{CG}^2$	1.28	1.28
terrestrial gravity model	$\hat{\sigma}_{TG}^2$	3.31	3.26

Table 4.6 compares the five final variance components for two different models. The convergent solutions are achieved after 37 iterations for the classic 4-parameter model and after 36 iterations for the second-order polynomial model. As opposed to the IMINQUE algorithm, it is evident that the constrained IAUE approach is convergent (Figure 4.13) and thus, provides reliable results. The variance components are positive and are very similar for both parametric models. The final variance components can be interpreted as scales to enhance the initial CV matrices. Therefore and since the variance factor for the ellipsoidal height with a value of about 9 is higher than for any other height component, it is clear that the CV matrix for the ellipsoidal height will be affected the most by the VCE. This fact is not surprising as this initial matrix with an average standard deviation of 1.6 cm seemed to be overoptimistic anyway. The estimated variance factors for the satellite and combined gravity model, on the other hand, are quite close to one which validates that the initial CV matrices \mathbf{Q}_{SG} and \mathbf{Q}_{CG} are reasonable.

**Figure 4.13:** Variance estimates of IAUE algorithm (classic 4-parameter model)

4.4.3.2 IAUE with summarized N

In order to enhance the initial covariance matrices for the ellipsoidal, orthometric and geoid height component, the user of *OrthoPlus* can apply a variant of the IAUE which summarizes $\hat{\sigma}_{SG}^2$, $\hat{\sigma}_{TG}^2$ and $\hat{\sigma}_{CG}^2$ and finally provides only one component $\hat{\sigma}_N^2$ for the geoid height. In section 4.3., the performance of seven different parametric models was assessed. In this context, it is worth considering if the implementation of these various models will have any influence on the results of the VCE. Following Table 4.7 gives an overview of the three different final variance components $\hat{\sigma}_h^2$, $\hat{\sigma}_H^2$ and $\hat{\sigma}_N^2$ resulting from the variance component estimations with the models A – G and their corresponding number of iterations.

One condition for a VCE is that no systematic effects among the different height types are inherent in the dataset. Furthermore, exactly these systematic effects, among others, should be sufficiently eliminated through the use of a proper parametric model within the combined least-squares adjustment. Thus, if one of the final variance components shows a totally different value for one of the parametric models, it could be interpreted as an inappropriateness of this model. In this case, the VCE procedure could indicate unmodelled systematic errors which propagate into the final variance components and provide the user with unreliable results (Koch, 1999).

Table 4.7: Final variance components for IAUE with summarized N for various parametric models

parametric model	number of terms	final variance components			number of iterations
		$\hat{\sigma}_h^2$	$\hat{\sigma}_H^2$	$\hat{\sigma}_N^2$	
A. 1 st order polynomial	3	9.34	5.53	2.44	22
B. classic 4-parameter	4	9.34	5.54	2.43	22
C. classic 5-parameter	5	9.33	5.63	2.42	22
D. 2 nd order polynomial	6	9.37	5.51	2.37	22
E. differential similarity	7	9.37	5.51	2.36	22
F. 3 rd order polynomial	10	9.43	5.38	2.36	22
G. 4 th order polynomial	15	9.48	5.56	2.32	22

From Table 4.7, it is evident that all models obtain very similar results for the final variance components with an average of 9.38 for the ellipsoidal height, 5.52 for the orthometric height and 2.39 for the geoid height. The convergence is always reached after 22 iterations. These results lead to the assumption that all parametric models seem to be appropriate for the Canadian dataset and verify the conclusions made from the performance assessment (section 4.3.4) that the selection of the appropriate

analytical model is quite arbitrary. Just the fourth-order polynomial has a slightly higher value for the ellipsoidal variance component with 9.48 and a minimum value with 2.32 for the geoid height. The scale for the initial orthometric CV matrix is minimal for the third-order polynomial. Therefore, both models are not expected to be the best choice for absorbing systematic effects and datum inconsistencies due to possible over-parameterization. However, the similar results of the various models are not only an indicator for the appropriateness of the parametric model but also show the capability of the IAUE procedure for providing reliable results. This conclusion can be confirmed by comparing the final variance components for both alternative definitions of the IAUE, with either three ($\hat{\sigma}_h^2$, $\hat{\sigma}_H^2$, $\hat{\sigma}_N^2$) or five different components ($\hat{\sigma}_h^2$, $\hat{\sigma}_H^2$, $\hat{\sigma}_{SG}^2$, $\hat{\sigma}_{TG}^2$, $\hat{\sigma}_{CG}^2$). With the classic 4-parameter model, for example, both IAUE approaches provide approximately the same results for the ellipsoidal component with 9.34 (summarized N) and 9.05 (separated N) and for the orthometric component with 5.54 (summarized N) and 5.55 (separated N). The geoid component $\hat{\sigma}_N^2$ is 2.43 for the IAUE algorithm with a summarized N and is divided into three components with $\hat{\sigma}_{SG}^2 = 1.40$, $\hat{\sigma}_{TG}^2 = 3.31$ and $\hat{\sigma}_{CG}^2 = 1.28$ for the IAUE algorithm with N separated in components.

Since all parametric models will lead to approximately the same results and in order to present a clear overview, further investigation of the results from the IAUE procedure will only concentrate on one of the models. For this purpose, the classic 4-parameter model was chosen in order to keep the number of parameters low and avoid over-parameterization.

The variance component estimation was mainly implemented in order to obtain final variance component factors which can then be used for calibrating the initial covariance matrices of the GPS height, the orthometric height and the geoid height. These scaled CV matrices will present estimates of the accuracy which are more reliable. The difference between the original and scaled average standard deviation (Eq. 4.1) before and after the VCE in Table 4.8 represents the change in the corresponding accuracy level. After taking a quick look at Table 4.8, it becomes clear that the original average standard deviations of the initial CV matrices were overoptimistic for all three height types h , H and N since all scaling factors, i.e., the final variance components, are bigger than one. This applies especially for the GPS heights where, on the one hand, the original average standard deviation is minimal with 1.6 cm and, on the other hand, the final variance component is maximal with 9.34. This leads to an up-scaling of the original average standard deviation up to nearly 5 cm. Furthermore, the scaled average standard deviation is a maximum for the orthometric heights and therefore provides the major contribution to the overall error budget with approximately 14 cm. This might be due to difficulties which arise from spirit-levelling and gravity measurements within the inhomogeneous and partly sparse data distribution within the Canadian dataset. Last but not least, and with a minimum for the final

variance component with 2.43, the initial measure of accuracy provided before the VCE with 2.48 cm is most realistic for the geoid heights and results in a scaled average standard deviation of 3.87 cm.

Table 4.8: Final variance components for all three height types and average standard deviations of CV matrices

	GPS height	orthometric height	geoid height
original average standard deviation [cm]	1.60	5.91	2.48
final variance component	9.34	5.54	2.43
scaled average standard deviation [cm]	4.88	13.91	3.87

Practically, it is often the case that fully-populated CV matrices are either not available or simply not used with all correlations for very large datasets. Therefore, another issue to consider within the IAUE algorithm is whether the correlations of the initial CV matrices for the GPS heights \mathbf{Q}_h , for the orthometric heights \mathbf{Q}_H and for the geoid heights \mathbf{Q}_N play a role in the variance component estimation. The effect of neglecting correlations can be tested by only taking the diagonal elements of these initial CV matrices into consideration.

Table 4.9: Effect of correlations on final variance components

covariance matrices	final variance components			number of iterations
	$\hat{\sigma}_h^2$	$\hat{\sigma}_H^2$	$\hat{\sigma}_N^2$	
fully-populated	9.34	5.54	2.43	22
diagonal	3.36	0.01	4.48	105

Table 4.9 compares the final variance components for the fully-populated covariance matrices in the first row and for the diagonal matrices in the second row. Taking a quick look at Figure 4.10, it can be noticed that the orthometric CV matrix contains relatively high correlations between neighbouring benchmark points. In fact, especially the final variance component of the orthometric height in the case of a diagonal cofactor matrix attracts attention. Its value is very small with 0.01, i.e., almost zero. That means that the initial CV matrix \mathbf{Q}_H will get extremely down-scaled. Thus, if the diagonal-only matrix was used and if the correlations were ignored, the user would erroneously think that the given initial CV matrix was too pessimistic. With a value of 3.36, the scale for the diagonal CV matrix of the ellipsoidal height is also much smaller than for the fully-populated CV matrix. As a consequence, the scaled average standard deviation is too small, i.e., overoptimistic, and therefore less reliable. The

geoid height is the only case for which the final variance component with the diagonal matrix is larger than with the original CV matrix. In conclusion, neglecting significant correlations of initial CV matrices will mostly result in overoptimistic accuracy values.

Another aspect of using diagonal-only CV matrices is that more iterations are needed for a convergence of the IAUE algorithm, respectively 105 compared to 22 for the fully-populated CV matrices. This can be validated by Figure 4.14 and Figure 4.15. The figures which reveal the convergence behaviour of the three different variance components additionally lead to the assumption that each factor converges at different rates. For the implementation of fully-populated CV matrices, the variance factor estimate for the ellipsoidal heights converges the fastest with 16 iterations, followed by the orthometric heights with 21 iterations and finally leads to an end of the iterative VCE process after one more iteration for the geoid heights. As already mentioned, the variant with the diagonal CV matrix will take more iterations. While the geoid height factor levels out after 28 iterations and the ellipsoidal height factor after 49 iterations, the orthometric height is the restraining factor for the entire approach which converges not until 105 iterations. This was expected since the difference between the diagonal and the ‘true’ CV matrix is even bigger than between the fully-populated and the ‘true’ CV matrix. That is why approaching unity will take even longer with diagonal-only CV matrices.

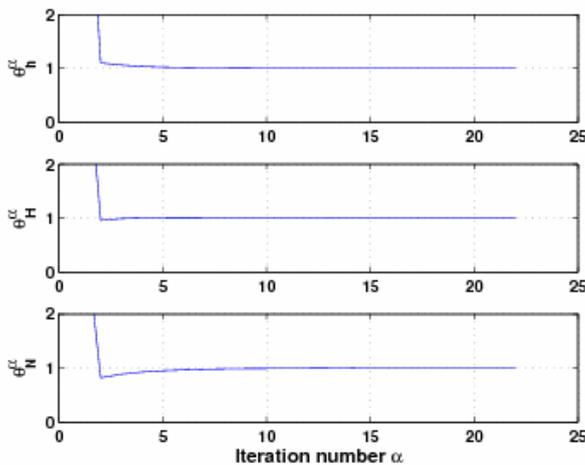


Figure 4.14: Variance estimates and iterations (fully populated CV matrices)

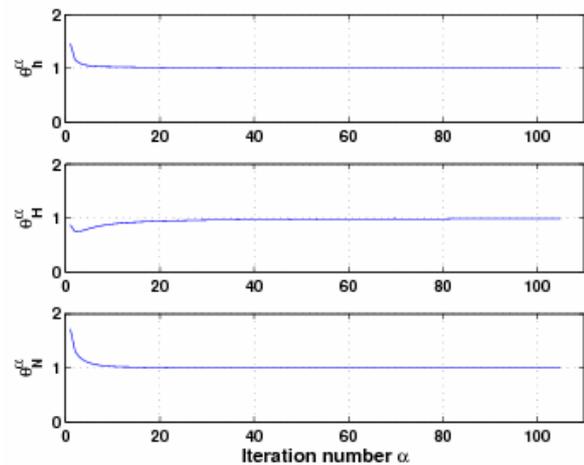


Figure 4.15: Variance estimates and iterations (diagonal CV matrices)

As a conclusion, the incorporation of the iterative variance component estimation should be definitely understood as a potentially powerful statistical tool in the mixed height network adjustment. In the Canadian dataset, especially the IAUE algorithm can be used for determining and validating an appropriate scale for the initial covariance matrices which were found to be too optimistic in most cases. By re-scaling these matrices, reliable accuracy measures can be obtained which are important for

determining the contribution of the ellipsoidal, the orthometric and the geoid height to the overall error budget.

4.5 Regional analysis

The main purpose of the semi-automated program *OrthoPlus* is to provide significant results of an assessment process for parametric surface models in order to facilitate the user's choice of an appropriate model for a particular region.

However, the results of the assessment within the whole Canadian dataset (section 4.3) still seem to be quite arbitrary. Thus, the user will not be able to come to an obvious decision about which parameter model would be best to apply in this specific region. It is assumed that the arbitrariness is due to the distribution of the benchmark points which are spread over the whole Canadian network. One idea to solve this problem is to divide the original dataset of Canada into smaller regions. In this study, four different regions in different parts of Canada have been chosen with respect to changing geographical characteristics from the West to the East coast. The following Table 4.10 as well as Figure 4.16 show the four regions within their defined interval given in degrees of longitude, the corresponding number of benchmark points within each region and the RMS value of the height misclosure before the fit.

Table 4.10: Basic network characteristics of four different regions in Canada

region	longitude	number of points	RMS [cm]
Western Canada	[135.1°W, 110°W]	148	51.0
Central Canada]110°W, 95°W]	76	40.4
Eastern Canada]95°W, 68°W]	176	31.8
The Maritimes]68°W, 60.8°W]	30	32.5

Eastern Canada contains most of the benchmark points and has the lowest RMS value of all regions at the same time which is an indication for a relatively good distribution of the GPS-benchmark data on a quite smooth surface. The region of Western Canada, by contrast, is known for its rough landscape dominated by the varying slopes within the Rocky Mountains. That is why here, the RMS value reaches a maximum with about 51 cm which is even 10 cm higher than for the entire Canadian dataset. For all other regions, the RMS values are smaller.

By applying a complete performance assessment with the pre-selected models A – G, the sections 4.5.1 – 4.5.4 will demonstrate the process of selecting the most appropriate model for every single region. Section 4.5.5 gives a short comparison between these different appropriate models and

compares the results to those of the entire Canadian dataset. Since initial covariance matrices have not been available for these four chosen regions, identity matrices were used instead. That is why variance component estimations could not be implemented in this case.

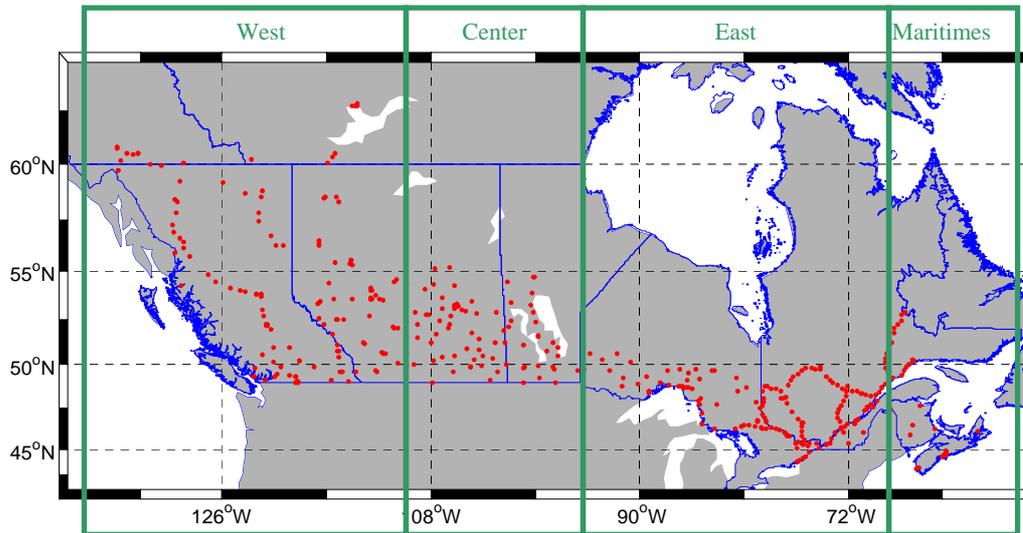


Figure 4.16: Division of Canadian dataset into four different regions

4.5.1 Results for the test network of Western Canada

The distribution of benchmark points within Western Canada is rather sparse and inhomogeneous due to a rough landscape.

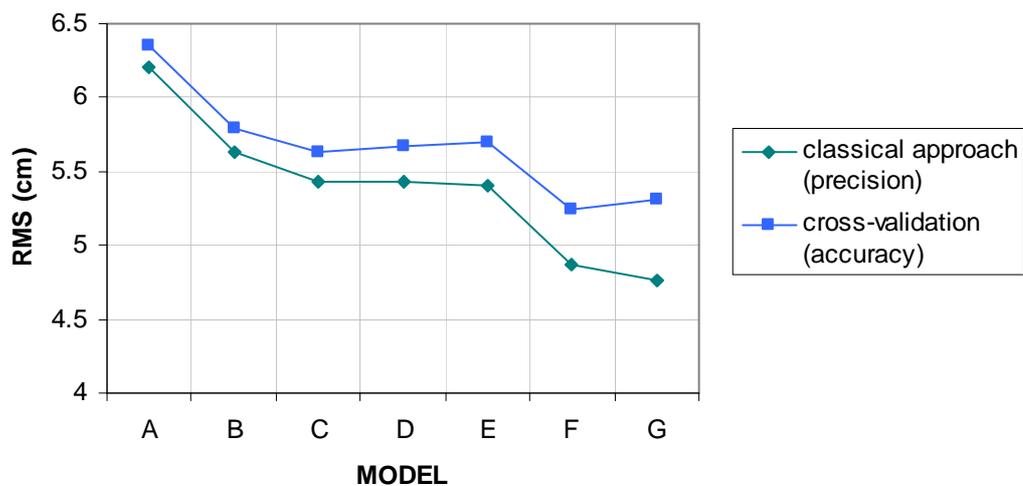


Figure 4.17: Values for precision and accuracy within Western Canada

By fitting a parametric surface model, the RMS of the misclosure vector is almost reduced to one tenth of the value before the fit, i.e., from 51 cm to about 6.2 cm down to 4.8 cm depending on the implied model. These RMS values of the classical approach are, however, too optimistic as they are only a measure of precision. The RMS values of the cross-validation procedure are a little bit higher ranging from about 6.4 cm to 5.2 cm but are a more realistic measure of the actual accuracy of the height misclosure after the fit (section 3.3.4.3). The two different RMS measures for Western Canada are plotted in Figure 4.17. As model F, i.e., the third-order polynomial model, has the smallest accuracy value with 5.2 cm, it could be interpreted as an appropriate model at first sight.

This assumption can be confirmed by Figure 4.18 which is a plot of the goodness of fit for Western Canada. For all models, the values are relatively small with values beyond 0.5 and even smaller than 0.06 for the first-order polynomial model. This reveals that the parametric surface fit is not very sufficient for any of the models A – G. Nevertheless, there is a significant jump between the differential similarity model E with an adjusted coefficient of determination R_α^2 of 0.25 and the third-order polynomial model F with a value of 0.38. This makes the latter one the model with the better parametric fit. Even if the fourth-order polynomial model G provides the highest measures of goodness of fit with a slightly higher value for R_α^2 of 0.39, the third-order polynomial model would be still identified as being suitable for the Western Canada region in combination with the results from the cross-validation (Figure 4.17).

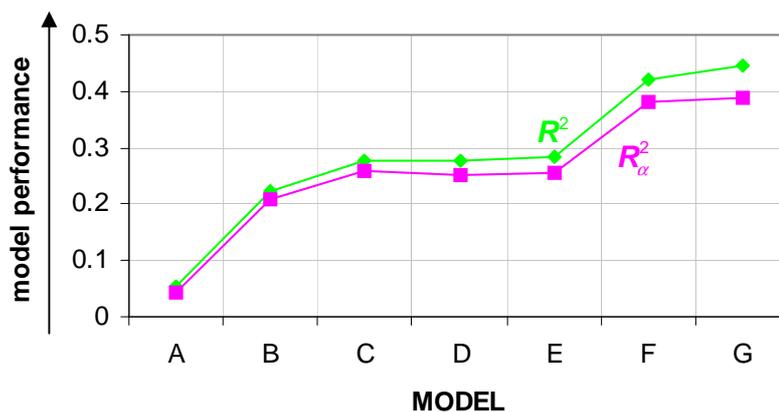


Figure 4.18: Statistical measures of goodness of fit for Western Canada

The only reason why one would rather choose a model with less parameters is the risk of over-parameterization and the improvement of the numerical stability. In fact, the condition number provides a new insight into the overall performance of model F. As its condition number is fairly high with a value of 2.9×10^8 , the user could also tend to select a polynomial model of lower order with a higher

stability. As a last step of the assessment process, the significance of the parameters used in the third-order polynomial model is tested which reveals that two parameters are insignificant. Although in the differential similarity model all parameters would pass the test and thus seem to be significant, the test indicates two insignificant parameters after decorrelation of the model by applying the Gram-Schmidt orthonormalization.

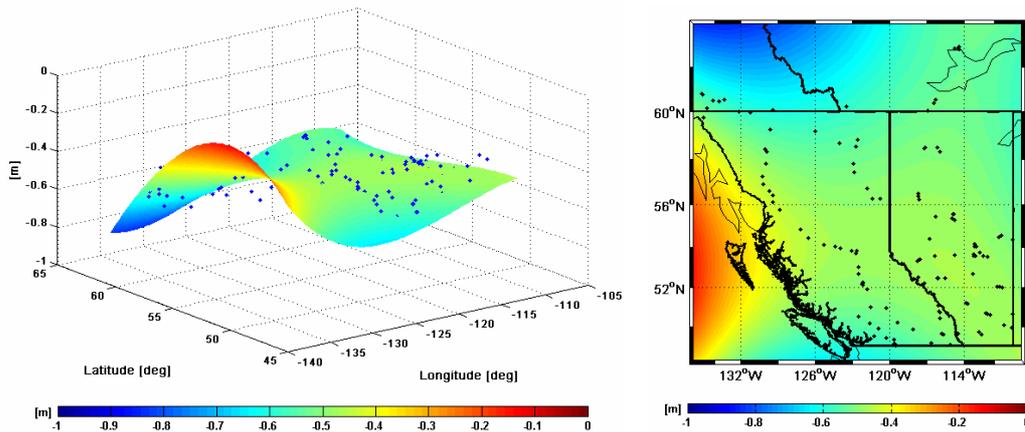


Figure 4.19: Third-order polynomial model and its parametric surface fit to Western Canada

As a consequence, the choice of the ‘best’ model is not obvious but rather depends on the user’s interests. That is why the user has to find a trade-off between the different results from the several statistical tests. The Figure 4.19 shows the third-order polynomial model and its fit to the Western Canadian dataset.

4.5.2 Results for the test network of Central Canada

Within the dataset of Central Canada, the selection of an appropriate parametric surface model will be more obvious. That is mainly because the benchmark points are more consistently distributed with rather small variations between the points due to a smoother landscape than in Western Canada. From Figure 4.20, it can be concluded that the models B and E have the lowest values for the accuracy with 4.5 cm for the first one and 4.4 cm for the latter one. Since the stability of model B is much better with a condition number of 1.5×10^7 compared to model E with a value of 1.3×10^{11} , this classic four-parameter model (model B) will be kept in mind as possible appropriate model for the following tests.

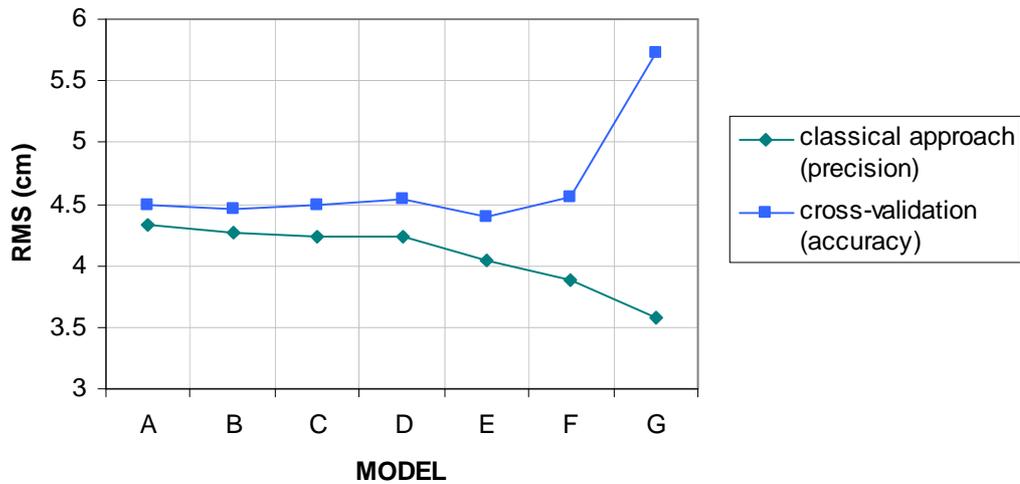


Figure 4.20: Values for precision and accuracy within Central Canada

The statistical measures of goodness of fit which are shown in Figure 4.21 are reaching a maximum for model G. The condition number of this model, however, is very high with a value of 2.9×10^{12} . At the same time, the parameter significance test reveals four insignificant parameters and even nine insignificant parameters after the Gram-Schmidt orthonormalization. Although the RMS for the classical approach gets minimal with 3.6 cm, this value only shows the internal precision of the model fit and is neglected by the cross-validation procedure which leads to a maximal value for the accuracy with 5.7 cm. If we take a closer look at Figure 4.21 with respect to the model B, it can be seen that the corresponding adjusted coefficient of determination R^2_{α} reaches a 'local maximum' with 0.22. Thus, the classical four-parameter model also has the best parametric fit within the lower order models A – D.

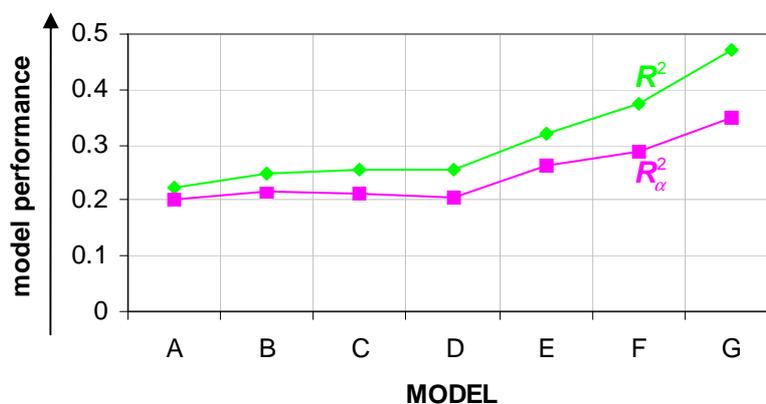


Figure 4.21: Statistical measures of goodness of fit for Central Canada

Finally, the parameter significance test was carried out to determine if any of the model parameters are insignificant. However, it became evident that this procedure was not necessary as the selected model is of low order and consists of only four terms.

Based on these collective results, it was deemed appropriate to make the final decision for the classic four-parameter model (model B) within the Central Canadian network (see Figure 4.22).

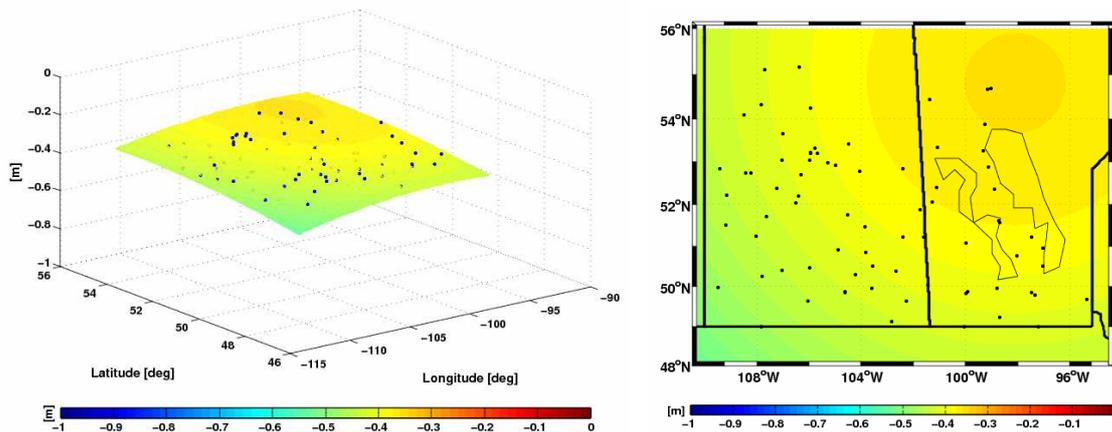


Figure 4.22: Classic four-parameter model and its parametric surface fit to Central Canada

4.5.3 Results for the test network of Eastern Canada

With respect to the network of Eastern Canada, the user of *OrthoPlus* is likely to choose the classic four-parameter model (model B) as it was already done for Central Canada, because of the following reasons:

- The overall RMS of the empirical cross-validation process at independent control points has the same minimal value with 4.9 cm as the first-order polynomial model and is thus, lower as for any other model. Therefore, the application of model B results into an acceptable accuracy for the fit (Figure 4.23).
- Even if the statistical measures of the goodness of fit are not maximal for model B, it can still be considered as appropriate model because there is only a minor difference between the performances of the different models with a marginal variability in the more indicative adjusted coefficient of determination R_{α}^2 (Figure 4.24). The values lie between 0.10 for model A and 0.14 for model F. The value for model B is 0.11.
- The condition number is 2.3×10^6 and thus slightly higher than for the models A and D which show the best numerical stability of all implied models.

- However, model B (Figure 4.25) is still the most appropriate model as all model parameters have been tested to be significant as opposed to the model parameters of the models A and D.

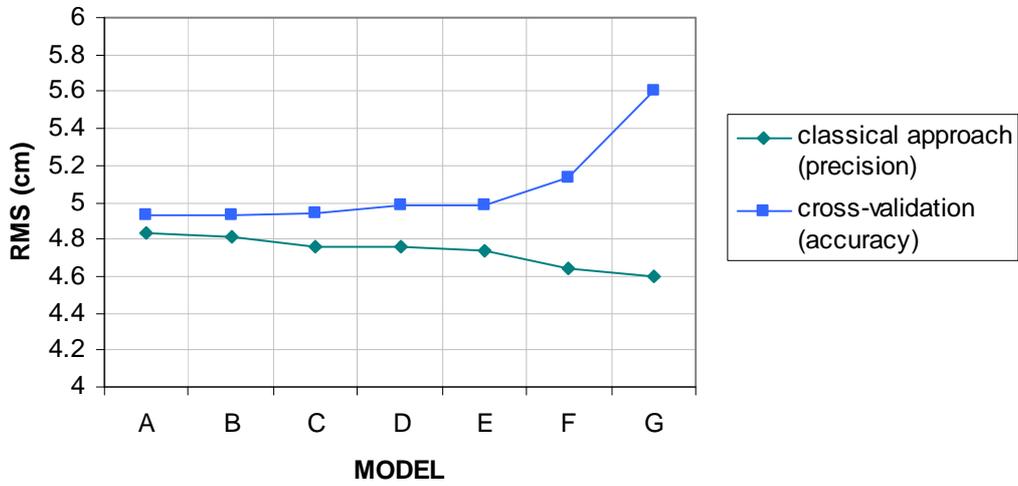


Figure 4.23: Values for precision and accuracy within Eastern Canada

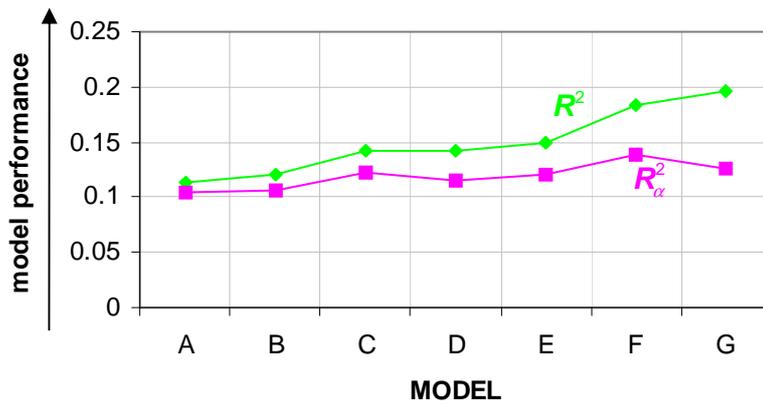


Figure 4.24: Statistical measures of goodness of fit for Eastern Canada

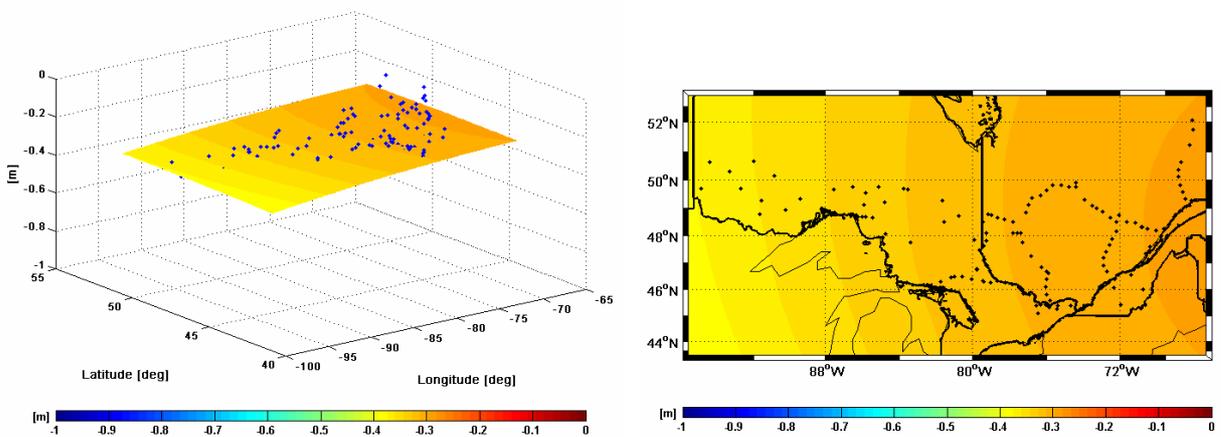


Figure 4.25: Classic four-parameter model and its parametric surface fit to Eastern Canada

4.5.4 Results for the test network of the Maritimes

According to Figure 4.26, the lowest RMS value with 4.9 cm for the cross-validation can be achieved if the first-order polynomial model (model A) is implemented within the height adjustment for the Maritimes. This is the same accuracy which was obtained with this type of model in Eastern Canada. The classic four-parameter model (model B) and the differential similarity model (model E) only result in slightly higher RMS values with 5.1 cm for the first one and 5.6 cm for the latter one. While the RMS values computed with the classical approach are decreasing with an increasing number of model parameters, the RMS values computed with the cross-validation approach are getting worse especially for the third and fourth-order polynomial model (model F and G). These visible effects of over-parameterization fairly show the difference between the precision and the accuracy for a small number of benchmark points with a poor distribution as it is the case in the Maritimes dataset. The implementation of model F already provides a low accuracy with an RMS of 13.2 cm. For model G, the RMS grows up to almost 56 cm. That is why it is not shown in Figure 4.26. It is even an inferior result to not applying any parametric model compared to the RMS value prior to the fit which was 32.5 cm. Not only the condition number with 3.8×10^{16} but also the warning of Matlab that the design matrix is badly scaled give proof for the inappropriateness of the fourth-order polynomial model. Therefore, the models F and G will not be taken into consideration for the further model performance assessment within this dataset.

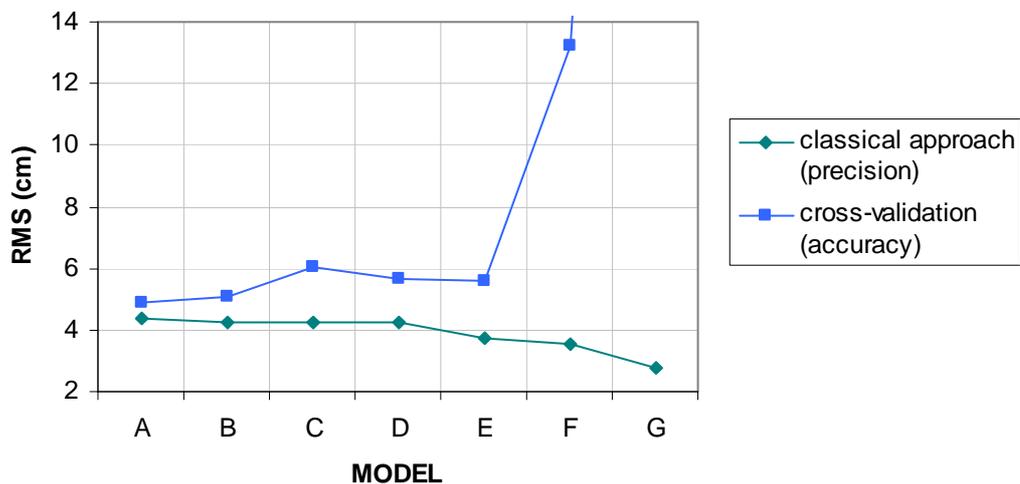


Figure 4.26: Values for precision and accuracy within the Maritimes

The measures of goodness of fit, R^2 and R_α^2 , depicted in Figure 4.27 clearly indicate the differential similarity model (model E) as a better choice than the models A – D with a value for the adjusted

coefficient of determination R_α^2 of 0.57. The classic four-parameter model is the second best model with a slightly lower value of 0.51.

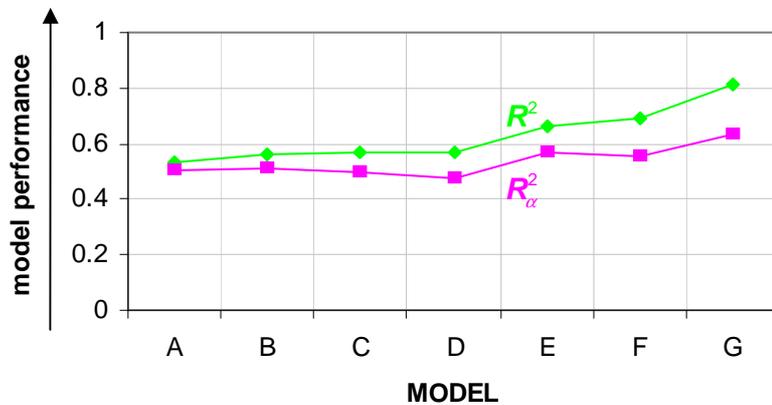


Figure 4.27: Statistical measures of goodness of fit for the Maritimes

Up to here, the user is likely to select the differential similarity model E as the most suitable trend surface. As this model with seven parameters is the extended form of the classic four-parameter model, the significance of the additional parameters should be tested by the backward elimination. The test shows that all seven parameters are statistically significant. However, there are three out of seven insignificant parameters for a new set of orthonormal base functions which was created by the Gram-Schmidt orthonormalization whereas the classic four-parameter model B has only two insignificant parameters in the orthonormal case and none in the original model. Another reason why one would rather choose the model B is its better numerical stability with a condition number of 4.6×10^6 compared to the value of 2.0×10^{10} for model E. Once more, the classic four-parameter model and its parametric surface fit to the Maritimes are shown in Figure 4.28.

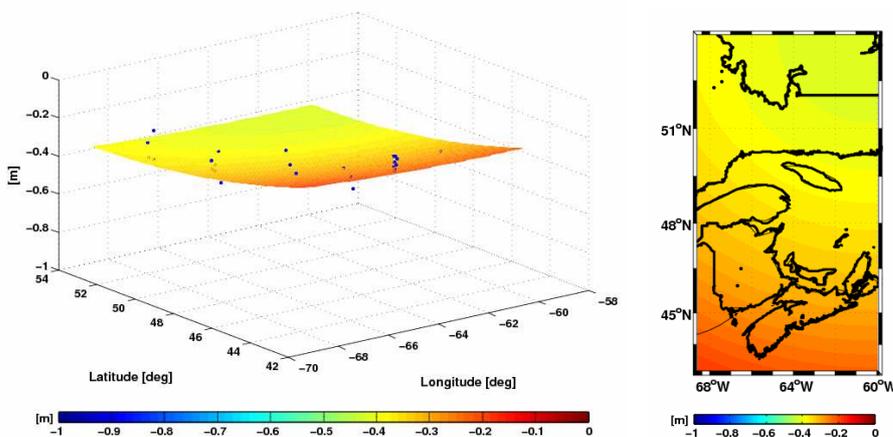


Figure 4.28: Classic four-parameter model and its parametric surface fit to the Maritimes

4.5.5 Comparison and overview

Table 4.11 summarizes the results from the regional analysis for the four different regions of Canada (sections 4.5.1 – 4.5.4) and compares them to the results of the performance assessment (section 4.3.4) of the original Canadian network.

Table 4.11: Comparison of the results from the regional analysis

region	model	RMS before fit [cm]	RMS after fit (accuracy) [cm]	goodness of fit (adjusted)	condition number
Canada	C. classic 5-parameter	40.9	4.95	0.70	2.7×10^6
Western Canada	F. 3 rd order polynomial	51.0	5.24	0.38	2.9×10^8
Central Canada	B. classic 4-parameter	40.4	4.46	0.22	1.5×10^7
Eastern Canada	B. classic 4-parameter	31.8	4.93	0.11	2.3×10^6
The Maritimes	B. classic 4-parameter	32.5	5.08	0.51	4.6×10^6

While the choice of the model type and extension has been relatively arbitrary within the whole Canadian dataset, the results of the regional analysis give a better understanding of which model fits best to the data. However, it becomes evident that *OrthoPlus* only provides a semi-automated tool for comparing parametric surface models and assessing their performance as still some investigation from the user is necessary. The Western Canadian dataset leads to a polynomial model with ten parameters and is therefore the region with the highest number of model parameters compared to the four parameters of the other regions and the five parameters of the Canadian dataset (Table 4.11). This seems to be appropriate as the chosen model should take the relatively inhomogeneous landscape of Western Canada into account which not only contains the Rocky Mountains but also has a maximum RMS value before the fit of 51 cm. In this way, the cross-validation RMS after the fit with a value of 5.24 cm is still higher than from the other regions and Canada but at least, it could be reduced to an acceptable level. A second interesting fact is that for Central Canada the RMS before the fit is almost as high as the one for entire Canada. After the fit with the classic four-parameter model, however, the RMS value within Central Canada gets minimal which is likely to be due to the mostly homogeneous distribution of the GPS benchmark points within this region. Taking a quick look at the results of the Eastern Canadian dataset shows that the RMS value before the fit is minimal compared to the other networks but its relative decrease after the fit is not as high as it is, for example, within Central Canada. Together with the minimal and very low goodness of fit (adjusted coefficient of determination of 0.11), this means that for this case of data distribution the classic four-parameter model might not be the best

choice. Surprisingly, the classic five-parameter model for the Canadian dataset provides the best adjusted measure of goodness of fit with a value of 0.70.

The condition numbers lie between 2.3×10^6 and 2.9×10^8 . Thus, the user should be aware of numerical instability problems. The models for all five networks have been chosen in that way that all parameters pass the significance test. By applying the Gram-Schmidt orthonormalization, it could be the case though that one or another parameter is deemed to be insignificant due to correlations.

In conclusion, it has to be mentioned that *OrthoPlus* is designed to provide the most revealing results from independent empirical tests such as the cross-validation procedure. The user, however, has to keep in mind that the quality of the data will also have an influence on the testing results. Every new dataset will have different characteristics as well as specific types of systematic errors and datum inconsistencies which have to be investigated before using the semi-automated process. This will make it possible to interpret the results from the performance assessment of *OrthoPlus* in an appropriate way and to finally select a suitable parameter model for the specific network.

5 Radial Basis Functions and Genetic Algorithm

So far, this work concentrates on the incorporation of parametric models for an optimal combination of ellipsoidal, orthometric and geoid heights. Different parametric models have been tested and implemented in the semi-automated software *OrthoPlus*. Now, this section introduces an alternative approach: the incorporation of radial basis functions (RBF) which make optimal use of the data distribution. As a refinement, the solution will be optimized by the application of a genetic algorithm (GA).

5.1 Radial basis functions

The radial basis function (RBF) is defined as a weighted sum of translations of a radially symmetric single base function. Richard L. Hardy was one of the first geodesists who applied that approach for the interpolation of surface data (Hardy, 1971). The main advantages of using RBF are the:

- ability to interpolate scattered data, i.e., points which are arbitrarily distributed without particular spacing or density (Franke, 1982)
- simple generalization from one to multi-dimensional spaces as the RBF involve a single independent variable regardless of the dimension of the problem
- flexible adjustment of rough surfaces
- dependence on the distribution of data points.

The most important characteristic of an RBF is that its function values at random points are only dependent on the distance from the origins, i.e., the function is isotropic. In combination with the surface fit, the known benchmark locations will be used as origins in the form of Cartesian coordinates. Therefore, the benchmark points given in geodetic coordinates (φ, λ, h) have to be converted to Cartesian coordinates (x, y, z) for the GRS80 (see Eq. 2.1, 2.2 and `fgeod2cart.m`). In this case, the approximating function, the RBF, will be dependent on the distribution of data points as opposed to the

parametric models which are used in *OrthoPlus*. Furthermore, the implementation of RBF requires using distinct data centers. Within this work, this prerequisite is guaranteed by the implementation of the function `fdistinct_center.m` which provides a number of n_0 distinct data points and a number of n_d deleted points. Only benchmark points which fulfill the following condition, will be included:

$$\rho = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} > 800.0m . \quad (5.1)$$

That means that the distances between all pairs of benchmark points $(\mathbf{x}_i, \mathbf{x}_j)$ will be calculated and if two points are closer than 800m, only one point will be used for the surface fit. The other point will be stored as a deleted point. The decision which one of the two points is kept and which one is deleted is random in this work due to simplicity and demonstration purposes. The reader should, however, keep in mind that the decision must be adapted if one of the two points is an outlier.

The remaining set of n_0 distinct data points $\mathbf{x}_i \in \mathfrak{R}^n$, the so called data centers, and their corresponding scalar values from the height misclosure $f(\mathbf{x}_i)$ are the input values for the basic RBF interpolant which is in general a pure radial sum. It is defined as (Sharifi, 2006):

$$\tilde{f}^{\text{RBF}}(\mathbf{x}) = \sum_{i=1}^{n_0} \alpha_i \varphi_i(\rho) = \sum_{i=1}^{n_0} \alpha_i \varphi_i(\mathbf{x}, \mathbf{x}_i, c^2), \quad (5.2)$$

where α_i are the unknown expansion coefficients with $i = 1, 2, \dots, n_0$. Each of them belongs to one of the data centers. c^2 is the shape parameter of the radial basis φ_i and defines the shape of the base function (Hardy, 1990). $\varphi_i(\rho)$ is the interpolant kernel function which is only dependent on the Euclidian distance from its origin \mathbf{x}_i , the i -th data center respectively:

$$\rho_i = \|\mathbf{x} - \mathbf{x}_i\| = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} \quad i = 1, 2, \dots, n_0 . \quad (5.3)$$

There are different types of kernels which can be used as interpolation kernel. The most common ones which are forming an infinitely smooth RBF are shown in the following Table 5.1 (Sharifi, 2006):

Table 5.1: Examples of more commonly employed RBF kernels

type of kernel	$\varphi(\rho)$
Hardy's Multiquadric (MQ)	$\sqrt{\rho^2 + c^2}, c \geq 0$
Inverse Multiquadric (IMQ)	$\frac{1}{\sqrt{\rho^2 + c^2}}, c > 0$
Inverse Quadratic (IQ)	$\frac{1}{\rho^2 + c^2}, c > 0$
Gaussian	$e^{-c^2\rho^2}, c > 0$

All of these kernel types can be used as an approach to derive a unique solution for the expansion coefficients α_i and all of them will approximately lead to the same results. As the Inverse Multiquadric RBF (IMQ) is more localized in contrast to the other radial basis functions, it is the most suitable type of kernel to be implemented for the surface fit in Canada and its regions and will therefore be used in this work. For further details and applications of the other RBF kernels, the user is referred to Sharifi (2006) and Franke (1982). The basic IMQ interpolant is then defined as:

$$\tilde{f}^{\text{IMQ}}(\mathbf{x}) = \sum_{i=1}^{n_0} \alpha_i \varphi_i(\mathbf{x}, \mathbf{x}_i, c^2) = \sum_{i=1}^{n_0} \alpha_i \frac{1}{\sqrt{\|\mathbf{x} - \mathbf{x}_i\|^2 + c^2}}. \quad (5.4)$$

The computation of an optimal shape parameter c^2 will be explained in the next section and is therefore assumed to be known for now.

Eq. (5.4) can be rewritten as the following symmetric linear system (Sharifi, 2006):

$$\underset{n_0 \times n_0}{\mathbf{A}} \cdot \underset{n_0 \times 1}{\boldsymbol{\alpha}} = \underset{n_0 \times 1}{\mathbf{1}}, \quad (5.5)$$

where $\mathbf{1}$ is the $n_0 \times 1$ observation vector of known height misclosures at the distinct data points and \mathbf{A} the design matrix with entries:

$$a_{ji} = \varphi_j(\mathbf{x}_j, \mathbf{x}_i, c^2) \quad i, j = 1, \dots, n_0. \quad (5.6)$$

The coefficients α_i from the unknown $n_0 \times 1$ vector $\boldsymbol{\alpha}$ can be estimated by solving this system by:

$$\boldsymbol{\alpha} = \mathbf{A}^{-1}\mathbf{1}. \quad (5.7)$$

With these derived coefficients, with the known shape parameter and with the base function, the IMQ interpolant can be estimated at a number of m_0 new random points \mathbf{x}^{new} within the area of the known benchmark points:

$$\hat{\mathbf{1}}_{m_0 \times 1} = \mathbf{A}_{m_0 \times n_0}^{\text{new}} \boldsymbol{\alpha}_{n_0 \times 1} = \sum_{i=1}^{n_0} \alpha_i \varphi_i(\mathbf{x}^{\text{new}}, \mathbf{x}_i, c^2). \quad (5.8)$$

This method will only compensate for short wavelengths as the RBF kernel with the most local support is chosen. In order to also take regional influences, i.e., medium wavelengths, into account, the augmented form of the RBF can be applied. In addition to the basic RBF, the augmented RBF is extended by a polynomial term (Sharifi, 2006). This approach could be part of a future work.

The next important step within the implementation procedure of an RBF interpolant is the determination of an appropriate shape parameter c^2 . Two possible realizations will be provided in the following section.

5.2 Computation of the shape parameter for a RBF kernel

Most of the interpolation kernels contain a shape parameter c^2 . This can either be a single parameter which is constant for all data centers or a parameter vector with different values for each data center.

The mean quadratic distance between the data centers is one possible definition of a single shape parameter. Its computation is implemented in `fsingle_shape_parameter.m` and is defined as follows (Shul'min and Mitel'man, 1974):

$$c^2 = \frac{1}{n_0(n_0 - 1)} \sum_{i=1}^{n_0} \sum_{j=1}^{n_0} \rho_{ij}^2. \quad (5.9)$$

If the local distribution of the known points should have an influence in order to improve the flexibility of the basis function, the following approach called minimum distance shape parameter vector has been realized in the Matlab function `fshape_parameter_min.m` (Ruprecht and Müller, 1995):

$$c_k = \min\{\rho_{i,k}\} \quad i, k = 1, 2, \dots, n_0. \quad (5.10)$$

This parameter c_k represents the distance from the data point $P(x_k, y_k, z_k)$ to its nearest neighbour. The result is a vector \mathbf{c} with n_0 individual shape parameters c_k^2 for each distinct data center. For both methods, however, the shape parameters are derived independently of the type of the interpolant kernel which makes it quite difficult to guarantee an acceptable accuracy for all types of radial basis functions. As a consequence, the performance of an RBF is quite arbitrary, mainly due to the random selection of one or more shape parameters. Hence, the objective of the next sections will be to find an appropriate criterion for the evaluation of the shape parameter, as well as an optimization algorithm based on this criteria rather than using the simple computations explained above.

5.3 Evaluation of the shape parameter

In order to compare the outcomes of different interpolation methods of scattered data, let it be by a parametric or polynomial model with varying orders as it was implemented in the *OrthoPlus* software or by different setups of radial basis functions with varying definitions of shape parameters, it is most important to find an appropriate assessment criterion. The next sections will give four different possibilities how to evaluate the accuracy of the results of different interpolation methods.

5.3.1 Cross-validation procedure (one-leave-out method)

The first assessment criterion is the cross-validation procedure or also called one-leave-out method which has already been introduced in section 3.3.4.3. and which is implemented in the Matlab function `fcrossvalid.m` within *OrthoPlus*.

By leaving-out the coordinates of one data center P_i together with its height misclosure l_i from the distinct data set, one can estimate the unknown coefficients and the shape parameter of the RBF with the new subset of the remaining data points $n_0 - 1$. These estimated parameters can be used to predict a new height misclosure \hat{l}_i at the excluded point. The difference between the predicted and known height misclosure is the interpolation error $\Delta_i = \hat{l}_i - l_i$ at the excluded point P_i . If these steps are repeated for every of the n_0 data points, the result will be a vector $\mathbf{\Delta}^{\text{cross}} = (\Delta_1, \dots, \Delta_{n_0})^T$ of interpolation errors for all data centers. The root mean square of this vector is:

$$RMS_{\Delta}^{\text{cross}} = \sqrt{\mu_{\Delta}^2 + \sigma_{\Delta}^2}. \quad (5.11)$$

In the same way as this $RMS_{\Delta}^{\text{cross}}$ was already implemented in *OrthoPlus*, it will be used again as a measure to evaluate the accuracy of the selected RBF with respect to their performance of predicting new points. The disadvantage of this procedure is that the computation of the $RMS_{\Delta}^{\text{cross}}$ is only based on the discrete and specific known distinct data points but not on the deleted points (Eq. 5.1).

5.3.2 Evaluation at check points

In order to take the benchmark points into account which are not distinct and therefore have been deleted by applying the function `fdistinct_center.m`, a slightly different evaluation process can be defined. This procedure works in the same way as the cross-validation. This time, however, the interpolation error is not computed by leaving out one single observation of the n_0 data points successively, but is based on the group of the n_d deleted points which have been stored but not been used in the estimation process. These deleted points will be referred to as check points in the sequel.

After determination of the unknown coefficient vector α , the height misclosures \hat{l}_i at the deleted check points P_i with $i=1, \dots, n_d$ can be predicted by applying any specific interpolation function and can be compared to their known height misclosures l_i . Thus, the interpolation error is defined as:

$$\Delta_i = \hat{l}_i - l_i \quad i = 1, 2, \dots, n_d, \quad (5.12)$$

which results in an RMS value derived from the vector $\mathbf{\Delta}^{\text{check}} = (\Delta_1, \dots, \Delta_{n_d})^T$ of interpolation errors by applying Eq. (5.11). This measure is an alternative to the RMS of the cross-validation procedure and uses none of the evaluated check points in the estimation process of the unknown coefficients. However, it might be too optimistic at the same time as only a few points in random parts of the dataset are evaluated. Furthermore, each point which was left out, i.e., each check point, is located close to another point, i.e., one of the distinct centers. So, the interpolation error Δ_i is connected to the error in the estimated point or to the difference between used point and left-out point.

5.3.3 Modified cross-validation procedure: single-step scheme

With respect to the evaluation of radial basis functions and to an optimization algorithm for the shape parameter, the computation of the RMS value as it was described in section 5.3.1 is far too complex. The reason for this complexity is that in the implementation of RBF one has to solve a linear system of order $(n_0 - 1) \times (n_0 - 1)$ for every single excluded point P_i with $i = 1, 2, \dots, n_0$. That means that n_0 different design matrices have to be inverted in order to calculate the differences between estimated and known height misclosures at n_0 excluded points. Even for a modest number of benchmark points, this process is a very time-consuming factor. Therefore, the implied cross-validation procedure should be modified if RBF interpolation methods are used in combination with an optimization algorithm. One suggestion is to apply the single-step scheme. Rippa (1999) derived this innovative and simplified evaluation criterion specifically for the RBF interpolation method. The interpolation error $\Delta_i = \hat{l}_i - l_i$ between estimated (\hat{l}_i) and known (l_i) height misclosures at the excluded point P_i can be computed as follows:

$$\Delta_i = \frac{\alpha_i}{B_{ii}} \quad i = 1, \dots, n_0, \quad (5.13)$$

where α_i represents the coefficients of the unknown $n_0 \times 1$ vector $\boldsymbol{\alpha}$ which can be estimated by applying Eq. (5.7). B_{ii} is the i -th diagonal entry of the inverse design matrix of \mathbf{A} (Eq. 5.6) with $\mathbf{B} = \mathbf{A}^{-1}$.

The advantage of this kind of evaluation criterion is that the computation process can be reduced to only one inversion of the initial design matrix \mathbf{A} with a size of $n_0 \times n_0$ and without deleting any rows or columns. That means that the unknown coefficients $\boldsymbol{\alpha}$ are only computed once based on the full set of distinct observations. By dividing the entries of this estimated unknown vector by the corresponding diagonal entries of the inverted design matrix (Eq. 5.13), one can compute the complete error vector $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_{n_0})^T$. This is a significant reduction in complexity compared to the n_0 inversions of the cross-validation procedure. The $RMS_{\Delta}^{\text{single}}$ value can again be computed by Eq. (5.11).

From the definition of the interpolation error in Eq. (5.13), it becomes clear that the $RMS_{\Delta}^{\text{single}}$ of the single-step evaluation scheme can be only used for RBF models but not for the evaluation of any parametric models due to the different implementation of the design matrix \mathbf{A} . By contrast, the RMS value of the cross-validation procedure $RMS_{\Delta}^{\text{cross}}$ can be computed for the parametric as well as for the

RBF models. While it takes very similar values as the $RMS_{\Delta}^{\text{single}}$ in the case of an RBF interpolation, it is, however, too complex to be used in combination with a GA optimization. In this case, the single-step scheme, implemented in the function `feval_multiquadric.m`, can be proposed as an appropriate evaluation function for an optimization of the shape parameter.

5.3.4 Combined evaluation

The fourth type of assessment criterion which is investigated in this work and set up in `feval_multiquadric_with_check_points.m` combines the results from the check point evaluation (Eq. 5.12) with the RMS value of the cross-validation procedure (Eq. 5.11) if parametric models are applied or with the RMS of the single-step scheme (Eq. 5.13) if RBF interpolation is used. The resulting RMS values will be derived as follows for the parametric models:

$$E_{\Delta} = RMS_{\Delta}^{\text{cross}} + \sqrt{(\mathbf{\Delta}^{\text{check}})^T \cdot \mathbf{\Delta}^{\text{check}}}, \quad (5.14)$$

and for the RBF model:

$$E_{\Delta} = RMS_{\Delta}^{\text{single}} + \sqrt{(\mathbf{\Delta}^{\text{check}})^T \cdot \mathbf{\Delta}^{\text{check}}}. \quad (5.15)$$

These evaluation values will be definitely larger than all the other RMS values described in the sections 5.3.1 – 5.3.3 but at the same time will guarantee that the outcome is not too optimistic as firstly all data points and secondly all deleted non-distinct points will be taken into account.

By using the Matlab code `RBF_without_genetics.m`, all of these four different evaluation criteria have been tested for the RBF model with a single shape parameter (`fsingle_shape_parameter.m`), with a minimum distance shape parameter vector (`fshape_parameter_min.m`) and for a second-order polynomial model whereas the single-step scheme could not be applied in this case. The tests have been applied to the Canadian dataset and to its four regional datasets, Western, Central, Eastern Canada and the Maritimes respectively. The results and a comparison are shown in section 5.5.1. The main outcome of these tests is a comparison between the performance of the parametric model and the radial basis function. It will be seen that the choice of the shape parameter c^2 , which can either be one common value for all data centers or a vector with one value for each data center, has a considerable influence on the accuracy that can be achieved when

interpolating scattered data with RBF. Therefore, it is important to find an optimization algorithm for c^2 in order to minimize the interpolation error. Such an algorithm has to take into consideration that the value of the shape parameter not only depends on the type of radial basis interpolant, but also on the number and distribution of data centers which determines the conditioning of the design matrix \mathbf{A} , on the values of the height misclosure vector and on the precision of the computation (Rippa, 1999).

5.4 Optimization algorithm for the shape parameter

There are several different methods for the optimization of the shape parameter in order to obtain a reasonable interpolation accuracy of radial basis functions. One possibility, for example, is to compare the RMS values of a specific assessment criterion for different values of c^2 . The one c^2 -value for which the RMS becomes minimal is seen as an approximation of the optimal shape parameter. An example for such an algorithm can be found in Carlson and Foley (1991).

Another algorithm which is investigated in this work is the genetic algorithm (GA) which is a powerful optimization technique. Every genetic algorithm requires the implementation of a fitness, evaluation or object function. All evaluation measures described in the last section 5.3 are mainly dependent on the selection of the specific shape parameter which will be proven in section 5.5.1 and can theoretically be used as fitness functions. They should be, however, rather called misfit functions as the objective is not to maximize the fit, but to minimize the misfit. Both implementations are basically identical. The misfit is the output of the misfit function in the form of an RMS value. The reader should keep in mind that it is important to minimize the RMS values as outcomes of the misfit function in order to optimize the shape parameters c^2 . The misfit function is commonly also referred to as cost function. As the performance of the GA should be as good and fast as possible, especially the RMS values of the single-step evaluation scheme (section 5.3.3) and of the combined evaluation (section 5.3.4) are appropriate to be used as misfit functions within the genetic algorithm.

In the sequel, a general overview of the theory of the genetic algorithm will be presented at first (section 5.4.1). In order to efficiently implement an optimization of the shape parameter by using the genetic algorithm, a Matlab code called `genetics.m` was set up within the scope of this work. It uses the “*Genetic Algorithm and Direct Search Toolbox*” of Matlab. Section 5.4.2 will focus on a more detailed description of this practical implementation and the corresponding code.

5.4.1 Theoretical overview of the genetic algorithm (GA)

This section will give a short overview of some theoretical aspects of the GA and is not meant to be a detailed investigation. For further details, the reader is referred to Beasley et al. (1993a), Haupt and Haupt (2004) and Sharifi (2006).

5.4.1.1 Introduction

“Optimization is the process of making something better.” (Haupt and Haupt, 2004). In this work, the objective of applying an optimization algorithm is to find an optimal shape parameter for each data centre which minimizes any kind of interpolation error and the misfit function respectively, and improves the interpolation with the radial basis functions at the same time. The reason why especially the genetic algorithm (GA) is suitable for this work is that it is a global optimization method. That means, it is a minimization technique which will find the true minimum value of the whole misfit function and not only the lowest value within a finite neighbourhood as a local search method would do. A misfit function can have more than one local minimum but only one global minimum. Therefore, it is much more complicated to look for a global than for a local minimum.

The speciality of the GA is that this global and adaptive search algorithm is inspired by the genetic processes of biological organisms and therefore belongs to the evolutionary algorithms. The method works analogue to the biological optimization and uses some of the same specialist terms (section 5.4.1.2). It is based on the idea of natural selection and survival of the fittest which was developed by Charles Darwin in “The Origin of species” (1859). The rules of evolution are applied to the individuals. First pioneered by the computer scientist John Holland in the 1960s and 1970s, GAs have been widely studied, experimented and applied in many fields such as biogenetics, computer science, engineering, economics, chemistry, manufacturing, mathematics, physics and others. The advantages of GAs compared to traditional optimization approaches are listed in the following (Haupt and Haupt, 2004):

1. optimization with continuous or discrete variables
2. no requirement of derivative information
3. simultaneous searches from a wide sampling of the fitness functions
4. possibility of processing with a large number of variables
5. suitability for parallel computers
6. optimization of variables with extremely complex fitness functions (no getting stuck at a local minimum)
7. disposal of a list of optimum solutions instead of a single solution
8. possibility of optimization with encoded variables
9. ability of processing simulated data, real observations or analytical functions.

The algorithm looks at first sight like a random search technique to find an optimal value. However, it is a highly intelligent technique which explores new areas and unknown points with possibly better solutions within the search space by exploiting the information of already processed points. This strategy helps to prevent the solution being stuck in a local minimum and guarantees an efficient search. It is also referred to as a heuristic method as the result gradually converges to an optimal solution within several iterations which are called generations in the sequel. The potential solutions are represented on a chromosome-like structure.

However, the reader should be aware of the fact that GAs are not always the best choice for global optimization. A major drawback is the expensive computational effort which might not be efficient for simpler problems with less variables, real time use or problems with only one local minimum which then is the global minimum at the same time, for example.

5.4.1.2 Terminology

The introduction and explanation of the terminology within this section provides a short background on evolutionary biology and natural selection processes. By means of the most important key terms in the following table together with a description of their meaning in biological optimization on the one hand and in the numerical optimization (GA) on the other hand, it should be relatively easy for the reader to understand the GA in the sequel of this section.

Table 5.2: Key terms within biological and numerical optimization (GA)

term	meaning in biological optimization	meaning in numerical optimization (GA)
gene	basic unit of heredity, located on a chromosome	variable
chromosome	carries organism's genes in form of DNA	array of variables (genes), IMQ chromosomes consistent of genes (one shape parameter per data center)
generation	act of producing offspring	one iteration of the GA
individual	any possible appearance of an organism, characteristics determined by chromosomes	any possible solution, here synonym to chromosome
population	group of all interbreeding individuals	group of individuals / solutions with which is currently dealt

In biology, *evolution* is a change in the gene pool of a population from generation to generation in order to better suit the environment by passing on genetic material (DNA). It does not mean that the individual itself will change, but that the fitter individuals within a population will survive longer since they will have a bigger chance to find enough resources, such as food or water. If these individuals with traits fitting best into the environment will also find a mating partner, their genes will be passed on to an increasing number of offsprings. This choice of individuals for further reproduction is called *natural selection*. By contrast, there is a high risk for other poorly performed individuals of the population to die out because of a smaller opportunity of reproduction. After some generations of this evolutionary process, optimized superfit offsprings will be reproduced which become more and more suited to their environment and have an even better fitness than their parents and grand-parents. In the same way, the actual numerical genetic algorithm will lead to a continuous improvement of the solution until finally an optimal solution of the underlying problem is achieved. The methodology will be described in the following section.

5.4.1.3 Methodology

Before starting a genetic algorithm, the user has to decide in which way the variables, i.e., the genes, are represented best. There are mainly two different possibilities: the binary or continuous data coding. If the genes are encoded as binary strings, the corresponding binary genetic algorithm (BGA) uses these binary strings for optimization purposes. As the shape parameter within this work is expected to be real-valued, it would have to be encoded before applying the BGA. However, the solution of an optimization problem with a continuous variable like the shape parameter can be implemented directly without quantization and consequently in an easier and faster way: the continuous genetic algorithm (CGA). The CGA uses directly the floating-point number of the shape parameter as a single real-valued gene representation and skips its encryption into binary strings. Here, the explanation is focused on the CGA although the implementation of the BGA is very similar. The representation of a floating-point number is limited due to internal precision and the round-off error of the computer. For further information on the BGA and differences between both algorithms, mainly in the crossover and the mutation operators, the reader is referred to Sharifi (2006).

The flowchart in Figure 5.1 shows the easiest and most common scheme of such a CGA. The following paragraphs explain each step and give concrete solutions for the implementation with respect to the optimization of a shape parameter of the RBF.

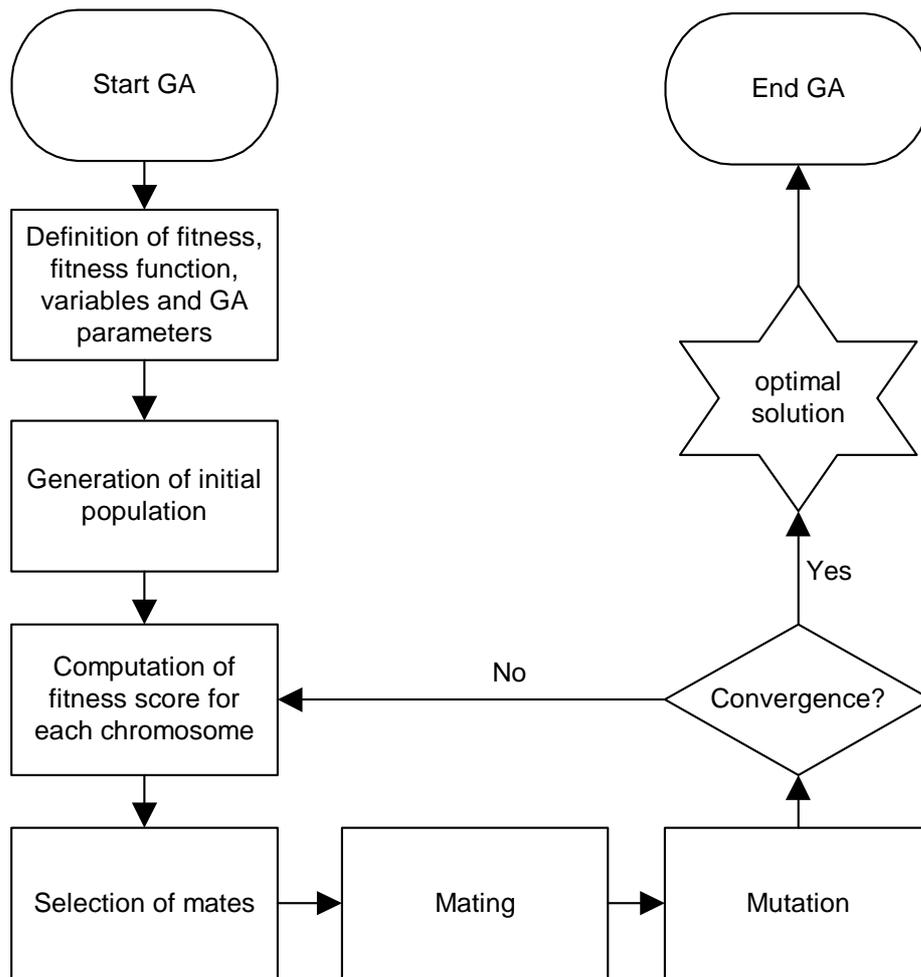


Figure 5.1: Flowchart of a continuous GA

Definition of variables, fitness function and GA parameters As an initialization step, the user first has to define the variables as well as their fitness, or their misfit function respectively. Usually, every variable represents a gene and all genes together form a vector, i.e., the chromosome. The chromosome in this study consists of $n_{\text{var}} = n_0$ individual shape parameters c_k^2 where each of them is ascribed to one of the n_0 data centers within the RBF:

$$\text{chromosome} = [c_1^2, c_2^2, c_3^2, \dots, c_{n_0}^2]. \quad (5.16)$$

The fitness function computes the fitness as an output dependent on the input variables, i.e., the chromosome, which should be optimized:

$$fitness = g(chromosome) = g(c_1^2, c_2^2, c_3^2, \dots, c_{n_0}^2). \quad (5.17)$$

The higher the fitness score, the better is the optimization of the shape parameters. The fitness function is provided by the programmer and gives each chromosome a score based on how well it performs at the given task.

This study uses two different types of fitness functions. It is very important to keep in mind that an optimization of the shape parameters c^2 leads to a maximal fitness score but to a minimal RMS value which is the output of the misfit function. Due to performance issues, especially the RMS values of the single-step scheme (section 5.3.3) and of the combined evaluation (section 5.3.4) were deemed to be appropriate.

The parameters used within the GA should also be selected carefully in the beginning as they can have an evident impact on the GA's performance. These parameters include, for example, the size of the initial population, the selection and mutation rate and the convergence criteria. As the Matlab toolbox provides a large choice, the specific settings of these and other variables will be explained in the next section 5.4.2.

Initial population The initial population consists of n_{pop} randomly generated individuals. As in biology, the chromosomes only determine the specific appearance of an individual. The choice of the number of initial individuals, the chromosomes respectively, has an influence on how fast the GA process will converge. At a certain number, adding more individuals does not increase the speed or the efficiency of the GA. A too low number for n_{pop} , however, involves the danger of premature convergence.

Fitness score After creating a random group of individuals from a search space to form the initial population, each of the population members is evaluated by using them as an input into the misfit function. Thus, the corresponding score will be computed for each individual according to its misfit. Individuals with low RMS values receive a low cost, i.e., they have a low misfit and a high fit. The search space contains a number of possible random solutions for the shape parameters in the form of individuals which can be chosen by the user prior to the start of the GA. For a successful optimization, it is not necessary to provide concrete a-priori information for the shape parameters as the GA searches narrowly for an optimal solution within the whole search space.

Selection The individuals with the highest fitness scores will then be selected as parents into the mating pool of size n_{keep} corresponding to the principles of the survival of the fittest or the natural selection of the best individuals. This step leaves out the $n_{\text{pop}} - n_{\text{keep}}$ individuals of lower fitness scores and thus, creates space for new offsprings descending from the higher fitness individuals from the mating pool in order to improve and continue the process. It is quite common to keep half of the population for the mating pool which means that two parents will have two offsprings. These n_{keep} fittest individuals will be divided into pairs in order to form the parents in the mating pool. The statistical selection of potential partners is one of the genetic operators. There are several schemes for the selection process, such as the pairing from top to bottom, the normal random pairing, the random pairing by rank or cost weighting (roulette wheel weighting) or the tournament selection, to name the most important ones. For details, the reader is once again referred to Haupt and Haupt (2004). Although roulette wheel and tournament selection are standard for most GAs, Matlab uses another method as default, the stochastic uniform selection, which will be discussed in section 5.4.2.

Mating The selected parents are "reproducing" with a certain probability to create one or more offsprings with better fitness values. During this mating process, the genetic material between two parents is interchanged and recombined in a crossover procedure. Thus, the resulting offsprings will carry characteristics of both mating partners. If the selection step sorts out $n_{\text{pop}} - n_{\text{keep}}$ individuals, then enough, i.e., at least $n_{\text{pop}} - n_{\text{keep}}$ offsprings, are generated during the whole mating process of all individuals in order to bring the population back to its original size n_{pop} . The user can choose between different methods for the recombination of the genetic information of two selected individuals. The easiest method is the single-point or simple crossover in which the chromosomes of each parent are cut at a randomly selected crossover point and the originating tail chromosome segments, the variables between the crossover point respectively, are merely swapped between the two parents. The chromosomes of the two new individual offsprings will each have a unique combination of genetic material with a part from the father and a part from the mother. Alternatively and for an even better recombination of genetic material, two or more crossover points can be chosen and the methods are analogously called two-point or multi-point crossover (Figure 5.2).

Application of further crossover methods such as the generalized or uniform crossover can be found in Beasley et al. (1993b). In this work, however, the arithmetic recombination is used which is especially suitable for the continuous GA as it introduces additional new genetic information by combining genes from the two parents through a specific mathematical relationship into new variable values in the offsprings. This algorithm is also implemented in the Matlab toolbox under the name intermediate

crossover (see section 5.4.2). If all genes within a chromosome are manipulated, then the mating is called whole arithmetic or flat crossover. Variations are the single arithmetic and the simple arithmetic crossovers where only a few genes will be changed for recombination (Sharifi, 2006).

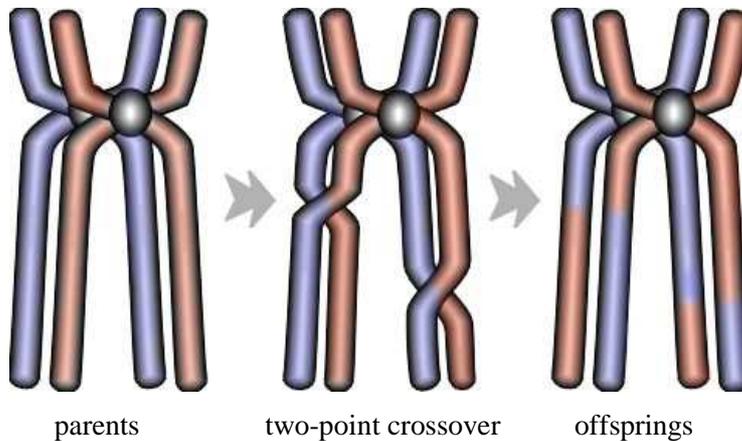


Figure 5.2: Two-point crossover

Mutations Additionally to the crossover, the GA will cause small random changes in some of the genes. These so-called mutations might be passed along to the offsprings. The mutation is the second step in the recombination process of forming a new generation. As external and spontaneous force, which is, for example, caused by environmental factors in biology, it allows for genetic diversity in the population and enables the GA to search a broader space in order to prevent the GA process of converging prematurely. One option is the uniform mutation which modifies a randomly selected individual from the mating pool with a small mutation probability p_m usually between 5% - 20% by adding a vector of uniformly generated numbers as random noise. If this vector contains normally distributed random numbers, the procedure is called Gaussian mutation instead. The genes of the individuals with the highest fitness scores, however, are deemed to be the elite solution and will therefore not be mutated.

Iterative process This whole GA process creates successively new generations by the application of the genetic operators such as the reproduction, crossover and mutation. The individuals of the mating pool and the new offsprings form the new population of the following generation. This continues iteratively until a certain convergence criterion is reached or a certain number of generations has passed, depending on the needs of the user. The final and usually optimal solution provides mainly new individuals with a better fitness while most of the individuals with lower fitness have died out.

5.4.2 Practical implementation of the shape parameter optimization

Within the scope of this work, the optimization of the shape parameter was implemented within Matlab by means of the “*Genetic Algorithm and Direct Search Toolbox*”. The following toolbox function minimizes the defined fitness function using a genetic algorithm:

```
[x, fval, reason] = ga(fitnessfcn, nvars, gaopt)
```

with the input variables:

```
fitnessfcn - fitness or cost function,
nvars      - number of variables,
gaopt      - algorithm option structure,
```

and the output variables:

```
x          - solution for optimized variables (shape parameters)
fval       - value of the fitness function at x
reason     - string containing the reason why the algorithm stops.
```

As already mentioned, two different cost functions were tested within the GA, the `feval_multiquadric` which is the single-step scheme evaluation and its augmentation `feval_multiquadric_with_check_points` which is the combined evaluation, both introduced in the sections 5.3.3 and 5.3.4. The variables which should be optimized are the $n_{\text{var}} = n_0$ individual shape parameters c_k^2 , i.e., one for every single distinct data centre, defined in Eq. (5.16).

The Matlab toolbox provides many different algorithm options which can be specified in the structure `gaopt`. The specifications presented in the following Table 5.3 allow to adapt every single step of the GA implemented in the function `ga` in order to solve the particular optimization problem of this work. The applied methods for the fitness scaling, the selection, the crossover and the mutation will be explained in detail afterwards. All settings and values were determined by trial-and-error until reasonable results were achieved. A systematic approach to find the best possible specifications goes beyond the scope of this thesis and will be part of future work.

Table 5.3: Specifications of algorithm options

step	specification
data type of individuals in population	double, continuous, floating point number
population size	100
creation of initial population	random initial population with a uniform distribution
fitness scaling	rank-based
selection	stochastic uniform
crossover	intermediate
crossover fraction	80%
mutation	adaptive feasible
number of elite children	2

Fitness scaling The *rank-based fitness scaling* converts the raw fitness scores into the rank of an individual which is its position in the sorted scores. This is done in order to facilitate the implementation of the selection function based on a more usable range of values as the spread of the scores is circumvented. For example, the rank of the fittest individual is 1, the second fittest individual gets the rank 2, and so on.

Selection The pairs of parents from the mating pool are chosen by the *stochastic uniform selection*. This method lays out a line in which each parent corresponds to a section of the line of length proportional to its rank. The algorithm allocates a parent by moving along the line in steps of equal size and choosing the specific individual from the section it lands on. The first step is a uniform random number less than the step size.

Crossover The *intermediate crossover* creates the offsprings O_1 and O_2 by taking a weighted average of the parents P_1 and P_2 :

$$\begin{aligned} O_1 &= P_1 + \kappa(P_2 - P_1), \\ O_2 &= P_1 + (1 - \kappa)(P_2 - P_1), \end{aligned} \tag{5.18}$$

where κ is a single random weight parameter within $[0,1]$. The crossover fraction parameter specifies that only 80% of the offsprings are produced by crossover.

Thus, the other 20% of the individuals from the next generation as well as the elite children are created by simply copying the chromosomes of the parents directly into the children population.

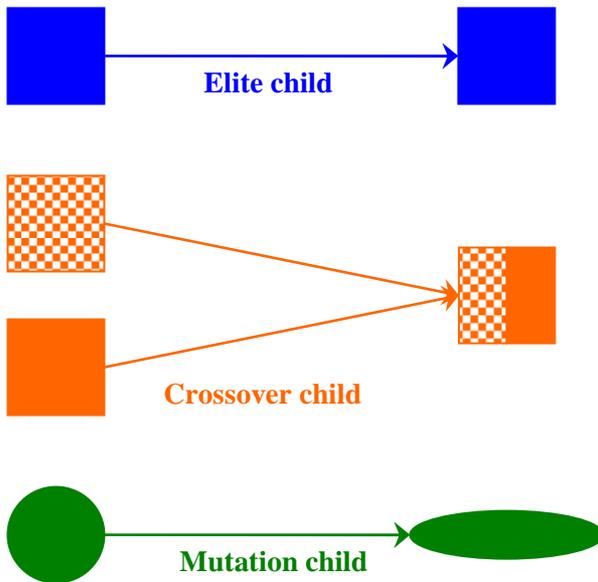


Figure 5.3: Reproduction options

Mutation In order to provide genetic diversity and enable the GA to search a broader space, the adaptive feasible mutation is applied. It creates mutation children in a feasible region by making small random changes in the individuals with randomly generated directions that are adaptive with respect to the last successful or unsuccessful generation. Furthermore, the two offsprings with the highest rank of one generation are defined as elite children and are guaranteed to survive to the next generation without being mutated. During the reproduction process, the GA creates a population of elite, crossover and mutation children (Figure 5.3) which will then replace the current population to form the next generation.

These GA steps are repeated iteratively and form a sequence of new populations until one of the following termination criteria is met. The algorithm terminates if:

- the maximum number of generations is reached, here defined as 5000
- the weighted average change in the fitness scores is less than function tolerance (10^{-16}) over (150) stall generations
- there is no improvement in the best fitness score for a specified interval of time (stall time > 200seconds).

Apart from the string containing the reason for stopping the algorithm, the function `ga` provides the user finally with the optimal solution of the variables, i.e., the shape parameter vector, as well as with the corresponding value of the fitness function, i.e., the RMS of the residuals, either for the single-step or combined scheme.

The actual results of this specific GA implementation within Matlab for RBF shape parameter optimization will be presented in section 5.5.2. The output will be also compared to the results of section 5.5.1 which show the RBF performance without using the GA.

5.5 Discussion of results

5.5.1 RBF without using GA

Before applying the genetic algorithm, this section provides the results of some tests which were only done with pure radial basis functions and different types of shape parameters without any optimization algorithm. This step is done in order to show that it is very difficult to achieve results with a reasonable accuracy if RBF are used as interpolant kernel without knowledge of an optimal shape parameter. The scheme in Figure 5.4 gives an overview of the test implementation within the main Matlab function `RBF_without_genetics.m`. The inverse multiquadric interpolant (IMQ) was chosen as kernel type for the RBF and has been applied to the five different test networks (Figure 4.16). The subfunction `fgeod2cart.m` converts the geodetic to Cartesian coordinates for the GRS80 and `fdistinct_center.m` saves the points which are not distinct (Eq. 5.1) as deleted check points.

The main objective of this chapter is to investigate the relationship between the choice of a shape parameter and the interpolation error of the IMQ kernel. As the shape parameters cannot be chosen totally randomly and as it was the objective of this section not to apply any optimization algorithms, the two different approaches as described in section 5.2 have been applied. This is first the computation of a single shape parameter as defined in Eq. (5.9) and implemented in `fsingle_shape_parameter.m`. It should be noted that in this case the resulting design matrix from Eq. (5.6) can easily become unstable and therefore some kind of regularization is required. The second possibility is the definition of a whole series of shape parameters (`fshape_parameter_min.m`), one for each data centre with a value equal to the distance between this data point and its nearest neighbour (Eq. 5.10).

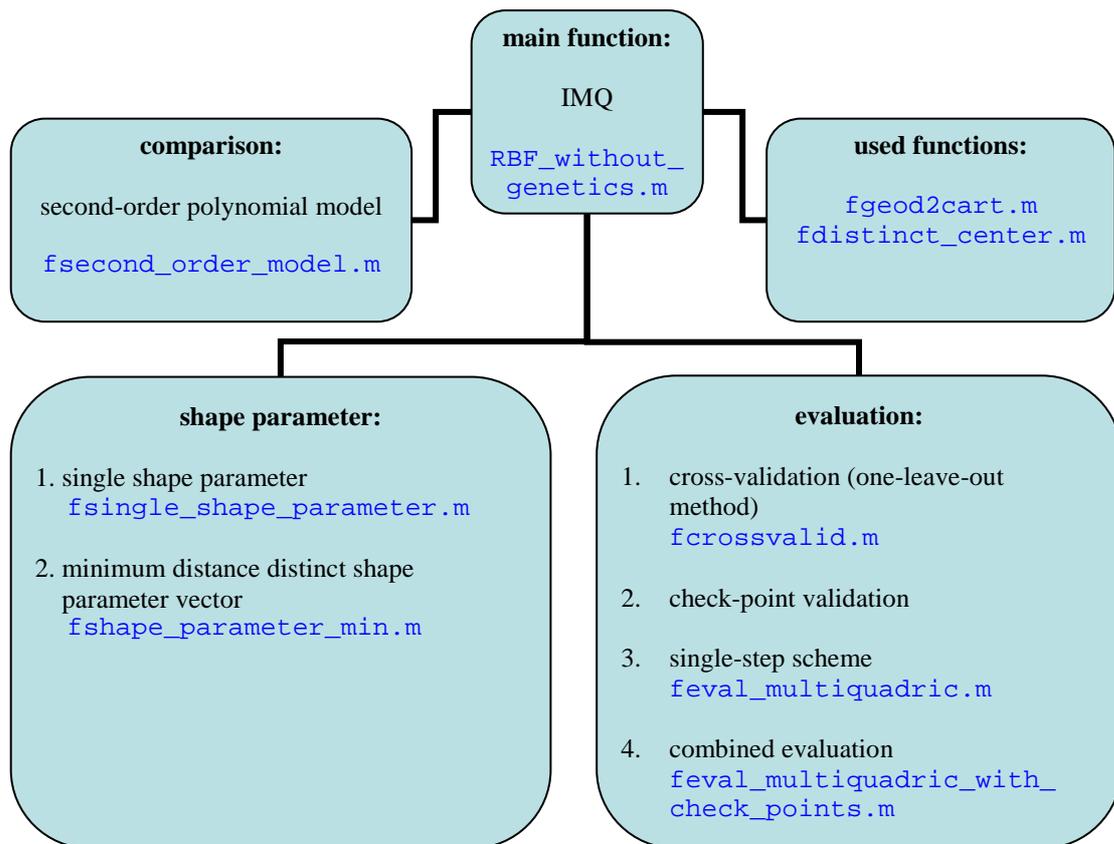


Figure 5.4: Scheme of applying radial basis functions without optimization

Both implementations of shape parameters were evaluated with all four assessment criteria and compared to the outcomes of an evaluation with a random parametric model for which the second-order model (`fsecond_order_model.m`) was chosen commonly for all of the five networks. The reader should keep in mind that the single-step evaluation can only be applied in case of an RBF interpolant but that the resulting RMS value is expected to be very similar to the RMS of the cross validation procedure. This makes a comparison between the application of an RBF interpolant and of parametric models possible.

The corresponding results are shown in the following tables (Table 5.4 - Table 5.7). The RMS values for the cross-validation (Table 5.4) get relatively large for the Maritimes network with more than 1 m if the RBF single shape parameter is used. After deleting six non-distinct points, only 24 points are left within the whole network. This little number of center points as well as their sparse distribution and the possibly resulting instability of the system might be the reason why especially the RBF cannot achieve good interpolation results for these scattered data. Even if the second-order polynomial model is used, the RMS is with almost 10cm worse than for the other networks.

It should be pointed out that the RMS values from the cross-validation with second-order polynomial model are slightly different to the values presented in the regional analysis section 4.5 due to the

deletion of the non-distinct data points. The level of these differences depends on the percentage of deleted points and the constellation of the new dataset of center points. Apart from the Maritimes network, the results achieved by `RBF_without_genetics.m` and those achieved by *OrthoPlus* diverge less than 5 mm. The difference within the Maritimes is almost 4 cm mainly due to the fact that 20.0% of the original benchmark points are not distinct and had to be removed. With the one-leave-out method, the second-order polynomial model seems to provide better RMS values than the RBF interpolant in general, no matter if just the single or the minimum distance shape parameters are used. Furthermore, the results seem to be quite random depending on the combination of shape parameter and network. It is expected that the RBF could achieve similar or even better results than any parametric model if an optimal shape parameter can be found. The Canadian dataset, for example, is the only case where the RMS value for the RBF with the single shape parameter becomes minimal with 5.2 cm and is thus smaller than the RMS for the second-order polynomial model with 5.4 cm. This leads to the assumption that the RBF model with a single shape parameter might be sufficient for the relatively large Canadian network, but not 'local' enough for the smaller networks. In this case, the RBF interpolation forms a relatively smooth surface with respect to the regional networks. That means that if two benchmark points are situated close to each other, they take similar estimated height misclosures and the columns of the corresponding design matrix (Eq. 5.6) will not differ much from each other. This will most likely lead to an instable design matrix since the columns are not completely independent. If this is the case, the RBF interpolation suffers from over-parameterization in the regional networks. The question in which networks, however, the RBF method without optimization is still sufficient, needs some further investigation in the future. Furthermore, it becomes evident that the smoother the region, the higher is the possibility for a rank-deficient design matrix. Consequently, the RMS values of an interpolation with RBF model and single shape parameter within the smooth Central Canadian network are expected to be higher than those within Western Canada with a rough terrain, for example. This assumption can in fact be verified for all four evaluation procedures in the following tables.

Table 5.4: Comparison of RMS values for cross-validation (one-leave-out method)

RMS values in [cm]	Canada	Western Canada	Central Canada	Eastern Canada	The Maritimes
RBF single shape parameter	5.19	6.25	22.98	7.55	110.30
RBF minimum distance shape parameters	11.70	12.14	6.87	7.39	96.75
2nd order polynomial model	5.37	5.72	4.53	5.12	9.45

Let us take a look at Table 5.5. It shows the RMS values from the check-point evaluation which pretends the non-distinct data points to be unknown. The results are much better than for the cross-validation procedure. This is especially the case for the RBF with the minimum distance shape parameters which can be seen in the Eastern Canadian network, for instance, with an RMS value of approximately 8 mm compared to almost 8 cm for the cross-validation. In the Maritimes network, the RMS value is 1.5 cm compared to almost 97 cm in Table 5.4. However, the results from this kind of validation are obviously too optimistic as the check points are very close to the data centers because they have been classed as non-distinct before. Thus, the RMS errors of the check points do not give a very good picture of the general interpolation error at some random points distributed all over the network.

Table 5.5: Comparison of RMS values for check-point validation

RMS values in [cm]	Canada	Western Canada	Central Canada	Eastern Canada	The Maritimes
number of check points	23 out of 430 (5.3%)	10 out of 148 (6.8%)	5 out of 76 (6.6%)	9 out of 176 (5.1%)	6 out of 30 (20.0%)
RBF single shape parameter	31.91	3.40	28.82	44.05	5.52
RBF minimum distance shape parameters	1.69	2.90	5.44	0.79	1.53
2nd order polynomial model	3.60	5.50	5.53	2.22	4.58

A more reliable assessment criterion of the performance of the RBF interpolator is realized by the single-step scheme. An advantage of this procedure is that the initial design matrix including all data centers only has to be inverted once. The fact that it cannot be applied to any parametric models is a drawback and makes a direct comparison impossible. However, it is feasible to compare the RMS values of the one-leave-out method (Table 5.4) for the parametric models with the RMS values of the single-step scheme (Table 5.6) for the RBF interpolant. In Table 5.6, it can be seen that the RMS values for the RBF with the single shape parameter are still quite random with a maximum of about 1.3 m for the Maritimes network but that the results for the RBF with the minimum distance shape parameter vector improved significantly compared to the results of the simple cross-validation. The RMS of the Maritimes, for example, becomes smaller than 7 cm. For Canada and Western Canada, the RMS values are approximately halved with ca. 5 cm and 6.1 cm if the single-step scheme instead of the one-leave-out method is used in case of the minimum distance shape parameters. This leads to the assumption that especially for this choice of the shape parameter setup as well as for these networks, the single-step scheme seems to provide a system with a higher stability than the one-leave-out method and will

therefore be compared to the cross-validation RMS values of the second-order polynomial model in the following. Even without optimization of the shape parameter, the RBF with the minimum distance shape parameter vector achieves better results than the second-order polynomial model for three out of five networks, with ca. 5.0 cm for Canada, 4.7 cm for Eastern Canada and 6.9 cm for the Maritimes. Just the RMS values for Western and Central Canada are slightly worse.

Table 5.6: Comparison of RMS values for single-step scheme

RMS values in [cm]	Canada	Western Canada	Central Canada	Eastern Canada	The Maritimes
RBF single shape parameter	5.60	14.37	52.05	7.07	130.88
RBF minimum distance shape parameters	5.03	6.13	4.66	4.73	6.91

The fourth and last method to evaluate the performance of an interpolant kernel is the combined evaluation (section 5.3.4). Although this kind of assessment criterion supplies higher RMS values and therefore the outcome seems to be worse than with the other criteria, it is nevertheless very suitable for a comparison between RBF and parametric interpolant since all components, i.e., check as well as data points, are taken into account. Let us first take a look at the two last rows of Table 5.7.

Table 5.7: Comparison of RMS values for combined evaluation

RMS values in [cm]	Canada	Western Canada	Central Canada	Eastern Canada	The Maritimes
RBF single shape parameter	19.86	26.29	72.34	8.23	139.28
RBF minimum distance shape parameters	12.96	14.84	16.27	7.06	10.34
2nd order polynomial model	22.24	22.29	16.25	11.66	20.58

If these RMS values are compared to each other, one can notice that the RBF with a minimum distance shape parameter vector with a minimum value of about 7 cm within Eastern Canada and a maximum value of 16.27 cm within Central Canada achieves usually better interpolation results than the second-order polynomial model. This last value is approximately the same as for the parametric model with 16.25 cm. The best RMS value which can be achieved with the second-order polynomial model is almost 11.7 cm within Eastern Canada due to a rather large number of benchmark points which are distributed relatively regularly. In contrast to the RBF with minimum distance shape parameters, the

RMS values of the RBF with a single shape parameter are significantly higher. The RMS within the Maritimes, for instance, reaches almost 1.4 m. Just for two out of five networks, the RBF with this special kind of shape parameter achieves better results than the second-order polynomial model with about 8.2 cm within Eastern Canada and almost 20 cm within the Canadian network. Thus, it can be concluded that the computed single shape parameters do not seem to be optimal for the RBF interpolation. This inappropriateness of the single shape parameter can also be seen in the following Figure 5.5 - Figure 5.9 which show the RMS values in meters dependent on the choice of the single shape parameters for all five different networks. The red cross within each plot represents the computed single shape parameter from Eq. (5.9). The shape parameters were chosen in an interval from 1 up to a value of twice the computed shape parameter.

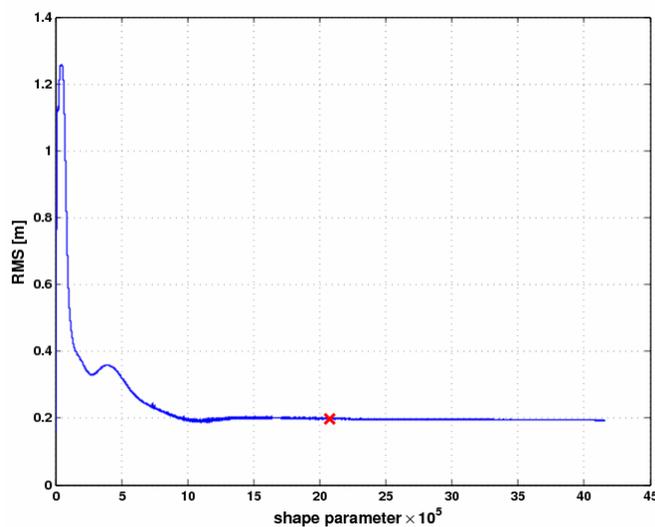


Figure 5.5: RMS values for combined evaluation dependent on varying single shape parameters for whole Canada

The computed minimum shape parameters of the Canadian network (Figure 5.5) and of the Eastern Canadian network (Figure 5.8) seem to be close to the optimal values as the corresponding RMS values are nearly reaching a minimum in these plots which just might be coincidence.

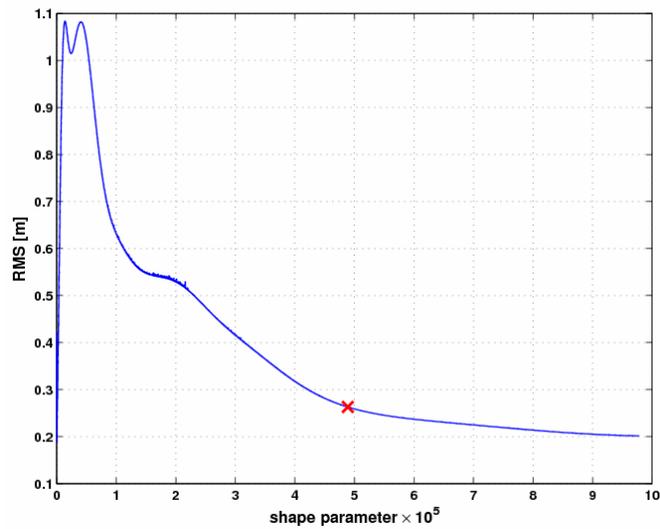


Figure 5.6: RMS values for combined evaluation dependent on varying single shape parameters for Western Canada

Within Western Canada (Figure 5.6), it also looks as if the RMS value of the computed minimum shape parameter is close to a minimum and with it the shape parameter almost optimal. Indeed, with a short look back to Table 5.7, one can see that the RMS value of the RBF interpolant with 26.3 cm is only 4 cm worse than for the parametric model.

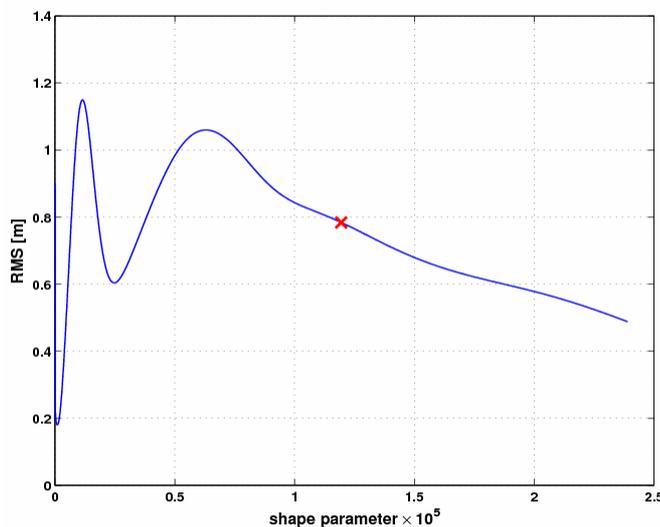


Figure 5.7: RMS values for combined evaluation dependent on varying single shape parameters for Central Canada

Figure 5.7 for Central Canada as well as the last Figure 5.9 for the Maritimes, show that the computed shape parameters (red crosses) seem to be far away from an optimum as in both figures, the RMS values are not located next to a minimum but within a strictly monotonic decreasing part of the RMS

curve. In fact, the minimum RMS values belonging to the optimal shape parameters are even higher than twice the computed shape parameters. In other words, the obtained shape parameters are obviously not the optimal solutions.

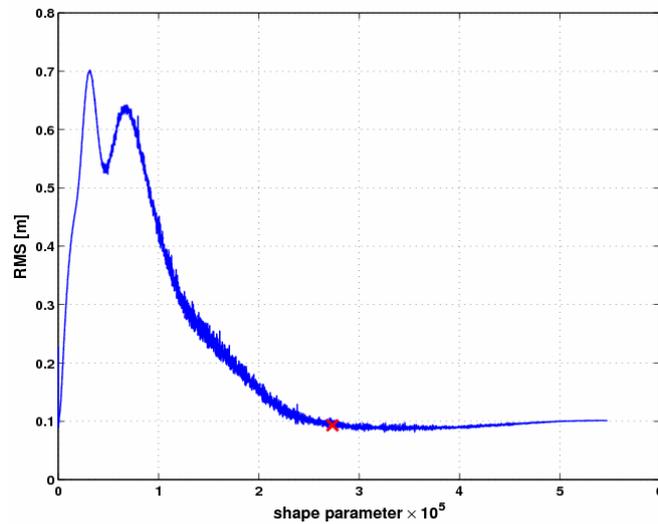


Figure 5.8: RMS values for combined evaluation dependent on varying single shape parameters for Eastern Canada

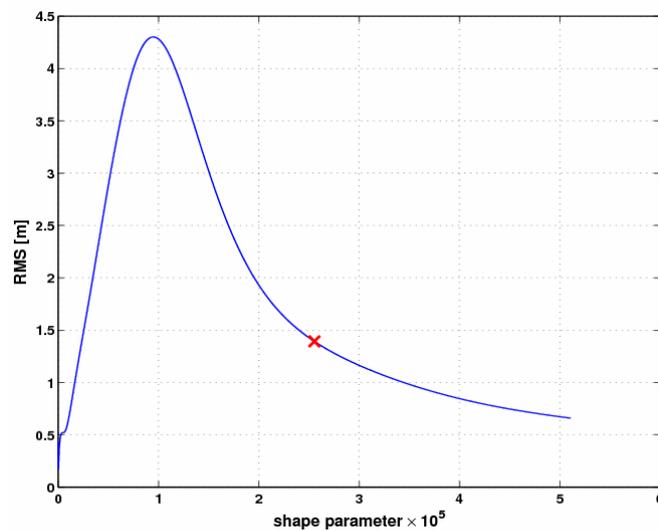


Figure 5.9: RMS values for combined evaluation dependent on varying single shape parameters for the Maritimes

The computation of the optimal shape parameter and thus the resulting performance of an RBF interpolant seems to be relatively random. Moreover, all fitness functions which compute the interpolation error in the form of an RMS value, are likely to have more than one local minimum.

As a conclusion, an optimization of the shape parameter in the form of a genetic algorithm is considered indispensable and will guarantee the global minimization of an interpolant error function.

5.5.2 RBF with GA optimization

As it was proven beforehand, the performance and accuracy of the inverse multiquadric RBF as kernel interpolant is closely dependent on the global optimization of the shape parameter vector with the entries representing one specific shape parameter for every distinct data centre.

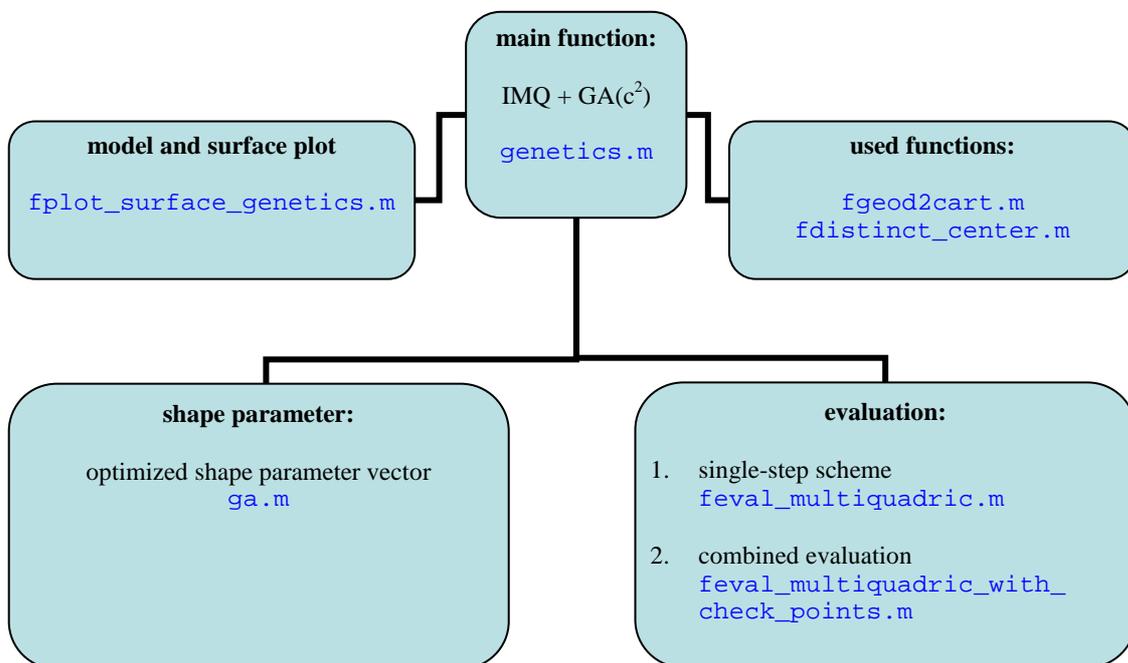


Figure 5.10: Scheme of applying radial basis functions with GA optimization

This optimization is implemented in Matlab by means of a genetic algorithm and is again applied to the Canadian network and to each of the four regional networks of Western, Central, Eastern Canada and the Maritimes. Figure 5.10 shows the structure of the main program `genetics.m`. The subfunctions `fgeod2cart.m` and `fdistinct_center.m` are again integrated in the main function and have the same purpose as for `RBF_without_genetics.m`. The actual optimization is realized within the toolbox function `ga.m` which has been already explained in section 5.4.2 together with the exact testing procedure. The main program is run twice for each dataset, once for the single-step evaluation (`feval_multiquadric.m`) and the second time for the combined evaluation (`feval_multiquadric_with_check_points.m`). In the following, the interpolation errors from both evaluation functions will be analyzed in comparison to the outcomes of the parametric model

interpolation, namely the second-order polynomial model, and to the RBF without optimization. Finally, the model itself and the projection of the derived height misclosures on a grid will be plotted by the Matlab function `fplot_surface_genetics.m`.

Table 5.8 shows in the first row the RMS values from the application of the RBF with the shape parameter vector being optimized by the genetic algorithm. The results are the outcome of the single-step evaluation. The best interpolation result with this configuration was achieved in the Eastern Canadian dataset with an RMS value of less than 3 cm. The reason is that the number of 167 distinct points is comparably large. Additionally, the distribution of the points is quite regular and dense compared to the networks of Western Canada including the Rocky Mountains or the relatively small dataset of the Maritimes with a very sparse distribution of benchmark points. Here, the RMS values become maximal with ca. 4.6 cm and 4.3 cm. They are even higher than for the much bigger dataset of whole Canada with 4.1 cm. That shows what many researchers, as for example Houck et al. (1995), have stated. GAs perform well for a global search but very poorly in a localized search. And since the performance of the RBF is strongly dependent on the optimization of the shape parameter, it is not surprising that this kind of interpolation works quite well in the Canadian dataset. Nevertheless, if these outcomes with the optimized shape parameter vector are compared to the RBF with a minimum distance shape parameter vector, the reader can immediately conclude that the optimization leads to an improvement of the interpolation for all networks and especially for Eastern Canada with an increase of almost 2 cm and for the Maritimes and Western Canada with an increase of more than 1.5 cm. Table 5.8 also shows the RMS values of the cross-validation as they are computed by *OrthoPlus* for the second-order polynomial model. Even in this comparison, the interpolation method of the RBF with optimized shape parameters outpaces the interpolation results of the second-order polynomial model and actually also any other kind of parametric model for all tested networks. Above all, this fact is again obvious in the Eastern Canadian network with an increase in accuracy of more than 2 cm from approximately. 5.0 cm to 2.7 cm.

Table 5.8: Comparison of RMS values from single-step scheme (with GA) and *OrthoPlus*

RMS values in [cm]	Canada	Western Canada	Central Canada	Eastern Canada	The Maritimes
RBF with optimized shape parameters	4.10	4.59	3.66	2.74	4.33
RBF minimum distance shape parameters	5.03	6.13	4.66	4.73	6.91
2nd order polynomial model (from <i>OrthoPlus</i>)	4.89	5.67	4.54	4.99	5.69

If the check points are included in the combined evaluation process (Table 5.9), the optimization of the shape parameters also leads to significantly improved RMS values, especially for Central Canada with an increase of 16.27 cm to 4.40 cm and for the Maritimes. For the last network, the combined evaluation yields an RMS value of 3.17 cm which is even smaller than the RMS value of 5.69 cm from the cross-validation procedure in *OrthoPlus*. This shows that the performance of the RBF in combination with the GA optimization is also strongly dependent on the choice of the evaluation function.

Table 5.9: Comparison of RMS values for combined evaluation (with GA)

RMS values in [cm]	Canada	Western Canada	Central Canada	Eastern Canada	The Maritimes
RBF with optimized shape parameters	11.11	10.21	4.40	4.94	3.17
RBF minimum distance shape parameters	12.96	14.84	16.27	7.06	10.34

Thus, the reader has to keep in mind that the RBF interpolation method together with the GA optimization of the shape parameters is not simple and straight forward. Some investigation prior to the application is necessary to find the best possible combination of evaluation function (single-step scheme, combined evaluation,...), shape parameter setup (single shape parameter or one shape parameter for every data centre) and GA operators (selection, cross-over and mutation type) corresponding to the dataset. On the other hand, if using any kind of parametric models, it is also crucial to find the best realization of a model as well as a suitable order to prevent over-parameterization.

Another major drawback of the optimized RBF interpolant is the relatively time-consuming computation in comparison to the parametric models. It depends on the user to find a trade-off and decide about the limit where an increase in interpolation performance is not any more redeeming the computational effort.

Finally, the following figures (Figure 5.11 - Figure 5.15) show the RBF models and their surface fit together with all benchmark points on a grid with a resolution of 0.5° in latitude and longitude for all five networks. The surface fit is based on the computed height misclosures on the grid points with the resulting model parameters plus the optimized shape parameters.

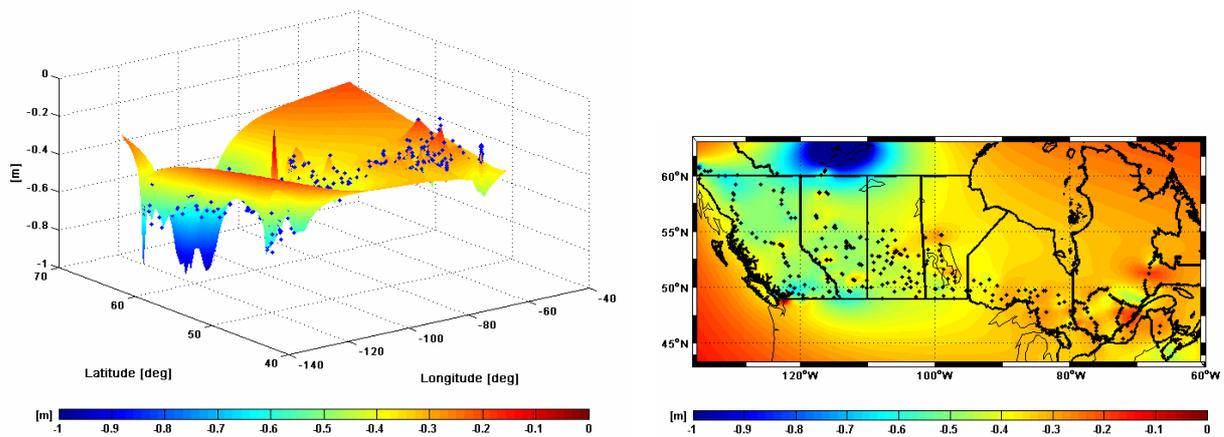


Figure 5.11: RBF model with GA optimized shape parameter and its surface fit to the Canadian network

Compared to section 4.3 which presents the corresponding plots to all kind of different parametric models interpolating the benchmark points of the Canadian dataset, Figure 5.11 visualizes that RBF models with GA optimized shape parameters are more flexible in adjusting rough surfaces than any other parametric model. The inherent features are less smooth and represent structures of the surface which are more adapted to the distribution of the benchmark points. The minimum peak in the North of Alberta as well as the maxima in the North of British Columbia and along its coast, for example, can not only be seen in the global network but also in the regional network of Western Canada in Figure 5.12. This proves that the application of the optimized RBFs is conform, no matter if applied to the global or regional network. It seems, however, that the performance of the RBF is worse if there is a lack of benchmark points on the edges of the networks. This effect can be seen in the Canadian dataset as well as in the five regional networks as opposed to the regional analysis of the parametric models (section 4.5). Maxima and minima mainly appear off-shore and in the North and South of the datasets. The only exception is Central Canada (Figure 5.13) where the benchmark points are distributed relatively regular all over the network as well as near the edges. The little peaks in between have significantly smaller magnitudes with about -80 cm than the peaks on the edges of the other networks. Thus, another drawback of the GA optimized RBF is the worse performance on the network edges whereas it has to be stressed that the strong dependence on the distribution of data points leads to very good interpolation results for regions surrounded by benchmark points.

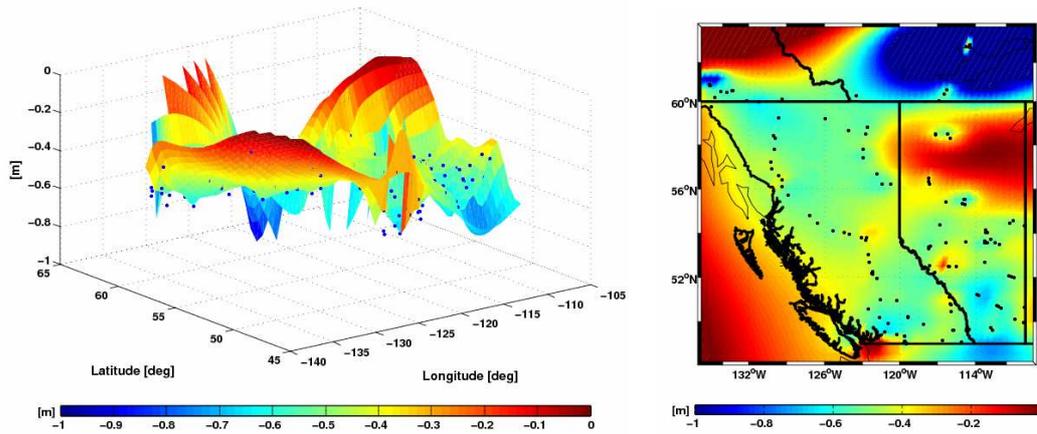


Figure 5.12: RBF model with GA optimized shape parameter and its surface fit to Western Canada

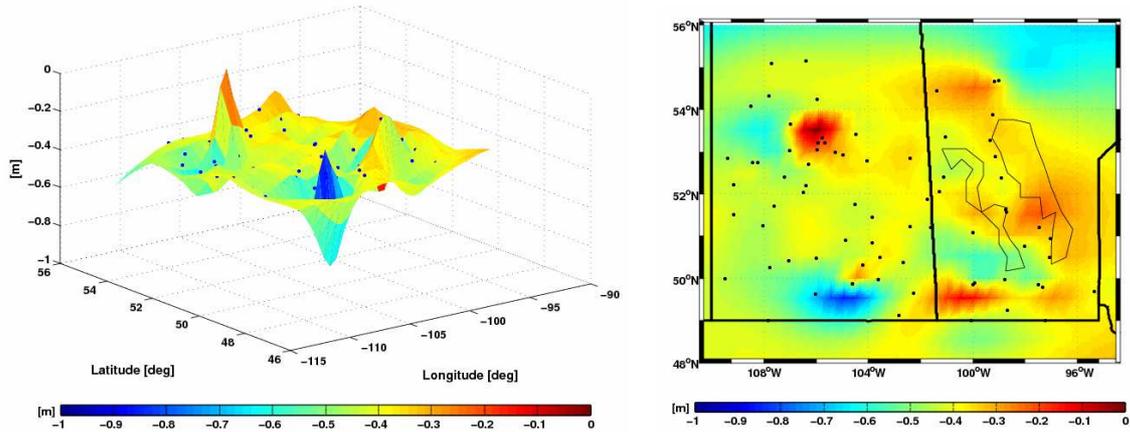


Figure 5.13: RBF model with GA optimized shape parameter and its surface fit to Central Canada

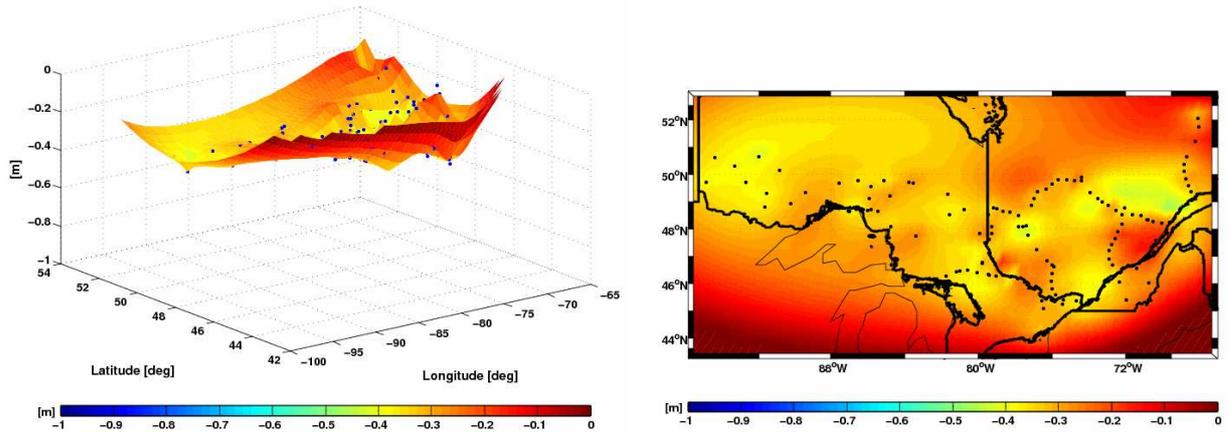


Figure 5.14: RBF model with GA optimized shape parameter and its surface fit to Eastern Canada

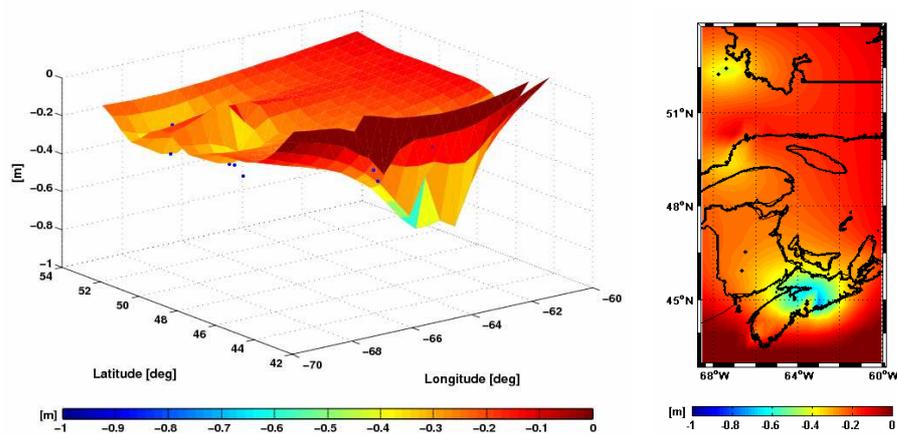


Figure 5.15: RBF model with GA optimized shape parameter and its surface fit to the Maritimes

As a conclusion, the user of interpolation methods always has to be aware of the advantages and disadvantages of optimized RBF on the one hand and parametric models on the other hand in order to adapt the application of one or the other to the expected outcome. It is not possible to simply decide which of the interpolation types is the best in general but it should be always thought about the most suitable solution to a specific problem.

6 Conclusions and future work

This final chapter summarizes the main contents and achievements of this study. Furthermore, some topics that deserve additional investigation with respect to GNSS-levelling are provided.

The main objective of this work was to find an optimal combination between ellipsoidal, orthometric and geoid heights. The discrepancies between the heterogeneous heights, mainly due to systematic errors and datum inconsistencies, were modelled by the incorporation of pre-selected parametric models in a combined height adjustment, on the one hand, and radial basis functions with GA optimized shape parameters, on the other. The new analysis tools implemented in the *OrthoPlus* software such as the model performance assessment, data screening and variance component estimation, allow for detailed tests with global and regional height data. All described approaches were evaluated using the most current geoid model, GPS and levelling data in Canada and its regions. The main outcome of this research is outlined in the sequel.

Evaluation of data screening In order to obtain reliable analysis results, an algorithm for the detection and localization of gross errors/blunders inherent in the dataset was implemented. As the Canadian vertical control network is expected to be ‘clean’, it is not surprising that no gross errors of large magnitude are detected by a routine 3-RMS test which screens the dataset prior to any network adjustment. The next step was the application of the post-adjustment data screening. All three implemented global tests, the two-tail and one-tail upper bound χ^2 -test as well as the F-test, yield the same result that the a-priori observation precision has been assumed wrongly. Therefore, a re-weighting of the a-priori covariance matrix is necessary. In order to also detect any possible gross errors with a smaller magnitude, every single residual was checked for outliers afterwards. The test shows that approximately two percent of the adjusted observations within the Canadian dataset have to be considered as outliers. The τ -test, which is a combination of global test and data snooping in the case that the a-priori precision of the observations is unknown, indicates all in all a slightly smaller percentage of outliers. Both tests flag those benchmark points containing outliers and maximum test

statistics as gross errors/blunders. Depending on the test and on the selected parametric model, mainly three different points situated in regions next to the coast are flagged as blunders in a first run of the post-adjustment data screening. As the tests are a-posteriori and the parametric surface is less supported by benchmark points on the edges, the adjusted model is probably introducing errors in these costal areas.

Assessment of parametric model performance Since the selection of an appropriate analytical model, i.e., the type and extent, is quite arbitrary, the performance of various types of parametric models was assessed thoroughly within the Canadian dataset by means of several empirical and statistical tests in the performance assessment. The tested models belong either to the pre-specified family of polynomial models up to 4th order or to the similarity-based transformation models and its simplified forms. The other two model types composed of orthogonal base functions, i.e., the Legendre polynomials and the Fourier series, are not included in the assessment as they are inappropriate due to their overmodelling effect. The first test in the assessment procedure was the comparison between the RMS value of the classical empirical approach and the RMS value of the cross-validation procedure. Both RMS values derived after the parametric surface fit are approximately 5 cm which are significantly reduced compared to the RMS of the misclosure vector before the fit with almost 41 cm. Above all the RMS of the classical empirical approach shows a trend of decreasing RMS value with increasing order and number of model parameters. Since this value indicates the precision of the observations only, the independent and more revealing measure is the RMS of the cross-validation which evaluates the model accuracy with respect to the model's performance in predicting new points. The latter one is also less sensitive to the model type. In a second step of the assessment, the coefficient of determination and the adjusted coefficient of determination are evaluated in order to determine the goodness of parametric model fit for a discrete set of points. For all model types, both measures take values of around 0.7 where the value 1 indicates best parametric model fit. The latter measure is always slightly smaller than the previous one with an increasing difference for an increasing number of model parameters since the adjusted coefficient of determination is corrected for the degrees of freedom. Testing the significance of each model parameter by backward elimination and the evaluation of the model stability are the last two steps in order to guarantee simplicity, computational efficiency and avoidance of over-parameterization. Executing this test on a model which was orthonormalized by a Gram-Schmidt orthonormalization before the adjustment leads to many more insignificant terms than testing the original model. Orthonormalization decorrelates the parameters and thus, the single parameters can be tested independently from each other. Otherwise, the test results seem to be distorted and overoptimistic due to the high correlations among the model parameters. Furthermore, orthonormalized models have the best possible numerical model stability with a condition number of

one. For the original models, however, the model stability is decreasing with increasing number of model terms. Applying the complete assessment procedure containing all tests, gives the user the possibility to check all kinds of different aspects and therefore should make the selection of an appropriate parametric model much easier. For the Canadian dataset, the choice would lie between a second-order polynomial or a classic 5-parameter model as they indicate the best test results.

Variance component estimation The purpose of the VCE is to test the statistical model of the combined height adjustment, to separate the random errors and to estimate the final variance components for ellipsoidal, orthometric and geoid heights. Latter can then be used as scaling factors in order to modify the initial covariance matrices of each height type. The evaluation of the initial CV matrices indicates that the levelling error component has the highest contribution to the overall accuracy and that there are relatively high correlations among the observations on the different benchmark points compared to the ellipsoidal and geoid height component. Both effects are due to the problems arising from the application of the spirit-levelling technique over a very large area including rough landscapes, like the Rocky Mountains, where the coverage of benchmark points is quite sparse and inhomogeneous. In order to ensure the reliability of the a-priori CV matrices, two different VCE algorithms are implemented in *OrthoPlus*. Using them iteratively, guarantees the globally best estimators. The first algorithm called IMINQUE does not lead to a convergent solution within the Canadian dataset and some final variance components become also negative. Consequently, the rigorous IMINQUE algorithm is not capable of improving the stochastic model. Nevertheless, it should be considered as an indicator for low redundancy in the dataset or for an incorrect stochastic model. Thus, the VCE is repeated with the alternative algorithm IAUE which constrains the final variance components to positive values and guarantees therefore a computationally more efficient, simpler and above all convergent solution. In our case, the final variance components of the IAUE show similar values for all parametric models. The variance components will be used as scaling factors for the a-posteriori CV matrices in order to obtain more realistic height accuracies. As they take values higher than one, the average of the standard deviation of all three down-scaled CV matrices will increase. Consequently, all three a-priori CV matrices have been overoptimistic. This is especially the case for the GPS-derived ellipsoidal heights since their final variance component is much higher than the other components and simultaneously their initial CV matrix is minimal because of the optimistic GPS post-processing outcome. Due to the difficulties of spirit-levelling over a large and inhomogeneous network, the scaled average standard deviation of the orthometric heights is, however, still the maximal contribution to the overall error budget. It is also shown that the geoid heights seem to be the most accurate height component in the Canadian network although their a-priori CV matrix is also too optimistic. Moreover, it can be concluded that neglecting significant correlations among the

heights of the same type by only applying the diagonal forms of the initial CV matrices in the VCE, does not only lead to a significant increase in the number of iterations, but will mainly result in smaller final variance components than applying the fully-populated form. Consequently, the average standard deviations will also be overoptimistic. The VCE is a powerful statistical tool in the combined height adjustment in order to obtain reliable accuracy measures and to determine the contribution of the ellipsoidal, orthometric and geoid heights to the overall error budget.

Regional analysis It was shown in the assessment of parametric model performance within the complete Canadian dataset that the determination of a most appropriate model is not very obvious for such a large network where the distribution of the benchmark points is relatively sparse and inhomogeneous. For a regional analysis, this Canadian network is divided into four smaller sub-regions which are characterized by changing geographical features from the West to the East coast. The performance assessment was applied to all regions. Obviously, the RMS of the height misclosure before the model fit depends very much on the data distribution and the topology. It becomes maximal with more than 50 cm for Western Canada with its rather sparse and inhomogeneous distribution of data points over rough landscapes and varying slopes and is minimal for the relatively good distribution of benchmark points on a smooth surface within Eastern Canada. This might be also the reason why in the assessment of the Western Canadian network, the third-order polynomial model is selected as best fitting model with a relatively high number of model parameters. For all the other regions, the classic four-parameter model with less parameters is deemed to be the most suitable one. Nevertheless, the initial RMS values are significantly reduced by the parametric model fit to an acceptable level of less than 5.5 cm for all regions. Furthermore, the difference between the two RMS values for precision and accuracy after the fit becomes evident. The first one computed by the classical-approach is typically decreasing for an increasing number of model parameters which is a visible effect of over-parameterization. The latter RMS defined by the cross-validation procedure, however, is increasing if models of higher orders and with more parameters are applied which makes it the more reliable measure.

It should be kept in mind that *OrthoPlus* is a semi-automated tool which is not meant to make its own autonomous selection of the most appropriate parametric model but rather just supports the user in the performance assessment of different models. It is the user's responsibility to find a trade-off between the various assessment results which is also dependent on the network type.

Radial basis functions and genetic algorithm The implementation of radial basis functions (RBF) instead of parametric models is an alternative approach in order to interpolate scattered surface data.

One advantage of the RBF over the parametric models is that the distribution of benchmark points is additionally taken into account. The performance and accuracy of the inverse multiquadric RBF as interpolation kernel is closely related to the global optimization of the shape parameters. Simple approximations of this parameter, such as the single shape parameter or the minimum distance shape parameter vector, lead to unacceptable performances with relatively high overall interpolation errors. Global minimization of these errors can be achieved by optimizing the shape parameter via a genetic algorithm. The performance of RBF without optimization with respect to accuracy was tested by four different evaluation criteria where the single-step scheme and the combined evaluation can additionally be used as fitness functions for the GA optimization. In comparison to the second-order polynomial model which serves as reference, the RBF with the minimum distance shape parameter vector achieves equivalent results even without optimization. The interpolation results of RBF with a single shape parameter are, however, significantly worse. Together with a GA optimized shape parameter, the interpolation method with RBFs outpaces the interpolation results of the second-order polynomial model and actually also any other kind of parametric models implemented in *OrthoPlus*. This can be recognized for Canada as well as for the other four tested regional networks. Since the GA is known to be a global optimization technique, it performs also well within the large dataset in Canada. Furthermore, optimized RBF interpolation is very flexible in adjusting rough surfaces and scattered data but shows problems in areas with only a few benchmark points, especially on network borders. It must be also considered that the optimization implies high computational effort which makes the interpolation with RBF much slower than with parametric models.

In conclusion, this study highlights the pros and cons of applying parametric surface models on the one hand and RBF models with GA optimized shape parameters on the other within a combined height adjustment in Canada and its regions. The test results depend on the quality of the data and are influenced by the different challenges in the networks. Accordingly, the implementation of a semi-automated algorithm in this study should rather support the user in the selection of a most appropriate model with respect to its performance for one specific dataset than providing one single general solution. The user is responsible for the correct interpretation of the test results and should therefore exactly know the challenges of the dataset used in order to fulfill most a-priori defined expectations.

In future, a move towards a modern vertical datum defined by a regional geoid model (for example in Canada) is expected which still requires the maintenance of some minimum-GPS levelling network for external accuracy assessment and monitoring changes over time. Within the framework of this new objective and as part of a future work, the software *OrthoPlus* might also be helpful in order to assist with the determination of the minimum GPS-levelling requirements. Furthermore, modules for an augmentation and improvement of a regional analysis of GNSS-levelling could, for example, include

the implementation of orthogonal polynomials of higher orders, the incorporation of augmented radial basis functions for a more local support, a method for a systematic determination of optimal GA specifications as well as an analysis of the eigenspectra of the design matrices in order to get better information about the numerical stability of the adjustment, among others. Another topic which needs further investigation with respect to the VCE, is the estimation of covariance additional to variance components whereas it should be noted that this will introduce more unknown parameters and reduce the data redundancy. Even an estimation of more than one final variance component for each height type can be considered in larger disconnected networks. Finally, the VCE procedure could be applied to calibrate new global geoid error models in compliance with the upcoming satellite-only global gravity field models such as GOCE.

References

- Beasley D, Bull DR and Martin RR (1993a): An overview of genetic algorithms: Part 1, Fundamentals. *University Computing*, Vol. 15, No. 2, pp. 58–69.
- Beasley D, Bull DR and Martin RR (1993b): An overview of genetic algorithms: Part 2, Research Topics. *University Computing*, Vol. 15, No. 4, pp. 170-181.
- Carlson RE and Foley TA (1991): The parameter R^2 in multiquadric interpolation. *Comput. Math. Appl.*, Vol. 21, pp. 29-42.
- Carroll JD and Green PE (1997): *Mathematical Tools for Applied Multivariate Analysis*. Revised Edition, Academic Press.
- Crain IK (1970): Computer interpolation and contouring of two-dimensional data: A review. *Geoexploration* 8, pp. 71-86.
- Darwin C (1859): *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. 1st edition, London: John Murray.
- Dermanis A and Rossikopoulos D (1991): Statistical Inference in Integrated Geodesy. Presented at the IUGG XXth General Assembly, International Association of Geodesy, Vienna, August 11-24.
- ESA (1999). Gravity Field and Steady-State Ocean Circulation Mission. Reports for Mission Selection. The four candidate Earth-explorer core missions, SP-1233(1).
- ESA, GOCE Mission Web Site, <http://www.esa.int/livingplanet/goce>
- Fotopoulos G (2003): An analysis on the optimal combination of geoid, orthometric and ellipsoidal height data. PhD Thesis, University of Calgary, Department of Geomatics Engineering, UCGE report No. 20185.
- Fotopoulos G and Sideris MG (2005): Spatial Modeling and Analysis of Adjusted Residuals over a Network of GPS-levelling Benchmarks. *Geomatica*, Vol. 59, No. 3, pp. 251-262.
- Franke R (1982): Scattered Data Interpolation: Tests of Some Method. *Mathematics of Computation*, Vol. 38, No. 157, pp. 181-200.
- Gao Y and Sideris MG (2005): Data Analysis in Engineering. University of Calgary, Department of Geomatics Engineering, Lecture notes to ENGO 563.
- Hardy RL (1971): Multiquadric equations of topography and other irregular surfaces. *J. Geophys. Res.*, Vol. 76, pp. 1905-1915.
- Hardy RL (1990): Theory and applications of the multiquadric-biharmonic method. *Computers and Mathematical Applications*, Vol. 19, pp. 163-208.

- Haupt RL and Haupt SE (2004): Practical Genetic Algorithms. 2nd Edition, Wiley-Interscience, John Wiley & Sons, Inc., Hoboken, New Jersey, USA, ISBN 0-471-45565-2.
- Hickernell FJ and Hon YC (1998): Radial basis function approximation of the surface wind field from scattered data. *Internat. J. Appl. Sci. Comput.*, Vol. 4, No. 3, pp. 221-247.
- Hofmann-Wellenhof B and Moritz H (2005): Physical Geodesy. SpringerWienNewYork.
- Huang J, Fotopoulos G, Cheng MK, Véronneau M and Sideris MG (2005): On the Estimation of the Regional Geoid Error in Canada, Dynamic Planet 2005 – Joint Assembly of the IAG, IAPSO and IABO, Cairns, Australia, August 22-26.
- Houck CR, Joines JA, and Kay MG (1995): A Genetic Algorithm for Function Optimization: A Matlab Implementation. NCSU-IE TR 95-09.
- IAG Comm.1, <http://iag.dgfi.badw.de/>
- Jekeli C (2000): Heights, the Geopotential, and Vertical Datums. Ohio State University, Geodetic Science and Surveying, Department of Civil and Environmental Engineering and Geodetic Science, Report No. 459.
- Koch KR (1999): Parameter Estimation and Hypothesis Testing in Linear Models, 2nd Edition, Springer-Verlag, Berlin Heidelberg, Germany.
- Kuang S (1996): Geodetic Network Analysis and Optimal Design: Concepts and Applications. Ann Arbor Press, Inc.
- Lancaster P and Salkauskas (1986): Curve and Surface Fitting: An Introduction, Academic Press Inc., London, England.
- Ling L (2003): Radial Basis Functions in Scientific Computing. PhD thesis, Simon Fraser University.
- Marti U and Schlatter A (2002): Höhenreferenzsysteme und –rahmen. *Vermessung, Photogrammetrie, Kulturtechnik*, Vol. 1, pp. 8-12.
- Mikhail E (1976): Observations and Least-Squares. Thomas Y. Crowell Company, Inc.
- Moore A and Lee MS (1994): Efficient Algorithms for Minimizing Cross Validation Error, Proceedings of the 11th International Conference on Machine Learning, pp. 190-198.
- Moritz H (1980): Geodetic Reference System 1980. *Bulletin Geodesique*, Vol. 54, pp. 395-405.
- Niemeier W (2002): Ausgleichsrechnung. Walter de Gruyter Verlag.
- Pope AJ (1976): The statistics of residuals and the detection of outliers. NOAA Technical Report NOS 65 NGS 1, U.S. Department of Commerce, Rockville, MD.
- Schneid S and Meichle H (2006): Normalhöhen in Baden-Württemberg. Arbeiten zur Einführung von Höhen im System des Deutschen Haupthöhennetzes 1992.

- Rippa S (1999): An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Advances in Computational Mathematics*, Vol. 11, pp. 193-210.
- Ruprecht D and Müller H (1995): Image warping with scattered data interpolation. *IEEE Computer Graphics and Applications*, Vol. 15, No. 2, pp. 37-43.
- Seeber G (1993): *Satellite Geodesy*. Walter de Gruyter Verlag.
- Sen A and Srivastava M (1990): *Regression Analysis: Theory, Methods, and Applications*. Springer Texts in Statistics, Springer, New York.
- Sharifi MA (2006): *Satellite to Satellite Tracking in the Space-wise Approach*. PhD thesis, Universität Stuttgart, Geodätisches Institut.
- Shul'min MV and Mitel'man YY (1974): The multiquadric method of approximating a topographic surface. *Geodesy Mapp. Photogramm*, Vol. 16, pp. 13-17.
- Suetin PK (1999): *Orthogonal Polynomials in Two Variables*. – (Analytical methods and special functions; vol. 3, Gordon and Breach Science Publishers.
- Torge W (1991): *Geodesy*. 2nd Edition, Walter de Gruyter, Berlin, New York.
- Véronneau M, Duval R and Huang J (2006): A Gravimetric Geoid Model as a Vertical Datum in Canada. *Geomatica*, Vol. 60, No. 2, pp. 165-172.
- Welsch W, Heunecke O, Kuhlmann H (2000): *Handbuch Ingenieurgeodäsie - Auswertung geodätischer Überwachungsmessungen*. Wichmann.