Improved Kriging Algorithms for Spatial Data Interpolation

Arsanchai Sukkuea

A Thesis Submitted in Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in

Environmental Management Technology

(International Program)

Prince of Songkla University

2022

Improved Kriging Algorithms for Spatial Data Interpolation

Arsanchai Sukkuea

A Thesis Submitted in Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in
Environmental Management Technology
(International Program)
Prince of Songkla University
2022

Thesis Title   Improved Kriging Algorithms for Spatial Data Interpolation

Author   Mr. Arsanchai Sukkuea

Major Program   Environmental Management Technology (International Program)

Major Advisor

…….…..………………….....…..
(Assoc.Prof.Dr. Apichat Heednacram)

Co-Advisor

…….…..…………………....….
(Dr. Rawee Rattanakom)

Co-Advisor

…….…..…………………....….
(Dr. Siriwan Ruamkeaw)

Examining Committee:

…..…..………………….…Chairperson
(Prof.Dr. Srilert Chotpantarat)

…..…..………………....…Committee
(Assoc.Prof.Dr. Apichat Heednacram)

…..…..………………....…Committee
(Dr. Rawee Rattanakom)

…..…..………………....…Committee
(Dr. Siriwan Ruamkeaw)

…..…..………………....…Committee
(Dr. Jinda Sawattawee)

       The Graduate School, Prince of Songkla University, has approved this thesis as fulfillment of the requirements for the Doctor of Philosophy Degree in Environmental Management Technology (International Program).

……………….………………………
(Prof.Dr. Damrangsak Faroongsarng)
Dean of the Graduate School

This is to certify that the work here submitted is the result of the candidate's own investigations. Due acknowledgement has been made of any assistance received.

………………..………………Signature

(Assoc.Prof.Dr. Apichat Heednacram)

Major Advisor

………………..………………Signature

(Mr. Arsanchai Sukkuea)

Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

…………….…………….Signature

(Mr. Arsanchai Sukkuea)

Candidate

ชื่อวิทยานิพนธ์	การปรับปรุงคริกกิ้งอัลกอริทึมสำหรับการประมาณช่วงข้อมูลเชิงพื้นที่
ผู้เขียน	นายอาสาฬห์ชัย สุขเกื้อ
สาขาวิชา	เทคโนโลยีการจัดการสิ่งแวดล้อม (หลักสูตรนานาชาติ)
ปีการศึกษา	2564

บทคัดย่อ

การสร้างภาพภูมิทัศน์มีความสำคัญในการวางแผนสิ่งแวดล้อม นักวางแผนด้านสิ่งแวดล้อมต้องการข้อมูลเชิงพื้นที่ที่แม่นยำเพื่อการตัดสินใจอย่างถูกต้องและเหมาะสม อย่างไรก็ตาม ข้อมูลดังกล่าวมักไม่พร้อมใช้งาน หาได้ยากและมีราคาสูง โดยเฉพาะอย่างยิ่งข้อมูลพื้นที่บริเวณภูเขาหรือบริเวณใต้ทะเลลึก นอกจากนี้ ข้อมูลสิ่งแวดล้อมที่รวบรวมจากการสำรวจภาคสนามมักได้มาจากการกำหนดจุดต่าง ๆ ในบริเวณที่กำหนด ดังนั้น ค่าของตำแหน่งจุดที่ยังไม่ทราบค่าจำเป็นต้องใช้วิธีการประมาณเพื่อสร้างข้อมูลต่อเนื่องเชิงพื้นที่ ในกรณีนี้นักวิจัยสามารถใช้วิธีการประมาณช่วงค่าเชิงพื้นที่เพื่อทำนายค่าความสูงของตำแหน่งจุดที่ไม่ได้เก็บตัวอย่างโดยใช้ข้อมูลจุดโดยรอบของตำแหน่งต่าง ๆ ที่สำรวจ

ในวิทยานิพนธ์นี้ ผู้วิจัยนำเสนออัลกอริทึมใหม่ 3 วิธี สำหรับวิธีการประมาณข้อมูลเชิงพื้นที่โดยใช้แบบจำลองคริกกิ้ง เนื่องจากการศึกษาที่ผ่านมายังไม่มีงานวิจัยที่แพร่หลายเกี่ยวกับการเลือกใช้คริกกิ้งพารามิเตอร์ในแบบจำลองเซมิแวริโอแกรมที่ส่งผลต่อประสิทธิภาพของการประมาณข้อมูลเชิงพื้นที่อย่างไร ดังนั้น แบบจำลอง 3 แบบจำลองของงานวิจัยจะถูกเปรียบเทียบกับแบบจำลองคริกกิ้ง ที่มีใช้กันอยู่ 5 แบบจำลอง และทำการประเมินประสิทธิภาพจากค่าความผิดพลาดน้อยที่สุดของทั้งแบบจำลอง 8 แบบจำลอง โดยจุดแข็งของแต่ละแบบจำลองจะได้รับการวิเคราะห์โดยพิจารณาจากชุดตัวอย่างที่มีตำแหน่งจุดต่าง ๆ ที่มาจากพื้นที่ศึกษาแตกต่างกัน จากการศึกษาจะเห็นว่าค่าความคลาดเคลื่อนที่เกิดจากวิธีการและแบบจำลองที่ผู้วิจัยเสนอนั้นมีค่าค่อนข้างน้อย โดยค่าขอบเขตล่างของช่วงความเชื่อมั่น 95% ของแบบจำลองของผู้วิจัยส่วนใหญ่ต่ำกว่าแบบจำลองที่มีอยู่เดิมทั้ง 5 แบบจำลอง อย่างไรก็ตามผลลัพธ์ไม่แสดงความแตกต่างอย่างมีนัยสำคัญระหว่างแบบจำลองต่าง ๆ ประโยชน์ของงานวิจัยชิ้นนี้จะช่วยให้มีวิธีการใหม่ ๆ ที่มีความแม่นยำมากขึ้น ส่งผลให้ผู้ใช้งานมีความเชื่อมั่นในการตัดสินใจและสามารถต่อยอดงานวิจัยให้ตรงกับความต้องการ สำหรับประเภทพื้นที่ที่แตกต่างกันและสามารถนำมาใช้จริงเพื่อปรับปรุงแผนผังพื้นผิว 3 มิติในการวางแผนสิ่งแวดล้อม

　　　　แม้ว่าอัลกอริทึมใหม่ 3 วิธีของผู้วิจัยจะเป็นแนวทางที่ถูกต้อง แต่ผู้วิจัยพบว่าการนำ
อัลกอริทึมมาใช้ในการประเมินภาพ 3 มิตินั้นยังมีอุปสรรค์ในการนำมาประยุกต์ใช้งานได้จริง
เนื่องจากระยะเวลาในการทำแผนผังพื้นผิวแบบจำลองที่ใช้เวลานานถึง 5 วัน อีกทั้งยังพบอุปสรรค์
ในการลดเวลาในการคำนวณให้น้อยกว่า 5 นาที โดยคงไว้ซึ่งความแม่นยำหรือความแม่นยำลดลง
เล็กน้อย วิทยานิพนธ์นี้ได้นำเสนออัลกอริทึมเพิ่มเติม 2 อัลกอริทึม โดยอัลกอริทึมแรกใช้เทคนิค
การแบ่งพื้นที่เพื่อแยกการคำนวณก่อน และอัลกอริทึมที่สองมีการนำพารามิเตอร์ความชัน (ความ
แปรผันของภูมิประเทศ) มาใช้เพื่อเลือกพื้นที่คำนวณ ผลการทดลองครั้งล่าสุด ผลลัพธ์สุดท้าย
สามารถลดเวลาในการรอจนเหลือเวลาเพียง 4 นาทีเท่านั้น

คำสำคัญ: คริกกิ้ง, การประมาณค่า, เชิงพื้นที่, ความสูง, การทำนาย

Thesis Title   Improved Kriging Algorithms for Spatial Data Interpolation
Author   Mr. Arsanchai Sukkuea
Major Program   Environmental Management Technology (International Program)
Academic   2021

ABSTRACT

Landscape visualization is important in environmental planning. Environmental planners need accurate spatially continuous data across an area to make competent and confident decisions. Obtaining such information can be difficult and costly, particularly in mountainous or deep-sea regions. In addition, environmental data gathered from field surveys are frequently derived from point sources. To generate spatially continuous data, the values of an attribute at unsampled points must therefore be estimated. In such cases, spatial interpolation techniques can be employed to predict the height values at unsampled sites using data from point observations.

In this thesis, we propose three novel algorithms for spatial interpolation methods using kriging models. Since there are not many findings of how kriging parameters in the semivariance model affect the performance of the spatial interpolators, we explore the parameters of the kriging algorithm and propose different semivariogram models to improve the performance of the spatial interpolation technique. Our three new models are compared with five contemporary kriging models. The performance is evaluated by error reduction that eight models can perform. The strengths of each model are analyzed based on a different set of sample sizes coming from two zones of study areas. The resulting errors of our proposed methods are relatively small. The lower bounds of the 95% confidence interval of our models are mostly lower than all five contemporary models. However, in general, the result shows no much significant differences among models. The benefits of this work are that it contributes to better accuracy resulting in more reliable decision making; supports different needs of algorithms for different area types, and can be practically used to improves the 3D surface plot in environmental planning.

Although our three new algorithms are accurate approaches, we found that applying them for landscape 3D visual assessment is not so practical as the waiting time to complete the model surface plot can be up to 5 days. Therefore, the second challenge is to reduce the computational time to be less than 5 minutes while preserving the accuracy or reducing it down marginally. The thesis presents two additional algorithms by first applying the divide and conquer technique and later improving it by introducing a slope (terrain variation) threshold parameter. The final result reduces the waiting time further down to 4 minutes.

Keywords: Kriging, Interpolation, Spatial, Height, Prediction

ACKNOWLEDGEMENT

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF FIGURES

LIST OF FUGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF ABBREVIATIONS

CV   Cross-Validation

CTS   Controlled Trend Surface

DEM   Digital Elevation Model

DGPS   Differential Global Positioning System

DVC   Divide and Conquer

EKO   Exponential with k-Iterations Optimizer

EPO   Exponential with Parameter Optimizer

ETL   Exponential with Polynomial-Trend Line

GIS   Geographic Information System

GP   Gaussian Process

GPS   Global Positioning System

IDW   Inverse Distance Weighting

LOOCV   Leave-One-Out Cross-Validation

MAE   Mean Absolute Error

ME   Mean Error

ML   Maximum Likelihood

MLE   Maximum Likelihood estimation

MLF   Multi Log Function

MLR   Multiple Regression Analysis

MLRK   Multiple Regression Kriging

MLP   Multilayer Perceptron

MLPK   Multilayer Perceptron Kriging

MPE   Mean of Percent Error

MQF   Multi Quadric radial basis Function

LIST OF ABBREVIATIONS (Continued)

MSE	Mean Square Error
OK	Ordinary Kriging
OCK	Ordinary Co Kriging
PSU	Prince of Songkla University
REML	Restricted Maximum Likelihood
RMSE	Root Mean Square Error
RTK	Real-Time Kinematic
SDVC	Slope Divide and Conquer
TNM	The Nation Map
UTM	Universal Transverse Mercator coordinate system

CHAPTER 1
INTRODUCTION

1.1 Introduction and Motivation

The spatial continuous data has an important role in environmental sciences and engineering. Environmental planners normally require spatially continuous data over the region of interest to make competent and confident decisions. Planers also need accurate and smooth continuous data across an area to make justified interpretations. These data are, however, usually not always readily available and often difficult and expensive to obtain. Furthermore, environmental data during field surveys are frequently collected from single points. As a result, in order to generate spatially continuous data, the values of an attribute at unsampled points must be estimated. Spatial interpolation methods, in such cases, provide a tool for predicting the values of an environmental variable at unsampled sites using data from point observations. Points interpolation can help save tremendous time and costs of collecting samples in every location (Al-Mamoori et al., 2021).

There are many methods for spatial interpolation. These include the Inverse Distance Weighting (IDW) (Aguilar et al., 2005; Mitas and Mitasova, 2005; Bello-Pineda and Hernandez-Stefanoni, 2007; Arun, 2013; Setianto and Triandini, 2013; Bărbulescu et al., 2021), the spline interpolation (Arun, 2013; Ajvazi and Czimber, 2019), the trend surface estimation (Liang and Zhoub, 2010; Thanoon, 2018), and the kriging (Lam, 1983; Aguilar et al., 2005; Bello-Pineda and Hernandez-Stefanoni, 2007; Li and Heap, 2011; Setianto and Triandini, 2013; Mert and Dag, 2017; Hasanipanah et al., 2021; Meng, 2021; Nie et al., 2021; Shi and Wang, 2021). If comparing these four spatial interpolation methods, the kriging technique is the best spatial interpolation method (Li and Heap, 2011). The kriging operates in several steps by combining surveys, analyzing statistical values of data, and semivariogram modeling. However, there are not many findings on how kriging parameters in

semivariance modeling can improve the performance of spatial interpolators. In this thesis, we will explore and automatically compute the parameters of the semivariogram model in kriging algorithms and propose three new algorithms for spatial height prediction based on kriging interpolation models. The aim is to increase the performance of our kriging interpolators in comparison with the other five conventional kriging interpolators existing in the literature. The results will appear in Chapter 3 – Three Novel Kriging Algorithms. Here are important contributions will be the three novel kriging models based on exponential semivariograms; the high accuracy results of the proposed models in comparison with the previous five contemporary methods; and the resilience of choosing different algorithms for different area types.

The second half of the thesis presents two more algorithms with the aim of reducing the computational time of our proposed algorithms. The first algorithm applies the divide and conquers technique, and the second improves the first one by introducing a slope threshold parameter. The first algorithm has six different kriging models to choose from. The experimental results in Chapter 4 – Two Divide and Conquer Algorithms will reveal that our four exponential-based models outperform the Gaussian and Spherical models in terms of accuracy. The two best performances of exponential-based models are used to enhance the second algorithm, which will reduce the waiting time down to 4 minutes but with a trade-off in model accuracy. The benefits of this research are that it contributes to substantially reducing the decision time and practically enhancing the speed of the calculation in generating the 3D surface plots in environmental management such as flood protection, transport planning, and farming management.

Moreover, Chapter 5 – Application will describe the application in environmental management, user interface, codes, and results. We developed a program to help us simulate all models studied in this research and implemented a JavaScript user interface for entering the input data. The results from this thesis can be used for 3D landscape visualizations in environmental management. Finally, In Chapter 6 – Conclusion will summarize our work and contributions to the thesis and suggest future works.

1.2 Objective

1.2.1   To design novel algorithms for spatial interpolation techniques based on the kriging model.

1.2.2   To improve the accuracy of the kriging interpolation with our technique.

1.2.3   To implement the proposed algorithm and compare the performance with existing methods.

1.2.4   To improve the computational time of our proposed method.

1.2.5   To apply the proposed algorithm to create a better 3D surface plot.

1.3 Scope of Research

1.3.1   This research will propose novel algorithms in the field of kriging interpolation models in comparison with five conventional kriging models.

1.3.2   The performance will be assessed by error reduction our models can perform.

1.3.3   The strength of each model will be analyzed based on a different set of inputs. The study area will be divided into zones. The proper algorithms will be selected differently for each zone to suit the elevation variance of the zone.

1.3.4   The computational time will be investigated. The proposed algorithms should execute with a practical waiting time.

1.3.5   The final scope aims at improving the 3D surface plot.

1.4 Benefit

1.4.1   There are new emerging algorithms in the fields of environmental sciences and management.

1.4.2   The estimation of unknown regions can be more accurate using our proposed kriging interpolation techniques.

1.4.3   The applications can be used to improve the 3D surface plot.

1.4.4   The 3D plots can be used in various fields such as flood protection, transport planning, and farming management

CHAPTER 2
LITERATURE REVIEW

This chapter describes the theory and concept used for this thesis. The technical background of spatial interpolation, semivariance, various semivariogram models, kriging interpolation, and maximum likelihood estimation will be discussed. The geotechnical and equipment errors, error assessment methods, related work on spatial interpolation, and related work on divide and conquer methods will be described.

2.1 Spatial Interpolation

The spatial interpolation is the process of using known point data to estimate other unknown point data. Data commonly used in spatial interpolation are elevation data, rainfall data, meteorological data, topography, and population density (Webster and Oliver, 2007). The estimation method relies on the weight relationship of the known points where the nearby sampled points have a high weight relation and the distant sampled points have a low weight relation. The estimated value at an unknown data point can be computed as follows.

$$z_p(s) = \sum_{i=1}^{n} w_i \cdot z(S_i) \tag{2.1}$$

The $z_p(s)$ is a value of an unknown data point for which we would like to predict. The $z(S_i)$ are the values of the neighbouring sampled points S_i where $i = 1, 2, \ldots, n$ points. Each of the sampled points will have the weight w_i assigned to them, where the total weight must be equal to 1 as per the equation below.

$$\sum_{i=1}^{n} w_i = 1 \tag{2.2}$$

2.2 Semivariance

The semivariance is calculated similarly to the variance estimator; namely, it is computed as half the variance of the differences between all sampled points. Semivariance measures the degree of spatial dependence and relationship among samples (e.g., elevation) based on the distance between S_i and S_{i+d}. A lower distance yields a lower semivariance and a higher distance leads to a higher semivariance.

$$\gamma(d) = \frac{1}{2n} \sum (z(S_i) - z(S_{i+d}))^2 \tag{2.3}$$

The $\gamma(d)$ is the semivariance, and the $z(S_i)$ are the values of the neighbouring sampled points S_i where $i = 1, 2, \ldots, n$ points.

2.3 Semivariogram Models

Semivariogram shows the spatial relationship of the data and must be chosen carefully. The rate of changing the distance between the points is described by the semivariance. There are several equations for computing the semivariance. Each equation requires the setting up of three parameters which consist of nugget, sill, and range. From Figure 2.1, given that the *x*-axis is the distance and the *y*-axis is the semivariance, the nugget is the value that intersects the *y*-axis, the sill is the plot *y*-value where the semivariance starts to be constant, and the range is the distance *x*-value where the semivariance starts to be constant.
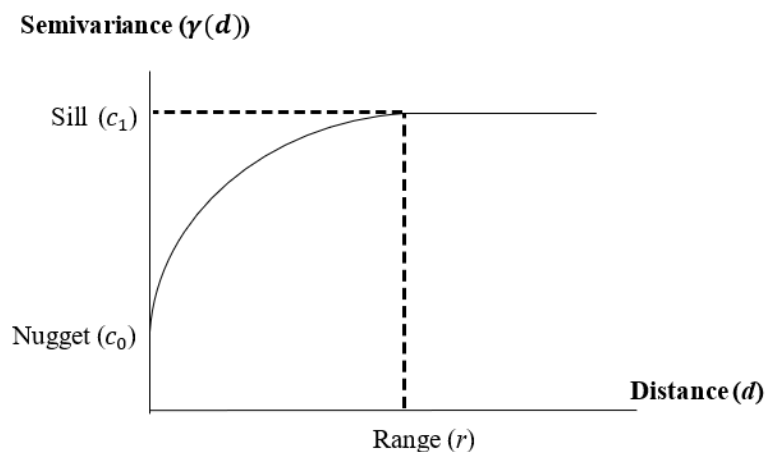
Figure 2.1 Semivariogram model showing nugget, sill and range.

Next, we will explain the common five semivariogram models that appeared in the literature (Webster and Oliver, 2007).

2.3.1 Linear Model

Linear interpolation is the method of curve fitting based on first-order polynomials or linear equations. The linear model takes little time to calculate and is effective for estimating the missing data from graphs with high linearity characteristics. The semivariogram model using the linear equation is shown below.

$$\gamma(d) = \begin{cases} c_0 + c_1 \left(\dfrac{d}{r} \right) & for\ d\ \leq\ r \\ c_0 + c_1 & for\ d\ >\ r \end{cases} \tag{2.4}$$

2.3.2 Spherical Model

Spherical model is one of the most frequently used models in geostatistics (Webster and Oliver, 2007). Semivariograms can be fitted for one, two, three, or more dimensions. The sphereical function in Equation (2.5) uses a cubic equation in a single dimension. The value of the semivariogram will increase rapidly at first, but when the distance equals the range, the semivariogram is parallel to the *x*-axis and constant.

$$\gamma(d) = \begin{cases} c_0 + c_1 \left(\dfrac{3d}{2r} - \dfrac{1}{2} \left(\dfrac{d}{r} \right)^3 \right) & for\ d\ \leq\ r \\ c_0 + c_1 & for\ d\ >\ r \end{cases} \tag{2.5}$$

2.3.3 Pentaspherical Model

Pentaspherical function can be thought of as the five-dimensional analogue of the circular and spherical models (Webster and Oliver, 2007). The model is useful in that the function curve in Equation (2.6) is somewhat more gradual than that of the spherical model.

$$\gamma(d) = \begin{cases} c_0 + c_1 \left(\frac{15d}{8r} - \frac{5}{4}\left(\frac{d}{r}\right)^3 + \frac{3}{8}\left(\frac{d}{r}\right)^5 \right) & for \; d \leq r \\ c_0 + c_1 & for \; d > r \end{cases} \tag{2.6}$$

2.3.4 Gaussian Model

If the variance is very smooth, the Gaussian model is normally used for the semivariogram. The semivariogram model using the Gaussian equation is given below.

$$\gamma(d) = \begin{cases} c_0 + c_1 \left(1 - exp\left(-\frac{d^2}{r^2} \right) \right) & for \; d \leq r \\ c_0 + c_1 & for \; d > r \end{cases} \tag{2.7}$$

2.3.5 Exponential Model

The exponential semivariogram model uses the exponential equation to fit the special data. The semivariogram model using the exponential equation is displayed below.

$$\gamma(d) = \begin{cases} c_0 + c_1 \left(1 - exp\left(-\frac{d}{r} \right) \right) & for \; d \leq r \\ c_0 + c_1 & for \; d > r \end{cases} \tag{2.8}$$

2.4 Kriging Interpolation

Kriging is named after D. G. Krige, the South African engineer who first developed this method (Lam, 1983). Kriging uses the semivariogram, in calculating estimates of the surface data at the unknown point. The kriging method selects mathematical equations that are appropriate to the selected sampled points or all sampled points within the specified radius, then calculates an estimated value at an unsampled point of the variable $z_p(s)$, based on the weighted average of neighbouring samples within a given region. Before the estimated value is computed by Equation (2.1), Ordinary kriging (Li and Heap, 2011) is applied to choose the optimal weights that produce the minimum estimation error. The optimal weights provide unbiased

estimates and have a minimum estimation variance. An example of three neighbouring sampled points will have the three associated weights (w_i for $i = 1, 2, 3$). These weights can then be computed by solving a set of simultaneous equations (Webster and Oliver, 2007), Equation (2.9) – Equation (2.11) using the inverse of a matrix as presented in Equation (2.13).

$$w_1\,\gamma(d_{11}) + w_2\gamma(d_{12}) + w_3\gamma(d_{13}) + \lambda = \gamma(d_{1p}) \tag{2.9}$$

$$w_1\,\gamma(d_{21}) + w_2\gamma(d_{22}) + w_3\gamma(d_{23}) + \lambda = \gamma(d_{2p}) \tag{2.10}$$

$$w_1\,\gamma(d_{31}) + w_2\gamma(d_{32}) + w_3\gamma(d_{33}) + \lambda = \gamma(d_{3p}) \tag{2.11}$$

And

$$w_1 + w_2 + w_3 = 1 \tag{2.12}$$

$$\begin{bmatrix} \gamma(d_{11}) & \gamma(d_{12}) & \gamma(d_{13}) & 1 \\ \gamma(d_{21}) & \gamma(d_{21}) & \gamma(d_{23}) & 1 \\ \gamma(d_{31}) & \gamma(d_{32}) & \gamma(d_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(d_{1p}) \\ \gamma(d_{2p}) \\ \gamma(d_{3p}) \\ 1 \end{bmatrix} \tag{2.13}$$

Note that $\gamma(d_{ij})$ is defined as the semivariance of the Euclidean distance between points i and j. A fourth variable in the weight vector of Equation (2.13) is called the Lagrange multiplier (λ) (Webster and Oliver, 2007). Once the individual weights are known, the estimation of the unknown value with three neighbouring sampled points is computed by the following equation.

$$z_p(s) = w_1 \cdot z(S_1) + w_2 \cdot z(S_2) + w_3 \cdot z(S_3) \tag{2.14}$$

The corresponding estimation variance of the above equation can be computed (O'Sullivan and Unwin, 2010) as follows.

$$\sigma_p^2 = w_1 \cdot \gamma(d_{1p}) + w_2 \cdot \gamma(d_{2p}) + w_3 \cdot \gamma(d_{3p}) + \lambda \tag{2.15}$$

2.5 Geostatistical Analysis Package (geoR)

The geoR (Diggle and Ribeiro, 2007) package performs geostatistical data analysis and spatial prediction, expanding the set of spatial data analysis methods and tools in R. It was created at Lancaster University's Department of Mathematics and Statistics in the United Kingdom. Some functions can be used at various stages of geostatistical data analysis. In this geoR, parameter estimation can be done in various ways. They are four methods for estimating the covariance model parameters implemented in geoR.

- Visual variogram fitting: using the function eyefit().

- Variogram fitting: using the function variofit() basically fitting a non-linear model to the empirical variogram.b

- Likelihood: using the function likfit(). The function likfit implements likelihood-based estimation, maximum likelihood (ML) and Restricted maximum likelihood (REML).

- Bayesian inference: using the function krige.bayes().

2.5.1 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation, often known as MLE, is a statistical method used to estimate the parameters of a presumed probability distribution based on specific observed data. This is performed by optimizing a likelihood function so that the assumed statistical model best fits the observed data, which creates the highest probability that the model is correct. The estimate with the highest possible likelihood is the maximum likelihood estimate, and it is located at the parameter space point that maximizes the likelihood function. Based on the presumption that most traditional kriging types (such as ordinary kriging, universal kriging, and others), random process interpolation may be viewed as an appropriate method for estimating values (David et al., 2008).

The maximum likelihood estimation (MLE) of model parameters is attained by optimizing the likelihood function L for the observed data $Y(s)$, provided by Equation 2.16. (Diggle and Ribeiro, 2007).

$$L(\beta, \tau^2, \sigma^2, \emptyset) \propto -\frac{1}{2}[\log|\Sigma| + (Y(s) - X\beta)'\Sigma^{-1}(Y(s) - X\beta)] \qquad (2.16)$$

Note that Σ represents the covariance matrix whose elements can be determined entirely by a semivariogram function with parameters τ^2, σ^2, and $\emptyset$ (nugget, sill, range). As well as the $\beta's$ in the component $X\beta$ are the model parameters.

2.5.2 Common semivariogram model in geoR

For calculating semivariogram models, geoR provides several alternatives. The semivariogram models include Matern, Exponential, Gaussian, Spherical, Circular, Cubic, Wave, Power, Powered Exponential, Cauchy, Gencauchy, and Gneiting, among others. In several of the studies in this thesis, geoR semivariogram models such as Matern, Exponential, Gaussian, and Spherical will be used to compare with our approaches. Equations (2.5), (2.7), and (2.8) illustrate the semivariogram model utilizing the Spherical, Gaussian, and Exponential, respectively. The Matern equation will be discussed in detail in the next section.

2.5.3 Matern semivariogram model

The Matern model is a class of semivariance models $\gamma(d)$ that emerges for different values of v (smoothing parameter). The Gaussian and Exponential semivariances are two frequently used members of the Matern class. In particular, the Exponential semivariance model is derived from the Matern class for $v = 0.5$. Also, when $v \to \infty$ then the Matern semivariance gives the Gaussian model. Given $T(v)$ is the Gamma function, and K_v is the modified Bessel function of the second kind, the Matern equation-based semivariogram model is shown below.

$$\gamma(d) = \begin{cases} c_0 + c_1\left(1 - \frac{2}{T(v)}\left(\frac{d\sqrt{v}}{r}\right)^v K_v(2\frac{d\sqrt{v}}{r})\right) & for\ d\ \leq\ r \\ c_0 + c_1 & for\ d\ >\ r \qquad (2.17) \end{cases}$$

2.6 Geotechnical and Equipment Errors

In the Cartesian coordinate system, the Universal Transverse Mercator (UTM) coordinate system is a grid-based geographic coordinate that represents the location of the earth's surface. This method is different from the typical method, which uses latitude and longitude coordinates. Within each of the 60 zones that make up the UTM grid system, we need to read the values of both the vertical grid (aligned with the east) and the flat grid (aligned with the north) in order for the two axes to be connected (Snyder and John, 1987).

The Global Positioning System, sometimes known as GPS, is a network-based system based on satellites that send signals from satellites to destination receivers to determine the position of a particular area. The Global Positioning System (GPS) transmits measurement signals using satellites that are precisely aware of their orbits. The signals will be broadcasted by satellites that only transmit in one direction, and receivers, as well as GPS receivers, will pick them up. The signal from the satellite will be translated into coordinates (X, Y, and Z), as well as time and speed. The GPS determines the precision of the location. The horizontal inaccuracy of the GPS is 10.305 meters, while the vertical error is 6.413 meters (Nazan and Semih, 2010).

Differential Global Positioning Systems (DGPS) is a satellite-based network system that employs two GPS satellites two to three hundred kilometres apart. The time on the satellite, the orbit, and the surrounding environment all have a degree of inaccuracy. The receiver picks up both satellite signals. Real-Time DGPS (Real-Time Differential Global Positioning System) is used to adjust position accuracy in real-time after the location survey, which determines precise coordinates and provides radio signal transmission services for the rover. The base and rover both receive accurate values measured in centimetres. To alter the rover, it will then use radio signals to connect. The DGPS error has several centimetre accuracies horizontally and 10-20 vertically (Gao et al., 2002).
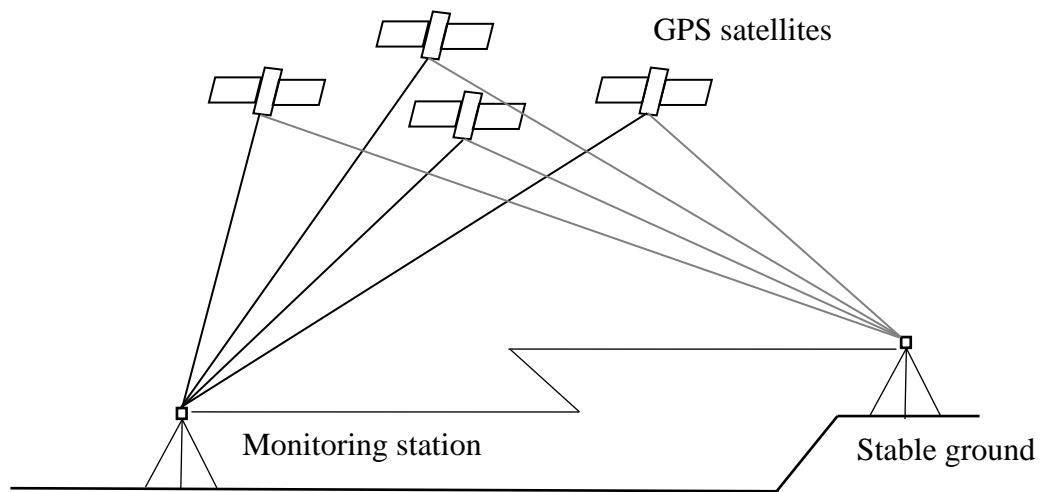
Figure 2.2 Differential global positioning system (DGPS).

To improve the accuracy of satellite-based positioning system data, real-time kinematic (RTK) positioning is a satellite navigation technique. Base Station and Rover Station are two types of receivers. Base Station is a stationary receiver, while Rover Station is a mobile one. Using a radio link, the Rover's location may be discovered and stored in a matter of seconds. RTK is used in applications that need a high degree of precision. The horizontal RTK inaccuracy is 9 millimetres, while the vertical RTK error is 2.2 centimetres (Ismat, 2017). However, the RTK is expensive and requires the installation of a base station with known coordinates.
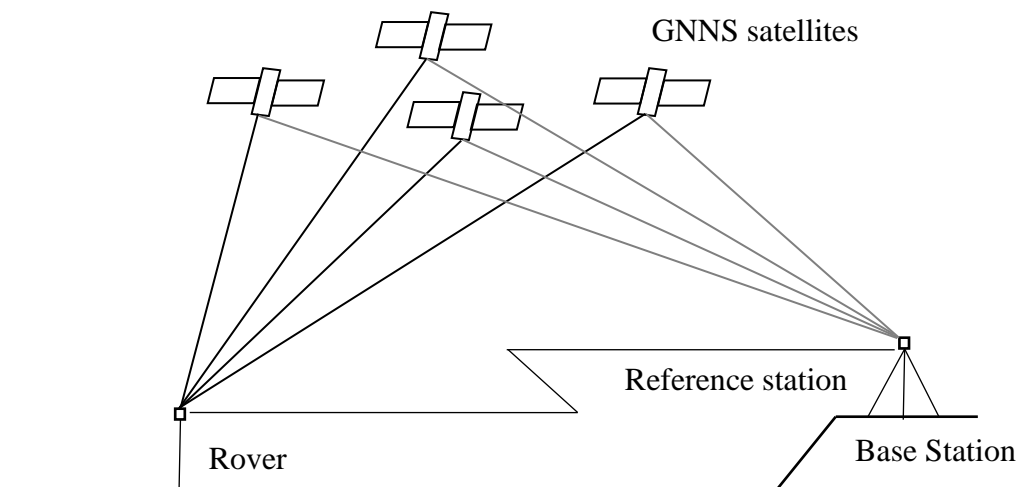
Figure 2.3 Real time kenematic (RTK).

A digital elevation model, often known as a DEM, is a depiction of the topographic surface of the earth that does not include vegetation, buildings, or any other surface features. DEM stands for digital elevation model and is a digital cartography dataset that is three-dimensional (XYZ) in format. It is produced using contour lines or photogrammetric techniques (Gandhi and Sarkar, 2016). There are various possible resolutions of the DEM, each of which has a unique level of accuracy. For instance, a DEM with a 30-meter resolution inaccurately is 16 meters horizontally and 20 meters vertically.

On the other hand, a 10-meter DEM error is 5 meters vertically (Kwanchai and Masataka, 2016). The DEM with a resolution of 1 meter is the highest-resolution standard DEM offered by the USGS. The DEM with a resolution of 1 meter has a horizontal accuracy of within 1 meter and a vertical accuracy of 19.6 cm at a confidence level of 95% (Arundel et al., 2015). For this thesis, we use the one-meter digital elevation model (DEM) input data, which may be acquired from the nation map (TNM) website of the United States Geological Survey.

Table 2.1 summarises six different types of equipment, along with a comparison of their errors. RTK has the most excellent accuracy of the six tools. But the RTK is costly and needs a coordinate-knowledgeable base station to function.

Table 2.1 Summary of equipment errors.

Equipment	Horizontal Error	Vertical Error	Reference
GPS	10.305 meters	6.413 meters	Nazan and Semih (2010)
DGPS	10-20 centimetres	10-20 centimetres	Gao et al. (2002)
RTK	9 millimetres	2.2 centimetres	Ismat (2017)
30-meter DEM	16 meters	20 meters	Gandhi and Sarkar (2016)
10-meter DEM	5 meters	5 meters	Kwanchai and Masataka (2016)
1-meter DEM	1 meter	19.6 centimetres	Arundel et al. (2015)

2.7 Measurement Error

The measurement errors can be separated into two distinct categories: systematic and random errors (Taylor, 1999). Systematic error is an error that, throughout several measurements performed under the same conditions of a given value and quantity. Systematic error is a type of measurement error. Random error is generated by changes in the readings of a measuring device or the experimenter's interpretation of those fundamentally unpredictable results. The measurements were repeated to prevent random error during the experiments, and the averages were taken as the final values. On certain occasions, human error is negligible (Carollo et al., 2007; Carollo et al., 2009; Azimi et al., 2019).

In addition, most studies in the literature review (Li and Heap, 2011) dismiss equipment inaccuracy and focus instead on model inaccuracy. On the other hand, Azimi et al. (2019) showed that the total uncertainty can be calculated by combining the experimental measurement error and model error. Therefore, this thesis will focus on model inaccuracy. For those interested in the complete uncertainty of our work, we recommend combining our model's inaccuracy with our measurements' inaccuracy.

2.8 Error Assessment Methods

The variety of quality measures of the fit between the actual data and the predicted data are listed in Equation (2.18) through Equation (2.22) (Webster and Oliver, 2007). The error is expressed in various forms of the mean differences between the predicted values p_i and the actual observed values o_i where $i = 1, 2, …, N$ prediction errors.

Mean Error (ME)

$$ME = \frac{1}{N} \sum_{i=1}^{N} (p_i - o_i) \tag{2.18}$$

Mean of Percent Error (MPE)

$$MPE = \frac{1}{N} \sum_{i=1}^{N} (p_i - o_i)/o_i \tag{2.19}$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |p_i - o_i| \tag{2.20}$$

Mean Square Error (MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (p_i - o_i)^2 \tag{2.21}$$

Root Mean Square Error (RMSE)

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} (p_i - o_i)^2 \right]^{1/2} \tag{2.22}$$

2.9 Related Work on Spatial Interpolation

2.9.1 Inverse Distance Weighting

The Inverse Distance Weighting (IDW) (Mitas and Mitasova, 2005) is the estimation method (weighted average) that assumes the distance of surrounding points inversely influences the elevation of a predicted point. A closer point has more importance or influence in determining the height of the predicted point more than a point that is far away. Mitas and Mitasova (2005) claimed that IDW is one of the simplest methods. Although this primary method is simple to perform and is available in nearly any GIS, it has some weaknesses. Aguila et al. (2005) highlighted the weakness of the IDW method with regard to the Multiquadric radial basis function (MQF) method and Multilog function (MLF) method. The weakness of IDW is because the interpolated altitudes are weighted averages that get values between the maximum and minimum points which decreases its effectiveness to estimate the highest or lowest levels when these levels are not in sampling points. As the terrain is steeper, MQF works better than IDW. In the mountainous area, the RMSE of IDW was 80.04, whereas the RMSE of the mentioned MQF was only 48.87.

Bello-Pineda and Hernández-Stefanoni (2007) compared the performances of IDW and the kriging method for generating a digital bathymetric model of the Yucatan submerged platform. The best RMSE of IDW was 187.25, whereas the best RMSE of the kriging was 153.21. Later, Setianto and Triandini 2013 developed the Digital Elevation Model (DEM) image for the area in Yogyakarta city. Two different methods, which are the kriging and IDW methods, were compared in this thesis. Again, the kriging method consistently outperformed the IDW for all levels of the factors mentioned in their article. For all the mentioned reasons, therefore, our thesis will focus on the kriging method rather than the IDW method.

2.9.2 Splines Method

The spline method represents two-dimensional curves on three-dimensional surfaces in a predictable manner. Using a mathematical function to fit a flexible surface over a collection of known points is comparable. From a limited

number of sample points, the spline may build highly precise and aesthetically pleasing surfaces. The output surface of the spline may have different minimum and maximum values than the input data set, are susceptible to outliers and lacks an error indication (Longley et al., 2005).

Laslett et al. (1987) analyzed and compared the performance of several spatial interpolation techniques, including kriging, IDW, spline, and others. On average, spline and kriging outperformed IDW, despite the fact that one technique may outperform the others in certain instances. Voltz and Webster (1990) compared kriging and spline for estimating soil characteristics and found that kriging performed better. Robinson and Metternicht (2006) explored spline, kriging, and IDW interpolation algorithms based on soil characteristics. They asserted that there was no universally applicable approach. Simpson and Wu (2014) investigated IDW, kriging, and spline for interpolating lake depth and found that spline produced the most accurate results with fewer sampled points than the ideal amount.

2.9.3 Trend Surface Estimation

A trend surface analysis is based on a regression function $f(x_i, y_i)$ that estimates the property value P_i at x_i and y_i coordinates. The trend surface estimation is given by $P_i = f(x_i, y_i)$. A trend surface estimation is a bivariate regression model with two independent variables, the x_i and y_i coordinates where a dependent variable is the P_i. If a linear regression function (first-order) is deployed, the modelled surface will correspond to a flat-oriented plane. When the spatial distribution is more complex, a polynomial (2nd, 3rd, …, or n-th order) can be applied, and the modelled surface will correspond to a curved surface with a growing number of curvatures (Mitas and Mitasova, 2005).

Liang and Zhou (2010) proposed a method of controlled trend surface (CTS) to account for large-scale spatial trends and non-spatial local effects concurrently. A geospatial model of forest dynamics was elaborated. The input was the Alaska boreal forest based on a large dataset which covers a wide range of forests. Two sets of validation plots representing temporal and spatial extensions were used to test the model. The results revealed that the CTS was more accurate than both the non-

spatial and traditional trend surface models. However, the limitation of this method is that the predictions were needed over areas of considerable sizes.

Thanoon (2018) studied the trend surface estimation with real spatial data in the context of rising groundwater levels. Three methods were compared. The first method estimated parameters by Maximum Likelihood. The second method used autocovariance estimates, and the third one used autocovariance function. The best method was proved to be the first one.

2.9.4 Kriging Method

The kriging method selects mathematical equations that are appropriate to the sampled points within the specified radius. It has a long history as it described in the early literature (Lam, 1983). In this classic technical review, spatial interpolation was categorized as point interpolation and areal interpolation. For point interpolation, the various methods were classified into exact and approximate. Kriging is a type of exact point interpolation along with other methods such as distance-weighting methods, spline interpolation, interpolating polynomials, and finite-difference methods. Kriging was mentioned to be the most distinctive of interpolation methods. Kriging has become an essential tool in geostatistics since the 1960s, and its applications to other fields are still increasing.

Simple kriging, ordinary kriging, universal kriging, and external trend kriging are all viable methods for performing kriging interpolation (Diggle and Ribeiro, 2007). Ordinary kriging is the most fundamental and extensively used kriging interpolation method. It estimates a variable at an unobserved location based on the weighted average of surrounding sites within a particular area. Standard kriging seeks to identify the optimal weights that provide the least estimated error. Solving a set of simultaneous equations produces optimal weights that produce fair estimates with a minor variance (Webster and Oliver, 2007).

As with conventional kriging, universal kriging, which is also a set of linear equations, is to forecast a point in an unsampled region. It separates the random function into a linear combination of deterministic functions, the smoothly evolving

and nonstationary trend, often called a drift, and a random component representing the residual random function (Wackernagel, 2003).

In this thesis, we use ordinary kriging to present three unique algorithms for spatial interpolation approaches based on kriging models. Many researches have been published in the field of kriging interpolation. Bello-Pineda and Hernández-Stefanoni (2007) showed that the exponential Kriging model produced the most accurate predictions, decreasing the error in 18.2% compared with IDW. This result has helped us to decide the type of interpolation technique and selection of the model for our work. Three semivariogram models were tested, namely, Spherical, Gaussian and Exponential models. The kriging parameters (nugget, sill and range) were fixed for each model. The Exponential model yielded the best output in terms of RMSE.

Although kriging is one of the most complex interpolators, Setianto and Triandini (2013) mentioned that for more informative data, kriging is preferable because it is more reliable. Kriging considers specific sampled points to obtain a value for spatial autocorrelation used for forecasting around that particular point rather than to allocate a universal distance power value.

The most sophisticated survey was conducted by Li and Heap (2011), where over 70 spatial interpolation methods were explored. The results revealed that ordinary co-kriging (OCK) was the most frequently used technique. The OCK was also listed as one of the most accurate methods in terms of RMAE. However, data variation is an influential impact factor and has essential effects on the accuracy of the methods. As the data variation rises, the efficiency reduces depending on the methods used.

Arun (2013) showed that the kriging method performed better than IDW, ANUDEM, Nearest Neighbor, and Spline approaches. RMSE values for three different terrain variations were compared. Kriging gave the minimal RMSE values in all three types of terrains which are mild slope area, steep slope area, and combined slope area.

In terms of software implementation, Mert and Dag (2017) developed a computer program (JeoStat) for Ordinary Kriging interpolation using Visual Basic. JeoStat can compute all main steps of basic geostatistical analysis, which are

semivariogram modelling, cross-validation and ordinary kriging. JeoStat offers seven semivariogram models, which are Spherical, Gaussian, Exponential, Linear, Generalized Linear, Hole Effect and Paddington Mix. Besides the software development, there were no new algorithms were presented. The thesis did not aim to increase the accuracy of the kriging interpolation, as our research does.

All well-known spatial interpolation methods were thoroughly surveyed and concluded in Table 2.2.

Table 2.2 Summary of literature survey on interpolation methods.

Year	Technique / Methodology	Assessment method	Value	Reference
	Inverse Distance Weighting			
2005	Interpolation on grid DEM accuracy	RMSE	8.80 (flat) 22.25 (rolling) 80.04 (mountainous)	Aguilar et al. (2005)
2007	IDW method for creating a digital bathymetric model of the Yucatan submerged platform	ME MAE RMSE	-12.33 42.56 187.92	Pineda and Stefanoni (2007)
2013	Analysis of different DEM interpolation IDW methods	RMSE	0.93 (mild slope) 1.45 (steep slope) 1.73 (combined slope)	Arun (2013)
2019	IDW analysis of different DEM interpolation in GIS	MAE RMSE	1.079 1.476	Ajvazi and Czimber (2019)
2021	IDW Interpolation Genetic Algorithm	MSE MAE	31.51 23.52	Bărbulescu et al. (2021)

Table 2.2 Summary of literature survey on interpolation methods (Continue).

	Spline			
1990	Comparison of kriging, cubic splines for predicting soil properties	RMSE		Voltz and Webster (1990)
2006	Spline interpolation techniques for mapping soil properties.	MSE		Robinson and Metternicht (2006)
2013	Analysis of different DEM interpolation Spline methods	RMSE	0.91 (mild slope) 1.37 (steep slope) 1.62 (combined slope)	Arun (2013)
2014	Spline interpolation and Accuracy			Simpson and Wu (2014)
2019	Spline analysis of different DEM interpolation in GIS	MAE RMSE	0.592 0.815	Ajvazi and Czimber (2019)
	Trend Surface Estimation			
2010	A geospatial model trend surface	RMSE	1.42 (conventional trend surface) 1.81(non-spatial) 3.88 (trend surface)	Liang and Zhoub (2010)
2018	Application trend surface models	MSE	1.9908 (method 1) 1.001 (method 2) 0.8921 (method 3)	Thanoon (2018)

Table 2.2 Summary of literature survey on interpolation methods (Continue).

	Kriging			
2007	Kriging method for creating a digital bathymetric model of the Yucatan submerged platform	ME MAE RMSE RI	-3.55 36.68 153.21 0	Pineda and Stefanoni (2007)
2011	Ordinary kriging in a review of comparative studies of spatial interpolation methods in environmental sciences	RMAE RRMSE	0.10-0.30 0.22-0.35	Li and Heap (2011)
2013	Analysis of different DEM interpolation kriging methods	RMSE	0.70 (mild slope) 1.31 (steep slope) 1.49 (combined slope)	Arun (2013)
2019	Kriging analysis of different DEM interpolation in GIS	MAE RMSE	0.602 0.838	Ajvazi and Czimber (2019)
2020	Kriging interpolation methods for spatial forest site index in pure beech forests: a case study from Turkey	RMSE	6.679 (MLR) 0.926 (MLRK) 2.968 (MLR) 2.326 (MLRK)	Günlü et al. (2019)

Table 2.2 Summary of literature survey on interpolation methods (Continue).

Year	Method	Metric	Value	Reference
2020	Kriging novel digital elevation model using an artificial neural network algorithm	RMSE	7.2	Behzadi and Jalilzadeh (2021)
2021	Analysis of groundwater table variability and trend using ordinary kriging: the case study of Sylhet, Bangladesh	RMSE	2.538	Hasan et al. (2021)
2021	Kriging Comparison of spatial interpolation methods for estimating the precipitation distribution in Portugal	RMSE	174.74	Antal et al. (2021)
2021	Kriging combined HASM and classical interpolation methods for DEM construction	RMSE	0.28-0.49	Wang et al. (2021)

Table 2.2 Summary of literature survey on interpolation methods (Continue).

2021	Kriging assessment of meteorology data in Mongolia using DEM and GIS	RMSE	6.16	Natsagdorj et al. (2021)
2021	Kriging for spot height DEM in Nigeria	RMSE	7.7944	Banjo et al. (2021)
2021	Non-parametric machine learning methods for interpolation of kriging (21CPT)	MAE MAPE	0.6 5.7	Shi and Wang (2021)
2021	Nonlinear models kriging interpolation	MAE RMSE	0.237 0.262	Hasanipanah el al. (2021)
2021	Soil salinity with geographically weighted regression kriging	ME MAE RMSE RI	-1.462 2.929 3.946 23.2	Nie el al. (2021)
2021	Spatial variation analysis of urban forest vegetation carbon storage kriging	RMSE	3 .8392 (cs) 0.8758 (gcs)	Ma et al. (2021)
2022	Ordinary kriging interpolation from hard soil depth	RMSE	5.22-5.75	Yanto et al. (2022)

2.10 Related Work on Divide and Conquer Methods

The divide and conquer algorithm solves a significant problem by dividing it into smaller sub-problems (Deng et al., 2017; Wang et al., 2020; Du and Deng, 2021). Figure 2.4 shows how the data is split into four sub-data. The sub-data are solved separately at this level before the subsolutions are merged into one complete answer.
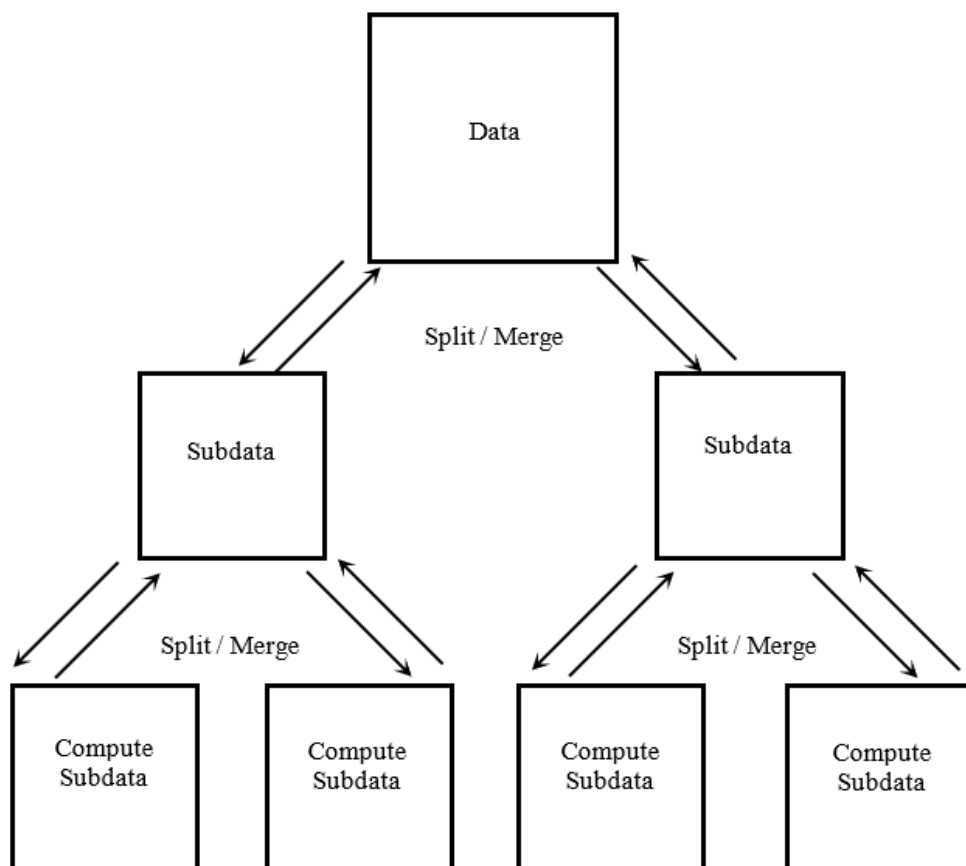
Figure 2.4 Divide and conquer algorithm.

The divide and conquer methods were thoroughly surveyed and concluded in Table 2.3.

Table 2.3 Summary of literature survey on divide and conquer technique.

Year	Technique	RMSE	Reference
2017	A divide-and-conquer method for space-time series prediction	30.8	Deng et al. (2017)
2020	SDC-depth: semantic divide-and-conquer network for monocular depth estimation	6.917	Wang et al. (2020)
2021	Unscented particle filter algorithm Based on divide-and-conquer sampling for target tracking	0.525 (x) 5.63 (y)	Du and Deng (2021)

CHAPTER 3
THREE NOVEL KRIGING ALGORITHMS

This chapter describes the three novel kriging algorithms based on the modified Exponential models. The methodology, study area, experimental results, discussion, and chapter summary will be explained inside the chapter.

3.1 Introduction

Spatially continuous data play a significant role in environmental sciences and environmental management. Environmental managers usually require spatially continuous data over the region of interest to make competent and confident decisions. Scientists also need accurate spatially continuous data across an area to make justified interpretations. Such data are, however, usually not always readily available and often difficult and expensive to acquire, especially for mountainous or deep marine regions. Moreover, environmental data collected from field surveys are often from point sources. Thus, the values of an attribute at unsampled points need to be estimated to generate spatially continuous data. In such instances, spatial interpolation methods provide a tool for predicting the values of an environmental variable at unsampled sites using data from point observations.

There are many methods for spatial interpolation (Cressie, 1993; Chilès and Delfiner, 1999; Wackernagel, 2003; Chilès and Desassis, 2018). These include the Inverse Distance Weighting (IDW) (Aguilar et al., 2005; Mitas and Mitasova, 2005; Bello-Pineda and Hernandez-Stefanoni, 2007; Arun, 2013; Setianto and Triandini, 2013; Bărbulescu et al., 2021), the spline interpolation (Arun, 2013; Ajvazi and Czimber, 2019), the trend surface estimation (Liang and Zhoub, 2010; Thanoon, 2018), and the kriging (Lam, 1983; Aguilar et al., 2005; Bello-Pineda and Hernandez-Stefanoni, 2007; Li and Heap, 2011; Setianto and Triandini, 2013; Mert and Dag, 2017; Hasanipanah et al., 2021; Meng, 2021; Nie et al., 2021; Shi and Wang, 2021). These well-known spatial interpolation methods were thoroughly surveyed in Table 2.2. If comparing these four spatial interpolation methods, the kriging technique is the best

spatial interpolation method (Li and Heap, 2011). The kriging operates in several steps by combining surveys, analyzing observed (point) data, and assuming a covariance-stationary Gaussian process (GP) with covariances estimated through semivariogram modeling. However, there are not many findings of how kriging parameters in semivariance modeling can improve the performance of spatial interpolators. In this thesis, we will explore and automatically compute the parameters of the semivariogram model in kriging algorithms and propose three new algorithms for spatial height prediction based on kriging interpolation models. The aim is to increase the performance of our kriging interpolators in comparison with the other five conventional kriging interpolators existing in the literature. Our important contributions are the three novel kriging models based on exponential semivariograms; the enhancement of conventional methods with the maximum likelihood estimation (MLE) method with four different correlation functions; and the robustness of choosing different algorithms for different area types. Our work can be used to improve the 3D surface plot in the application of environmental management.

3.2 Methodology

We propose three new algorithms in addition to the existing algorithms, which are in the field of kriging interpolation methods. Our unique algorithms do not require guessing important kriging parameters, namely, nuggets, sill, and range. We designed algorithms to compute the optimal values of kriging parameters automatically. Three variants of calculating the kriging parameters and weight vector for a predicted value at the unknown location are:

- Exponential with Parameter Optimizer (EPO) model,
- Exponential with k-Iterations Optimizer (EKO) model, and
- Exponential with Polynomial-Trend Line (ETL) model.

The three proposed algorithms start with acquiring the required input values. The input consists of the locations with the x and y coordinates called $L(x_i, y_i)$ and the altitudes $z(L(x_i, y_i))$ at the corresponding locations. After that, all three methods of the algorithms will configure the initial guess values for nuggets, sill, and range as the arrays of m sorted values. The algorithms create a semivariance matrix A from

distances between sampled points and generate a semivariance vector *b* from distances to a predicted point. All semivariances are computed based on the Exponential model in Equation (2.8). The exponential is chosen as the based model due to the positive results of previous research. (Mohd Aziz et al., 2019). They compare the performance of the Spherical, Gaussian, and Exponential models in their respective study areas. The findings indicate that the Exponential model provides the most excellent precision. In addition, we begin evaluating our data and find that the exponential model outperforms others.

After we obtain matrix *A* and vector *b*, we solve the weight vector using the equation $w = A^{-1}b$. Once we apply the weight vector, we will be able to find the predicted height $z_p(s)$ from the equation $\sum_{i=1}^{n} w_i \cdot z(S_i)$. After we get the value of $z_p(s)$, then we can calculate a predicted error σ_p^2. We aim to find the smallest amount of σ_p^2 because it indicates high accuracy. We vary the nugget, sill, and range values until the lowest error value is found within the conditions set in each algorithm.

The above process is called the Exponential with Parameter Optimizer (EPO) model. From this model, we develop an additional algorithm by adding iterations and narrowing down the nugget, sill, and range boundary values to find the minimal predicted error, we call this algorithm the Exponential with *k*-Iterations Optimizer (EKO) model.

In the third model, we improve the computational time by selecting a graph between distance and semivariance, which gives the least predicted error. We then compute the relation between the two variables in the form of the polynomial trend line equation: $a_3x^3 + a_2x^2 + a_1x + a_0$. Note that in this case, *x* becomes a distance that can give a *y*-value (a new semivariance). After that, we re-compute the semivariance matrix *A*, and vector *b* to solve the weight vector using the equation $w = A^{-1}b$ as before. We call this model, the Exponential with Polynomial-Trend Line (ETL) model. The detailed steps of each algorithm are explained in Figures 3.1 – 3.3.

Algorithm 1 Exponential with Parameter Optimizer (EPO)

Input: $L(x_i, y_i)$ and $z(S_i)$ /* locations and altitudes */

Output: $z_p(s)$ /* altitude of a predicted point */

Step 1: Build D from $L(x_i, y_i)$ /* D is a distance matrix */

Step 2: Set initial values for nugget, sill, and range, each with m sorted values /* m is the length of the array */

Step 3: Use D to find a semivariance matrix A

Step 4: Build a semivariance vector b from predicted points

Step 5: Solve a weight vector w where $w = A^{-1}b$

Step 6: Compute $z_p(s) = \sum_{i=1}^{n} w_i \cdot z(S_i)$

Step 7: Compute a predicted error σ_p^2

Step 8: Repeat Steps 3 – 7 until m^3 values of σ_p^2 are obtained

Step 9: Select $z_p(s)$ with minimum σ_p^2

Figure 3.1 Exponential with Parameter Optimizer (EPO) algorithm.

Algorithm 2 Exponential with k-Iterations Optimizer (EKO)

Input: $L(x_i, y_i)$ and $z(S_i)$ /* locations and altitudes */

Output: $z_p(s)$ /* altitude of a predicted point */

Step 1-7: as per Step 1-7 of Algorithm 1

Step 8: Repeat Step 3 – 7 until m^3 values of σ_p^2 are obtained

Step 9: Compute $\sigma_j^2 = \min(\sigma_1^2, \sigma_2^2, \dots, \sigma_{m*m*m}^2)$

Step 10: Select the nugget, sill, and range with σ_j^2

Step 11: Iteratively reduce m to $m_1, m_2, \dots, m_k$ and repeat Step 2 – 9 of Algorithm 1

Step 12: Select $z_p(s)$ with minimum σ_j^2

Figure 3.2 Exponential with k-Iterations Optimizer (EKO) algorithm.

Algorithm 3 Exponential with Polynomial Trend Line (ETL)

Input: $L(x_i, y_i)$ and $z(S_i)$ /* locations and altitudes */

Output: $z_p(s)$ /* altitude of a predicted point */

Step 1-10: as per Step 1-10 of Algorithm 2

Step 11: Construct the semivariogram model with nugget, sill, and range of σ_j^2

Step 12: Create a trend line equation: $a_3x^3 + a_2x^2 + a_1x + a_0$

Step 13: Apply the trend line to re-compute matrix A

Step 14: Apply the trend line to re-compute vector b

Step 15: Solve a weight vector w where $w = A^{-1}b$

Step 16: Re-compute $z_p(s) = \sum_{i=1}^{n} w_i \cdot z(S_i)$

Figure 3.3 Exponential with Polynomial-Trend Line (ETL) algorithm.

3.3 Study Areas and Experimental Results

Prince of Songkla University (PSU), Phuket Campus is an educational institution located in the area of Kathu District, Phuket Province, Thailand. The campus is located in the East longitude of 100°25′27″ – 100°32′58″ and North latitude of 13°42′30″ – 13°47′42″. We have chosen two study areas, as displayed in Figure 3.4. The aim of the experiments is to compare and validate the accuracy of the contemporary methods with the new interpolation algorithms that we proposed. We begin the experiment with the two zones, each having an input of 31 points. The input data are the DGPS data of the x and y coordinates in UTM coordinate systems and their altitudes in meters (m). The DGPS error is several centimetre accuracies horizontally and 10-20 centimetres vertically (Gao et al., 2002). All input data are collected by the same DPGS. We have verified and cross-checked the data with the RTK (Real-Time Kinematic), and then we apply the input data to the models. The output data are the predicted altitudes, which is calculated by using eight models, namely, the Linear, Spherical, Pentaspherical, Gaussian, and (unmodified) Exponential, EPO, EKO, and ETL models. Note that the latter three models are our proposed models.

The results of the first four models (Linear, Spherical, Pentaspherical, and Gaussian) are based on Equation (2.4) – Equation (2.7). The three parameters (nugget, sill, and range) are chosen by testing values from a number of empirical semivariogram plots, and the values giving the minimum error are used. The results of the latter four models (Exponential, EPO, EKO, and ETL) are based on Equation (2.8). The conventional Exponential model uses the fixed parameters with nugget = 0.1, sill = 0.2, and range = 250. On the other hand, the EPO, EKO, and ETL do not use the pre-computed parameters. The best parameter values are computed automatically based on algorithms presented earlier in Section 3.2. We understand that finding the optimal parameters would lead to better results than using the fixed ones. However, we include the conventional Exponential model with fixed parameters in the table of results only for a reference to see how much the three new exponential-based approaches without the fixed parameters have improved.

We will divide our studies into four experiments according to two different study areas and two different sets of inputs; 31 points and 51 points. The first two experiments use the input of 31 points, and by exploiting the LOOCV technique, 30 points will be served as a training set in each run. In the latter two experiments, we add 20 more input points into each zone, and 50 points will be served as a training set in each run. The characteristics of Zone 1 and Zone 2 are shown in Table 3.1. Notice that they are not the same since the sample standard deviation of Zone 1 is larger than the one of Zone 2. The range of Zone 1 is also wider than Zone 2. We select these areas because we want to see how our algorithms perform given different types of terrain variations. Moreover, when increasing the number of sampled points, the margin of error is reduced. Therefore, we will also explore how the performance of the algorithms is affected, given the larger sample size.
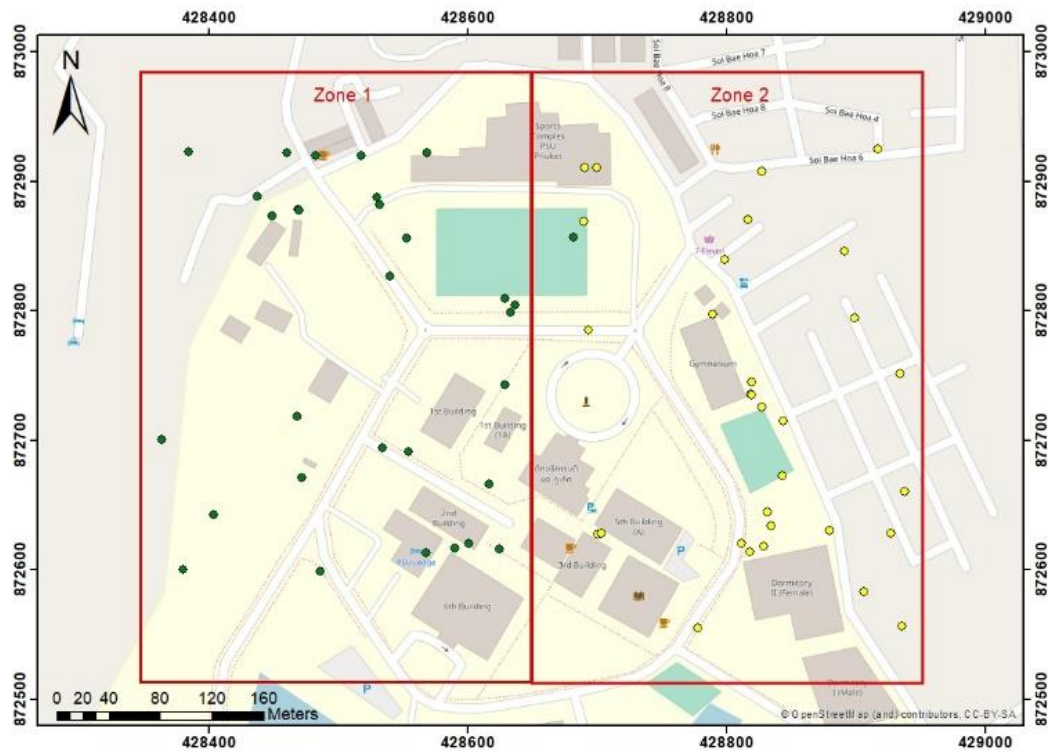
Figure 3.4 Sampled 31 points for each of two zones in PSU Phuket area.

Table 3.1 Characteristics of height-samples in zone 1 and zone 2.

Area	Zone 1 (31 points)	Zone 2 (31 points)	Zone 1 (51 points)	Zone 2 (51 points)
Sample Size	31	31	51	51
Average	34.6519	35.0273	34.7104	34.5571
Standard deviation	10.1745	4.0759	10.5448	3.9841
95% Margin of error	3.5817	1.4348	2.8941	1.0935
Upper bound	38.2336	36.4621	37.6044	35.6506
Lower bound	31.0702	33.5925	31.8163	33.4637
Max	65.2595	45.5822	74.8654	45.5822
Min	25.1687	26.8065	25.1687	26.8065
Range	40.0908	18.7757	49.6967	18.7757

*The unit of input data (meters).

3.3.1 Study area of zone 1 with 31 points

The accuracies of eight models are presented in Table 3.2 in the form of five error assessment methods discussed in Section 2.8. The values of RMSE for our methods, namely EPO, EKO, and ETL range between 3.7798 – 3.9749, the three lowest errors in the table. Other values of ME, MPE, MAE, and MSE yield similar results. We discuss these results further in the discussion section.

Table 3.2 Error assessment for the study area of zone 1 with 31 points.

Model	ME	MPE	MAE	MSE	RMSE
Linear	-6.2581	-0.2015	7.6397	72.2372	8.4992
Spherical	-6.9418	-0.2254	8.0175	77.5022	8.8035
Pentaspherical	-6.0399	-0.1957	7.4041	73.6677	8.5830
Gaussian	-2.1666	-0.0551	5.4422	44.9070	6.7013
Exponential	-0.2770	0.0191	3.4629	34.9012	5.9077
EPO	-0.2722	0.0072	2.1451	15.7998	3.9749
EKO	-1.5773	-0.0393	2.6633	14.2872	3.7798
ETL	-0.2722	0.0072	2.1451	15.7998	3.9749

*The unit of input data (meters).

3.3.2 Study area of zone 2 with 31 points

Table 3.3 shows the errors for the study area of Zone 2 with the same number of sampled points as the previous experiment. However, the terrain here is significantly flatter. RMSEs for EPO, EKO, and ETL models vary from 2.1499 – 2.2218, the three lowest errors in the table, and also lower than Zone 1. ME and MPE of our models are also small (ME = -0.0965 to -0.0626 and MPE = -0.0005 to 0.0041). The small negative MPE indicates that our predicted values are slightly lower than the actual values. However, it is not conclusive just by looking at ME or MPE. The other three types of errors (MAE, MSE, and RMSE) should also be examined together because in ME and MPE positive errors may compensate for negative errors. The discussion section will elaborate more on this result.

Table 3.3 Error assessment for the study area of zone 2 with 31 points.

Model	ME	MPE	MAE	MSE	RMSE
Linear	0.7138	0.0187	2.1991	8.4232	2.9023
Spherical	-0.2156	-0.0119	2.5141	13.6125	3.6895
Pentaspherical	-0.9810	-0.0359	3.3372	20.2478	4.4998
Gaussian	1.1484	0.0384	2.3909	8.6129	2.9348
Exponential	-0.2719	-0.0085	1.4860	6.1048	2.4708
EPO	-0.0626	0.0041	1.5041	4.9366	2.2218
EKO	-0.0965	-0.0005	1.1003	4.6223	2.1499
ETL	-0.0626	0.0041	1.5041	4.9366	2.2218

*The unit of input data (meters).

3.3.3 Study area of zone 1 with 51 points

In this experiment and the next experiment, we increase the number of sampled points to 51 points. Table 3.4 shows the errors for the study area of Zone 1. RMSEs for EPO, EKO, and ETL models are between 2.5704 – 3.1627, still the three lowest errors in the table. The impact of more sampled points may not be obvious in the table, but we will clarify this result in the discussion section.

Table 3.4 Error assessment for the study area of zone 1 with 51 points.

Model	ME	MPE	MAE	MSE	RMSE
Linear	-7.0240	-0.2360	8.4045	89.0390	9.4360
Spherical	-7.4717	-0.2580	9.3936	100.4659	10.0233
Pentaspherical	-6.0176	-0.2133	8.8407	100.2174	10.0109
Gaussian	-2.8717	-0.0814	5.7993	49.6288	7.0448
Exponential	-0.0982	0.0150	2.4538	20.1328	4.4870
EPO	-1.8730	-0.0618	2.6063	10.0024	3.1627
EKO	-0.0937	0.0043	1.2424	6.6071	2.5704
ETL	-1.8730	-0.0618	2.6063	10.0024	3.1627

3.3.4 Study area of zone 2 with 51 points

Table 3.5 shows the errors for the study area of Zone 2 with 51 points. The terrain in this zone is significantly flatter than in Zone 1. RMSEs for EPO, EKO, and ETL models are lower than in the previous three experiments ranging from 1.4353 – 1.6889.

Table 3.5 Error assessment for the study area of zone 2 with 51 points.

Model	ME	MPE	MAE	MSE	RMSE
Linear	0.6881	0.0172	1.7842	5.1597	2.2715
Spherical	-0.4577	-0.0206	2.2751	10.0405	3.1687
Pentaspherical	-1.3165	-0.0480	3.3137	17.0352	4.1274
Gaussian	1.2804	0.0422	2.2348	7.1474	2.6735
Exponential	-0.3151	-0.0110	1.1053	2.9224	1.7095
EPO	-0.0467	0.0024	1.1154	2.8525	1.6889
EKO	-0.0748	-0.0008	0.7325	2.0600	1.4353
ETL	-0.0467	0.0024	1.1154	2.8525	1.6889

*The unit of input data (meters).

3.3.5 Quality control of model evaluation

In this section, we will find a 95% confidence interval ($\alpha = 0.05$) for the previously calculated errors. However, RMSE is an estimate of the actual standard deviation or dispersion of σ. Thus, we want to estimate the confidence interval of σ, not RMSE. MSE is a square of RMSE, an estimate of σ^2. To get SE, we multiply MSE by the number of samples N. The ratio of SE and σ^2 theoretically follows Chi-square distribution with $N - 1$ df depicted by Equation (3.21).

$$\frac{SE}{\chi^2_{\alpha/2,N-1}} \leq \sigma^2 \leq \frac{SE}{\chi^2_{1-\alpha/2,N-1}}$$

(3.21)

The 95% confidence interval of σ with already known RMSE is therefore given by Equation (3.22).

$$\frac{\sqrt{N} \cdot RMSE}{\sqrt{\chi^2_{0.025,N-1}}} \leq \sigma \leq \frac{\sqrt{N} \cdot RMSE}{\sqrt{\chi^2_{0.975,N-1}}} \tag{3.22}$$

Table 3.6 shows the quality of the 95% confidence interval (CI) of our algorithms in comparison with the five benchmark algorithms for all study areas. We will interpret these results in the next section.

Table 3.6 The 95% confidence interval (lower-bound and upper-bound) of σ for all areas.

Model	Zone 1 (31 points)	Zone 2 (31 points)	Zone 1 (51 points)	Zone 2 (51 points)
Linear	6.9040-11.5487	2.3576-3.9436	7.9738-11.8459	1.9195-2.8516
Spherical	7.1512-11.9622	2.9970-5.0133	8.4700-12.5832	2.6777-3.9780
Pentaspherical	6.9721-11.6626	3.6553-6.1143	8.4596-12.5676	3.4878-5.1815
Gaussian	5.4436-9.1057	2.3840-3.9878	5.9531-8.8440	2.2592-3.3563
Exponential	4.7989-8.0274	2.0071-3.3573	3.7917-5.6330	1.4446-2.1461
EPO	3.2289-5.4011	1.8048-3.0190	2.6726-3.9704	1.4272-2.1202
EKO	3.0704-5.1360	1.7464-2.9213	2.1721-3.2269	1.2129-1.8019
ETL	3.2289-5.4011	1.8048-3.0190	2.6726-3.9704	1.4272-2.1202

3.3.6 Improved conventional methods with MLE

In the results of earlier experiments, we have investigated the prediction caused by a different type of semivaiogram models. However, there are more techniques for estimating the kriging parameters, for example, using the maximum likelihood estimation (MLE) approach to numerically solve the likelihood function and estimate the best value for the parameters. In this section, we will investigate the performance of conventional methods that use MLE. The tools are available in package geoR (Diggle and Ribeiro, 2007), which offers various models for the correlation

function such as Matern, Exponential, Gaussian, and Spherical. Kriging can be performed with options for simple kriging, ordinary kriging, universal kriging, and external trend kriging. We will use ordinary kriging in our experiment. Table 3.7 compares the performance of our three algorithms with the MLE method using four various correlation functions, namely, Matern, Exponential, Gaussian, and Spherical, which we call MLE_matern, MLE_exponential, MLE_gaussian, and MLE_spherical, respectively. The values of RMSE for all seven methods show non-significant differences. Other errors such as ME, MPE, MAE, and MSE are omitted here as their influences on the result discussion are similar to the previous four experiments.

Table 3.7 Comparison of the RMSE values for the four different MLE methods and our proposed models.

Model	Zone 1 (31 points)	Zone 2 (31 points)	Zone 1 (51 points)	Zone 2 (51 points)
MLE_matern	3.8035	2.1605	2.6005	1.5161
MLE_exponential	3.8035	2.1605	2.5946	1.4809
MLE_gaussian	3.7302	2.3680	2.5421	1.4181
MLE_spherical	3.7919	2.0921	3.1977	1.5268
EPO	3.9749	2.2218	3.1627	1.6889
EKO	3.7798	2.1499	2.5704	1.4353
ETL	3.9749	2.2218	3.1627	1.6889

*The unit of input data (meters).

3.4 Discussion

We will divide the discussion into 3 parts. Firstly, we will discuss the results of the prediction errors caused by different types of semivariograms. Secondly, we will compare the prediction errors caused by different correlation functions based on MLE. Finally, we will discuss the practical application in environmental management.

3.4.1 Different types of semivariograms

We have studied two areas at Prince of Songkla University, Phuket Campus. We compared and verified the performance of the original five methods with our three new methods. For the first area (Zone 1), we concluded that the errors of the original methods, namely, Linear, Spherical, Pentaspherical, Gaussian, and classical Exponential models, are higher than our proposed EPO, EKO, and ETL methods in almost all cases.

In the first experiment, the RMSE of the existing techniques were between 5.9077 and 8.8035, but the RMSE of our new methods were only between 3.7798 and 3.9749. The results of MSE follow the same direction as RMSE does because MSE is just the square of RMSE. However, RMSE is widely used because it is easy to interpret as the error is in the same unit as the predicted output. In addition, MAE is less sensitive to outliers, and it can be used together in the interpretation. Table 3.2 shows that our models have the lowest MAEs when compared to the other five conventional models. ME and MPE are one of the simplest forms but beware that positive errors may compensate for negative errors, and therefore a careful interpretation must be made. Nevertheless, they are useful in indicating whether the forecasts tend to be disproportionately negative or positive. In our case, they are more negative meaning our predicted values are mostly lower than the actual values. Note that our new methods are based on the Exponential model and in some cases when assessing with ME and MPE, the classical Exponential model exhibits a good performance as well (ME = -0.2770, MPE = 0.0191).
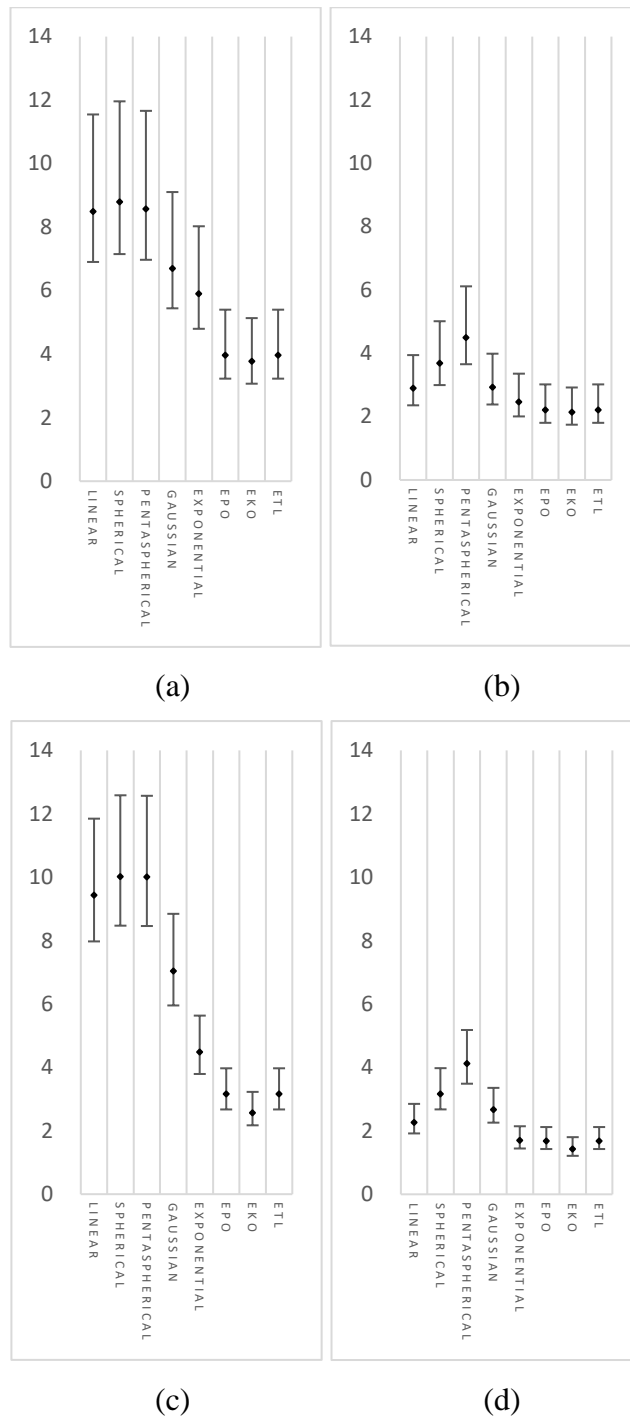
40

Figure 3.5 The 95% CI of errors for 31 points (a) Zone 1, (b) Zone 2 ; and 51 points (c) Zone 1, (d) Zone 2.

The performance of our new methods was again confirmed by the similar results of the study area in Zone 2, where the original methods gave the RMSE between 2.4708 and 4.4998, but our new methods gave the RMSE between 2.1499 and 2.2218. Therefore, the estimation of the height of test locations with our novel techniques is more accurate than the original ones. These results are in line with the other four error assessment methods. The by-product received for this second experiment is the area with lower terrain variations tends to yield smaller RMSE than the area with higher terrain variations as Zone 1. It is interesting to note that when the area is quite flat, the Linear model performed with quite an acceptable error, even though its RMSE of 2.9023 is slightly higher than our three models. It is possible that the linear semivariogram model gives the best fit for the small data variance and number of points (Tran and Nguyen, 2008)

When increasing the input points to 51 points, Table 3.4 and Table 3.5 show similar performance. In particular, the RMSEs of EKO were reduced to 2.5704 and 1.4353 for Zone 1 and Zone 2, respectively. As expected, the error of our three models decreases when the sample size increases.

According to the first four experiments, the problem still lies within the number of sampled points and the quality of our error assessment. To be statistically confident that our models are superior and to what extent they can be claimed, we compute the 95% CI for each sample size in each zone. Table 3.6 reveals that we can be 95% certain that the EPO, EKO, and ETL models are more accurate than the rest of the models except the Exponential model. To ease the interpretation, Figure 3.5 visually compares the 95% CI for all models. However, in Zone 2 with 31 points, we cannot conclusively claim that our three models perform the best as the intervals in Figure 3.5(b) overlap one another. We can only claim that our models perform better than the Pentasherical model. We hence increase the sample size to 51 points in the hope to better distinguishing the 95% CI when the sample size is larger. Indeed, we discover in Figure 3.5(d) that it is 95% certain that our three models perform better than the Spherical, Pentaspherical, and Gaussian models. Note that the comparison is not conclusive for the Linear model in the case of low terrain variations. In general, the result implies non-significant differences among models.

Overall, the Spherical and Pentaspherical models have the worst performance among the eight models whereas the exponential-based models perform very well. We recommend using our new interpolation algorithms because they offer higher accuracy, and most importantly, our methods do not require guessing of nugget, sill, and range values. However, for an area with minimal elevation variance, we do suggest using the Linear model because it takes little time to calculate, and the accuracy is still acceptable.

3.4.2 Different correlation functions based on MLE

The RMSE values of our proposed methods (EPO, EKO, and ETL) are compared with four variants of MLE methods. As shown in Table 3.7 the errors of MLE methods for Zone 1 were between 3.7302 and 3.8035, whereas the errors of our new methods were between 3.7798 and 3.9749. The result implies non-significant differences among models. When increasing the input points to 51 points, the errors of all seven methods relatively decrease as we expected. In particular, the performance of MLE using the Gaussian correlation function is marginally better than the rest of the models.

For Zone 2, the errors of the four MLE methods were reduced to between 2.0921 and 2.3680. When increasing the input points to 51 points, the performance of all seven models uncovered little difference; the RMSEs of our methods were between 1.4353 and 1.6889 whereas the RMSEs of the four MLE methods were between 1.4181 and 1.5268. In particular, the error of MLE using the Gaussian correlation function is marginally lower than the other six models.

In terms of 95% CI, Figure 3.6. reveals that we cannot be 95% certain that the EPO, EKO, and ETL methods are more accurate than the four MLE methods. The result implies non-significant differences among models. The bounds of all models were reduced when the input changed to 51 points (see Figures 3.6(b) and (d)). The errors for the four MLE methods are relatively close to our proposed methods, though they are marginally lower than ours in some cases. We concluded that conventional methods that use MLE perform better than the five conventional methods in Figure 3.6.
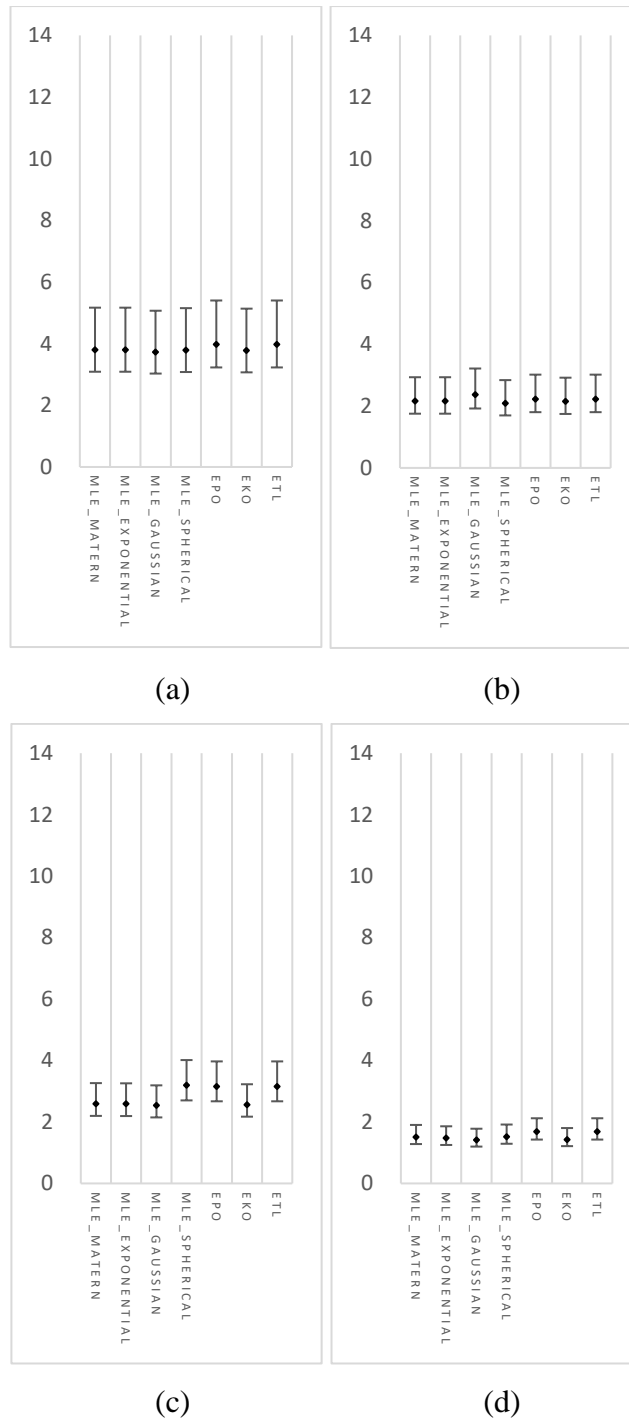
Figure 3.6 The 95% CI of errors on different estimation methods for 31 points (a) Zone 1, (b) Zone 2 ; and 51 points (c) Zone 1, (d) Zone 2.

3.5 Chapter Summary

We proposed three alternative kriging algorithms. These three variants compute the optimal values of kriging parameters automatically and predict the unknown value at the unsampled location. The proposed kriging models are Exponential with Parameter Optimizer (EPO) model, Exponential with k-Iterations Optimizer (EKO) model, and Exponential with Polynomial-Trend Line (ETL) model. We concluded that the errors of our three methods are exceptionally small when compared with the five conventional semivariogram models, namely, Linear, Spherical, Pentaspherical, Gaussian, and classical Exponential models. Statistically speaking, when the sample size is sufficiently large, it can be 95% certain that the EPO, EKO, and ETL models are more accurate than the Spherical, Pentaspherical, and Gaussian models. It is also discovered that for the region with low terrain variations, the Linear model is adequately usable. Hence, our work supports the different use of algorithms for different area types. For example, we use the exponential-based models for relatively large terrain variations and the linear ones for relatively flat areas. Our future work will deal with dynamic selections of suitable algorithms for different elevation variances.

For the conventional methods presented in the thesis, we recommended estimating the model parameters using the MLE method with the Gaussian correlation function, though there are non-significant differences among the four correlation functions presented in Section 3.3.6. In summary, our main contributions are the proposal of three novel kriging algorithms based on exponential semivariograms; the enhancement of conventional methods with the maximum likelihood estimation; and the resilience of applying different algorithms for different area types. Our algorithms are easy to use because there is no need to specify nugget, sill, and range values. The practical benefit of our results is in the area of improving the 3D surface plot as the number of unknown points can be predicted, so the plot can be filled with more points resulting in a more accurate plot.

CHAPTER 4
TWO DIVIDE AND CONQUER ALGORITHMS

Although the presented algorithms are accurate techniques in previously studied areas, in this chapter, we discovered that using them for different areas (other than PSU) with more data points is impractical due to the five-day wait required to finish the model surface plot. Hence, the next problems will be focusing on decreasing the computing time to less than 5 minutes while maintaining or slightly diminishing the accuracy. This chapter will introduce two divide and conquer algorithms by first applying the divide and conquer technique and later improving it by introducing a slope (terrain variation) threshold parameter. This chapter will cover the technique used, wider study areas, experiment results, discussion, and chapter summary.

4.1 Introduction

Environmental management is described as a part of a management system to achieve specific environmental behaviours that can reduce the impact on the environment. Rapid decisions and judgments can reduce the major damage in environmental management. The computer programing and interpolation methods have opened new doors for environmental management to reduce the computational time and improve the accuracy of data. Many interpolation techniques were developed to predict the spatial values in unsampled points (Chilès and Delfiner, 1999; Chilès and Desassis, 2018; Cressie, 1993; Wackernagel, 2003). Points interpolation helps save tremendous time and costs of collecting samples in every location.

The divide and conquer technique is one of an algorithm design paradigm that solves complex problems by dividing them into smaller problems and then merging them one by one (Comen et al., 2009). The data set is split into non-overlapping small sub-data, and the target model is estimated in each sub-data. We will incorporate this design paradigm into a field of spatial interpolation technique named kriging. There are many works in the area of kriging interpolation (Aguilar et al., 2005; Bello-Pineda and Hernandez-Stefanoni, 2007; Hasanipanah et al., 2021; Lam, 1983; Li and Heap, 2011; Meng, 2021; Mert and Dag, 2017; Hasan et al., 2021; Nie et al., 2021; Setianto and Triandini, 2013; Behzadi and Jalilzadeh, 2021; Antal et al., 2021; Antal et

al., 2021; Yanto et al., 2022; Wang et al., 2021; Natsagdorj et al., 2021; Banjo et al., 2021; Ma et al., 2021). Moreover, there have been advanced studies of the kriging surrogate models to improve optimization speed over recent years (Bhattacharyya, 2022; Hao et al., 2020; Hao et al., 2021). However, the divide and conquer strategy for spatial interpolation has been the subject of few investigations (Deng et al., 2017; Wang et al., 2020; Du and Deng, 2021). In this thesis, we focus on developing algorithms with the divide and conquer approach for fast handling of spatial data. The kriging models are used to predict heights at unknown locations. Our work can make landscape 3D visual assessment much more practical as the computational time to complete the model surface plot can be reduced from 5 days to 5 minutes with only a marginal trade-off in model accuracy

4.2 Methodology

In our previous work, we designed three new algorithms; Exponential with Parameter Optimizer (EPO) model, Exponential with k-Iterations Optimizer (EKO) model, and exponential with Polynomial-Trend Line (ETL) model. The EPO was designed to optimize kriging parameters for the minimal predicted error. The EKO iteratively calls EPO while narrowing down the kriging parameter values in each call. The ETL applied a rather different technique by computing the relation between distance and semivariance in the form of the polynomial trend line, then the line is used to re-compute the better version of the weight vector for point predictions. We also improved the existing Exponential model by incorporating a parameter estimation technique called MLE (Pan and Fang, 2002) in the process of optimizing the nugget, sill, and rage of the model. Therefore, we have four adaptations of exponential-based models (Exponential, EPO, EKO, and ETL).

Although these algorithms are accurate approaches in the previous work, we found that applying them for landscape 3D visual assessment is not so practical as the waiting time to complete the model surface plot can be up to 5 days. Therefore, the remaining challenges are to reduce the computational time to be less than 5 minutes while preserving the accuracy or reducing it down marginally. This thesis presents two algorithms by first applying the divide and conquer technique and later improving it by introducing a slope (terrain variation) threshold parameter.

Algorithm 4: DVC

Input: $L(x_i, y_i)$ and $z(S_i)$ /* locations and altitudes */

Output: $z_p(s)$ and RMSE   /* altitude of a predicted point and error */

Step 1: Input $L(x_i, y_i)$ and $z(S_i)$ where $i = 1, 2, …, n$

Step 2: Choose the grid size of gxg where $g = 1, 2, 3, 4$

Step 3: Divide data into gxg subdata based on latitude and longtitude of $L(x_i, y_i)$

Step 4: Input the first subdata (the 1st cell)

Step 5: Calculate $z_p(s)$ for each predicted point in the cell based on 6 kriging models;
Gaussian, Spherical, Exponential, EPO, EKO, and ETL

Step 6: Calculate RMSE for each algolithm

Step 7: Repeat Steps 4 – 6 for all cells

Step 8: Calculate total RMSE

Figure 4.1 Divide-and-conquer algorithm based on grid sizes.

Figure 4.1. shows DVC, which is the divide and conquer algorithm with 4 alternative grid sizes. The input consists of the locations with the x and y coordinates called $L(x_i, y_i)$ and the altitudes $z(L(x_i, y_i))$ at the corresponding locations. Users will select the size of gxg cells. After that, the point locations are divided according to latitude and longitude. Figure 4.4 depicts the different grid sizes. Figure 4.4(a) is the original grid 1x1 served as the reference when the divide and conquer concept is ignored. Figure 4.4(b), 4.4(c), and 4.4(d) are the divide and conquer with grid sizes of 2x2, 3x3, and 4x4.

The predicted points are interpolated based on six kriging models. In the conventional model, semivariances are computed based on Equations (2.5), (2.7), and (2.8), while EPO, EKO, and ETL were introduced in our previous work. The DVC design paradigm is meant to significantly reduce the computational time by decreasing the size of the distance matrix D in our proposed models. For example, 500 sampled points when dividing into grid size 4x4, the number of points in each cell is reduced to around 31 points (n/g^2 points). This has led to the D of size 31x31, which is considerably smaller than the D of size 500x500. Although the calculation must be made for every point in g^2 cells, the total calculation time is less than one large cell with n sampled points.

Figure 4.2 shows the improved algorithm of the divide and conquer by adding a new parameter called β, a slope of terrain in the given cell, which is calculated based on Equation (4.1).

Slope of terrain:

$$\beta = \left(\frac{z(S_{max})-z(S_{min})}{\sqrt{(x_{max}-x_{min})^2+(y_{max}-y_{min})^2}} \right) \times 100 \qquad (4.1)$$

SDVC requires the minimum number of cells to be four, and the divide and conquer paradigm is forced. The slope is used to control the accuracy of the model. If the β value in a cell exceeds the specified slope (α), the algorithm will select the EKO model that works well in areas with high terrain variations.

Algorithm 5: SDVC

Input: $L(x_i, y_i)$, $z(S_i)$, and α /* locations, altitudes, and a slope threshold */

Output: $z_p(s)$ and RMSE   /* altitude of a predicted point and error */

Step 1: Input $L(x_i, y_i)$ and $z(S_i)$ where $i = 1, 2, \ldots, n$

Step 2: Choose the grid size of $g \times g$ where $g = 2, 3, 4$

Step 3: Divide data into $g \times g$ subdata based on latitude and longitude of $L(x_i, y_i)$

Step 4: Input the first sub-data (the 1st cell)

Step 5: Calculate a slope β of the cell

Step 6: If $\beta < \alpha$, calculate $z_p(s)$ using the Exponential model, otherwise use EKO

Step 7-9: as per Step 6-8 of Algorithm 4: DVC

Figure 4.2 Improved divide-and-conquer algorithm based on slope (terrain variation) threshold.

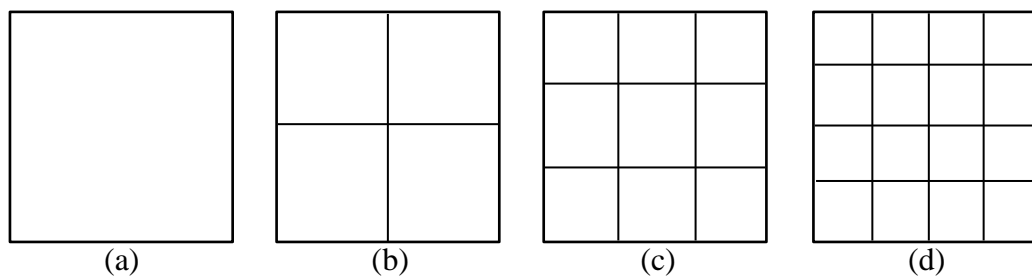(a)   (b)   (c)   (d)

Figure 4.3 Grid sizes; (a) 1x1, (b) 2x2, (c) 3x3, (d) 4x4.

By applying a slope threshold $\alpha = 5$ in the SDVC algorithm, examples in Figure 4.4(a), (b), and (c) show the sub-areas in the mountainous area where the EKO model indicated by the black cells are utilized relatively more often than the Exponential model indicated by the white cells. On the other hand, the EKO model will be used less when the area is relatively flat, as indicated by Figures 4.4 (d), (e), and (f). As a result of this design paradigm, the computational time in the flat area should be reduced because the SDVC algorithm splits the less complicated work for the Exponential model. We anticipate that the effect of increasing the slope threshold from 5% to 8% would reduce the computational time even more and hopefully still within the acceptable trade-off in model accuracy. Examples in Figure 4.5(a) – (f) simulate a slope threshold $\alpha = 8$; there are more white cells in the areas which means the Exponential model is used more often than the EKO model.
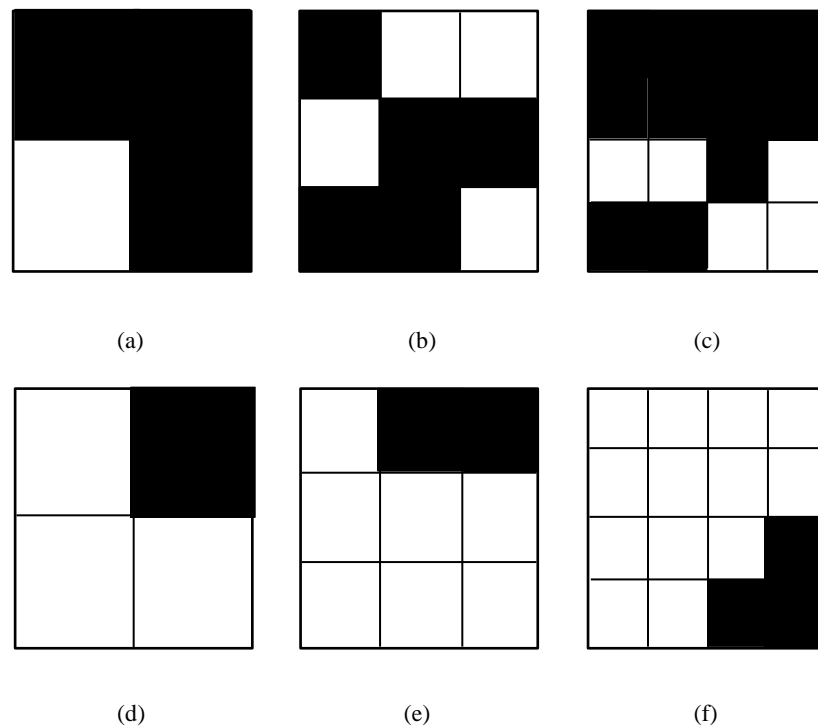

(a)　　　　　　　　(b)　　　　　　　　(c)

(d)　　　　　　　　(e)　　　　　　　　(f)

Figure 4.4 Slope threshold of 5% for mountainous area; (a) 2x2, (b) 3x3, (c) 4x4; and flat area (d) 2x2, (e) 3x3, (f) 4x4.
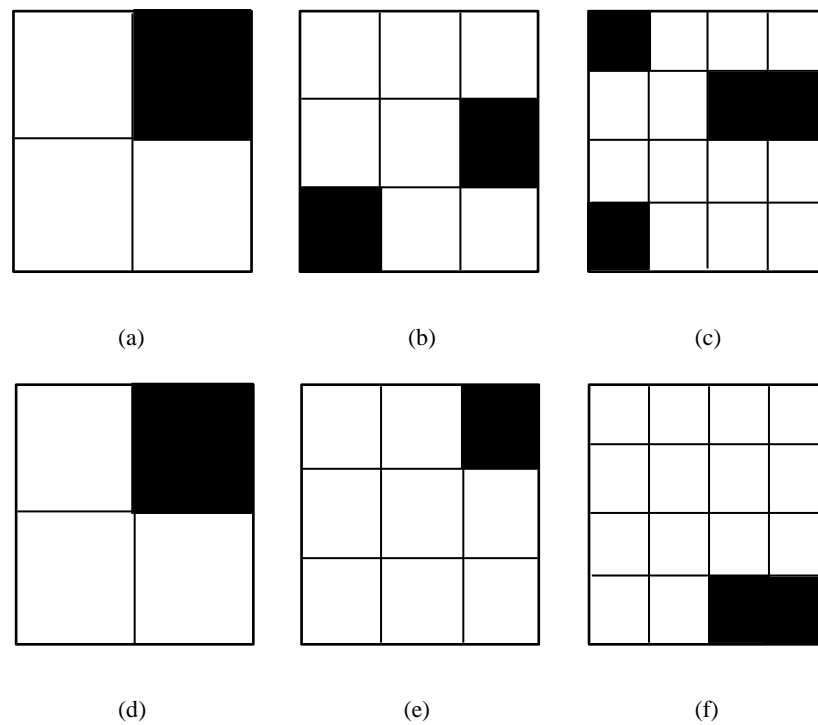
Figure 4.5 Slope threshold of 8% for mountainous area; (a) 2x2, (b) 3x3, (c) 4x4; and flat area (d) 2x2, (e) 3x3, (f) 4x4.

4.3 Study Areas and Experimental Results

Mount Adams and Yakama are the native American reservation area located in Washington state, USA. Mount Adams is the highest point in the Yakima region, located at a North latitude of 46°12′9″ and a West longitude of 121°29′27″. We have chosen two study areas, namely, mountainous and flat areas, as displayed in Figure 4.6 and Figure 4.7, respectively. The input of 500 points in each area is systematically sampled from the 1-meter DEM data of the x and y coordinates in UTM coordinate systems and their altitudes in meters (m). The 1-meter DEM has a horizontal accuracy within 1 meter, and vertical accuracy of 19.6 cm at the 95-percent confidence (Arundel et al., 2015). All 1-meter DEM input data are downloaded from the nation map (TNM) of the U.S. Geological Survey. Note that the average height and standard deviation of the flat area are significantly smaller than the one of the mountainous area. We chose these regions to examine how our algorithms perform in different types of terrain. As a result, we will investigate how the performances (accuracy and computational time) of the six models, namely, Gaussian, Spherical, Exponential, EPO, EKO, and ETL are affected given the two terrain types and four different grid sizes.
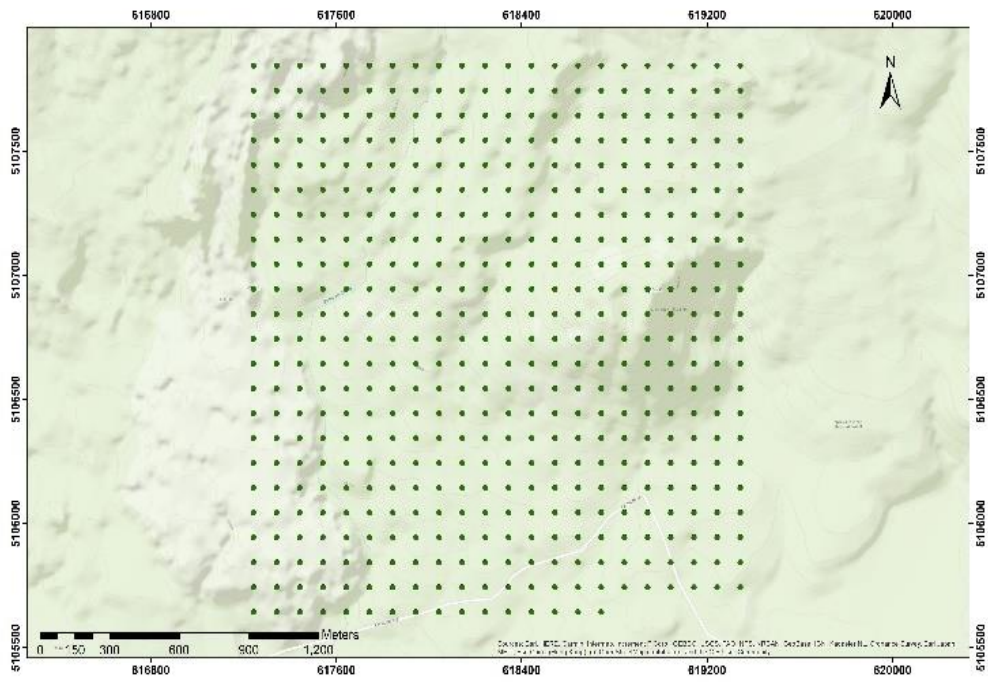
Figure 4.6 Sampled points for the mountainous area near Mount Adams.
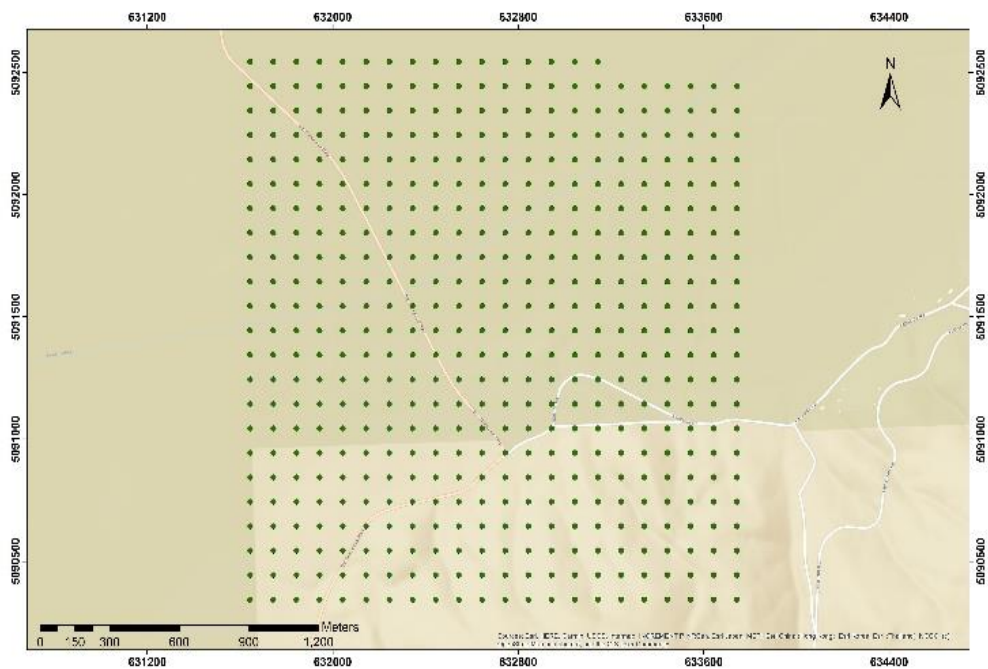

Figure 4.7 Sampled points for the flat area in Yakima, Washington.

The characteristics of the mountainous and flat areas are shown in Table 4.1. Next, we will examine how the algorithms perform with more samples.

Table 4.1 Characteristics of height-samples in mountainous and flat areas.

Area	Mountainous area	Flat area
Sample Size	500	500
Average	1307.3222	576.7202
Standard deviation	84.1536	44.3313
95% Margin of error	7.3764	3.8858
Upper bound	1314.6986	580.6060
Lower bound	1299.9458	572.8344
Max	1474.4000	767.0274
Min	1155.6300	553.5831
Range	318.7700	213.4443

*The unit of input data (meters).

4.3.1 Accuracies versus grid sizes based on DVC

4.3.1.1 Mountainous area

In the first experiment, the DVC algorithm is used to examine the accuracies of six models, and the result is presented in Table 4.2.

Table 4.2 The RMSE values for mountainous area based on DVC algorithm.

Model	Grid Size			
	1x1	2x2	3x3	4x4
Gaussian	25.3482	22.2799	24.0599	21.9945
Spherical	36.4505	34.8079	28.12647	19.4797
Exponential	5.0250	5.1944	5.3115	5.7556
EPO	5.1220	5.2375	5.4276	5.9978
EKO	5.0920	5.2163	5.3945	5.8938
ETL	5.1220	5.2375	5.4276	5.9978

*The unit of input data (meters).

We improved the Gaussian, Spherical, and Exponential model by incorporating a parameter estimation technique called MLE (Pan and Fang, 2002), shown in Tables 4.2 and 4.3. For Gaussian and Spherical models, the RMSEs of grid size 4x4 are lower than those of other grid sizes. For instance, the errors of Gaussian and Spherical models decrease by 13.23% and 46.56%, respectively. We believe that the accuracy increases due to the $\gamma(d)$ formula which quantifies the degree of spatial dependence between points fits the two models better when the regions are condensed. On the other hand, the RMSEs of the four exponential-based models (Exponential, EPO, EKO, and ETL) slightly increase in the range of 0.7306 – 0.8758 which corresponds to about 14.54% to 17.10%. We found that the $\gamma(d)$ formula for computing semivariance fits exponential models marginally less well in regions with fewer inter-sample distances. Nevertheless, the EKO and Exponential models are more accurate than the rest of the models.

4.3.1.2 Flat area

In the second experiment, the errors of six models for the flat area are presented in Table 4.3.

Table 4.3 The RMSE values for flat area based on DVC algorithm.

Model	Grid Size			
	1x1	2x2	3x3	4x4
Gaussian	15.4289	13.5957	14.4233	14.2205
Spherical	28.4958	27.3271	19.5480	13.8819
Exponential	4.0514	4.3451	4.5522	4.9203
EPO	4.1754	4.3765	4.6604	5.0942
EKO	4.0915	4.3093	4.5483	5.0525
ETL	4.1754	4.3765	4.6604	5.0942

*The unit of input data (meters).

As we expected, the flat area yields smaller RMSE than the area with higher terrain variations like the mountainous area. The power of divide and conquer works in favour of error reduction of the first two models. On the other hand, the errors

of the four exponential-based models slightly increase when the grid size increases, yet the RMSEs are all less than 5.095. Note that the EKO slightly outperforms the Exponential model in some cases. Although the divide and conquer did not help our four exponential cases in terms of accuracy, the hidden benefit lies in the computational time. Hence, we chose to further investigate the EKO and the Exponential models as they performed very well in both the flat and mountainous areas. Later, in the SDVC algorithm, we will utilize both the EKO and the Exponential models in the algorithm. The remaining challenge is to make the EKO much more practical in terms of computational time.

Table 4.4 The RMSE values and running time for two terrain types based on the EKO model of DVC algorithm.

Grid Size	RMSE		Running Time (mins)	
	Mountainous	Flat	Mountainous	Flat
1x1	5.0920	4.0915	7,705	7,650
2x2	5.2163	4.3093	1,540	1,440
3x3	5.3945	4.5483	104	109
4x4	5.8938	5.0525	33	33

*The unit of input data (meters).

Table 4.5 provides a summary of six distinct models, along with a comparison of their advantages and disadvantages. The accuracy of Gaussian and Spherical models is inferior to that of Exponential, EPO, EKO, and ETL models. However, it is standard and available in multiple software packages. EKO and Exponential models have the highest accuracy among exponential-based models; however, in some instances, EKO outperforms Exponential.

Therefore, we chose to further investigate the EKO and the Exponential models, given their accuracy is the highest among the six examined models. Later, in the SDVC algorithm, we will utilize both the EKO and the Exponential models in the algorithm. The remaining challenge is to make the EKO much more practical in terms of computational time.

Table 4.5 The pros and cons of six different models.

Model	Pros	Cons
Gaussian, Spherical	Standard and accessible in several software programs	Medium accuracy
Exponential	High accuracy	Defeated by EKO in some cases
EPO, ETL	High accuracy (less than EKO)	Time-consuming
EKO	High accuracy	Time-consuming

The third experiment in this section investigates the trade-off in accuracy and running time-based on the EKO model. Table 4.6 reveals that by increasing the grid size to 4x4, the computational time for the mountainous area is substantially reduced from 7,705 to 33 minutes (reduced by 233 times); and the computational time for the flat area is reduced from 7,650 to 33 minutes (reduced by 232 times). Both cases are in exchange for increasing errors where for some applications, these slight increases are acceptable. With these results, we are now ready to proceed with SDVC.

4.3.2 Computational time versus grid sizes based on SDVC

Except for a very large study area, keep dividing the area into grid sizes greater than 4x4 is not always a good idea as the errors also increase. In this experiment, we introduced a new way to improve the computational time. We will apply the SDVC algorithm presented in Section 4.3.3 with two different settings of slope threshold ($\alpha = 5$ and $\alpha = 8$). The choice of 5% and 8% were inspired by the slope transition values in the slope classification (Elewa and Qaddah, 2011), which are typically described as gentle undulating (2-5%), undulating (5-8%), and rolling (8-15%). For further reading about the slope classification, we suggest studies on the landscape analysis of geographical names (Chen et al., 2014), effective identification of terrain positions from gridded DEM data (Jiang et al., 2018), and terrain classification divided into uniform slopes and basins (Iwahashi and Yamazaki, 2022).

Table 4.6 The RMSE values and running time for two terrain types based on SDVC algorithm with slope thresholds of 5% and 8%

Grid size	RMSE		Running Time (mins)	
	Mountainous	Flat	Mountainous	Flat
$\alpha = 5$				
2x2	5.2262	4.3618	138	132
3x3	5.3422	4.6567	15	14
4x4	5.4435	4.8972	5	4
$\alpha = 8$				
2x2	5.2077	4.3618	128	125
3x3	5.3429	4.6061	14	13
4x4	5.3315	4.8439	4	4

4.3.3 Slope threshold of 5%

The result is shown in Table 4.6. By increasing the grid size from 2x2 to 4x4, the error is increased by 0.2173 for the mountainous area and 0.5354 for the flat area. However, the computational time for the mountainous area is significantly reduced from 138 to 5 minutes (reduced by 28 times); and the computational time for the flat area is reduced from 132 to 4 minutes (reduced by 33 times). The computational time in the flat area is reduced because the SDVC algorithm splits the less complicated work for the Exponential model. In the next experiment, we will investigate the effect of increasing the slope threshold from 5% to 8%. We expect to reduce the computational time even more because the Exponential model will be used more in the sub-areas, hopefully within the acceptable trade-off in model accuracy.

4.3.4 Slope threshold of 8%

Table 4.6 also shows that for the slope threshold of 8% the error of grid size 4x4 is increased by 0.1238 for the mountainous area and 0.4821 for the flat area. The running time is slightly decreased. By increasing the grid size to 4x4, the

computational time is substantially reduced to 4 minutes. The waiting time is now much more practical but with the trade-off in model accuracy. However, there are non-significant reductions in the running time. Therefore, we suggest not to go beyond the slope threshold of 8%. In fact, we recommend to trade-off at 8%, which we can gain reasonable computational time and acceptable accuracy.

4.4 Discussion

The discussion will be divided into 2 parts. Firstly, we will discuss the trade-off between accuracy and computational time. Secondly, the application in environmental management will be discussed in the section.

4.4.1 Trade-off between accuracy and computational time

Our proposed methods are designed specifically to handle both mountainous and flat areas. The accuracy of the proposed methods is significantly higher than the standard kriging models (Gaussian and Spherical). Considering the errors from the two areas, the RMSE values of Gaussian and Spherical models are in the range of 13.8819 – 36.4505, while our methods are only 4.0514 – 5.9978. In terms of the computational time, by increasing the grid size to 4x4 and applying the SDVC algorithm, the computational time is drastically reduced from 7,705 to 4 minutes for the mountainous area (reduced by 1,926 times); and from 7,650 to 4 minutes for the flat area (reduced by 1,913 times). However, it comes with a trade-off in model accuracy. For applications that require acceptably high accuracy and the waiting time is not an issue, we recommend using the DVC algorithm. For fast prediction, we suggest using the SDVC with a slope threshold of 8%, which we can gain relatively short computational time and reasonable accuracy.

In addition, we managed to conduct a separate experiment to demonstrate that our exponential-based models also perform better than one of the previously published ones. Based on the SINENVAP algorithm (Guisande et al., 2019), when the optimal grid size is used, and the model is set to Exponential, the computational times for the mountainous and flat areas (based on the same dataset) are 105 and 108 minutes, respectively. In terms of errors, there are non-significant

differences among algorithms. The RMSEs for the mountainous and flat areas are 5.3957 and 4.4505, respectively

For the equipment error and model error, most research (Li and Heap, 2011) ignore equipment error and concentrate on model inaccuracy. Azimi et al. (2019) determined the overall uncertainty by combining experimental measurement error and model error. This thesis focuses on model error. For the whole uncertainty of our work, combine our model's and measurements' errors.

4.5 Chapter Summary

We presented two divide and conquer algorithms for the prediction of spatial heights. The first algorithm is called DVC. It contains six different kriging models to choose from. The evaluation revealed that the four exponential-based models (Exponential, EPO, EKO, and ETL) outperform the Gaussian and Spherical models in terms of accuracy. Among the exponential-based models, the two best performances are the EKO and Exponential models, which are later used in the second algorithm called SDVC. In terms of the computational time, the DVC algorithm substantially reduced the waiting time from 7,705 to 33 minutes. Moreover, the SDVC algorithm can reduce the waiting time further down to 4 minutes. This result is now much more practical (than in the previous work) but with a trade-off in model accuracy. In summary, we recommend using the SDVC algorithm with a slope threshold of 8% where we can gain fast computational time with only a marginal trade-off in model accuracy. In addition, we found that the divide and conquer design paradigm truly helps improve both the accuracy (in the case of the Gaussian and Spherical models) and the running time (in the case of the EKO and Exponential models), resulting in a practical application for landscape 3D visual assessment.

CHAPTER 5

APPLICATION

This chapter describes the application in environmental management, user interface, codes, and results. We developed a program to help us simulate all models studied in this research and implemented a JavaScript user interface for ease of entering the input data.

5.1 Introduction

Practically, the results from this thesis can be used for 3D landscape visualizations in environmental management. Inspecting a 3D surface plot helps reduce the construction risk of building infrastructures blocking flood flows in the study area. The 3D plots are also used in various fields such as flood protection (Ackere et al., 2016; Bales et al., 2007; Krajewski et al., 2017; Yin et al., 2016) transport planning (Keler and Mazimpaka, 2016), and farming management (Reznik et al., 2017). However, environmental managers need accurate spatially continuous data across an area to make competent and confident decisions. Data collected from field surveys are often from sampled points. Thus, the values at unsampled points need to be interpolated to generate spatially continuous data. Accurately predicted points result in an accurate 3D surface plot. Numerous previous works use available tools and programs. (Guisande et al., 2019; Kaya et al., 2019; Yu et al., 2021; Zerdoumi et al., 2022; Yu et al., 2022). However, our algorithm is not yet available anywhere. Therefore, we developed a program to help us simulate all models studied in this research.

5.2 User Interface

We choose to implement our model with Javascript as it is on the web application (Delcev et al., 2018; Kredpattanakul and Limpiyakorn, 2018; Ranjan et al., 2020). For the first experiment with three novel kriging algorithms, we implemented a JavaScript user interface for entering the input data, as shown in Figure 5.1.

The program allows uploading the excel file from the Choose File button. Users can choose to manually guess the nugget, sill, and range values to compare and verify the accuracy of the original methods and our new interpolation methods.

Figure 5.1 JavaScript user interface with three novel kriging algorithms.

For the second experiment with two divide and conquer algorithms, we also developed an application for landscape 3D visual assessment which is implemented in JavaScript. On the top left of the program, users can select the grid sizes of 1x1, 2x2, 3x3, and 4x4 depending on their preferences in Figure 5.2. After selecting the grid sizes, users can select whether to use the DVC or SDVC algorithms

Figure 5.2 JavaScript user interface with two divide and conquer algorithms.

5.3 Source Code

We developed a program to help us simulate all models studied in this research and implemented it with JavaScript. The project creates by ReactJS. Firstly, users should install NodeJS on their computers. To start the project, in the project directory, users must run: npm install and npm start, then run the application in the development mode by opening the http://localhost:3000 to view it in the browser. The page will reload the application if users make edits. Users will also see any list errors in the console.

The first experiment of the thesis presents the three novel kriging algorithms. The source code of these algorithms is made accessible by following the link https://github.com/sukkuea/improved-kriging-algorithms.

The second experiment of the thesis presents two more algorithms to reduce the computational time of our proposed algorithms. The source code of these algorithms is made accessible by following the link https://github.com/sukkuea/kriging-divide-and-conquer-algorithms

The third experiment of the thesis presents the additional interpolated points algorithm. The source code of these algorithms is made accessible by following the link https://github.com/sukkuea/kriging-additional-interpolated-points-algorithm

5.4 Results

The output windows are displayed in Figure 5.3 – Figure 5.5. The bottom part of the output window also indicates how the polynomial trend line fits in the ETL model. Note that in Algorithm 3, we use this kind of line to re-compute the semivariance and the better version of the weight vector. Finally, the predicted heights, together with error assessments, semivariograms and 3D surface plots, are available for supporting decision-making.

For the second experiment with two divide and conquer algorithms, the output window is displayed in Figure 5.6. The contour examples of Gaussian, Spherical, Exponential, EPO, EKO, and ETL models for the mountainous and flat areas are shown in Figures 5.7 and 5.8. The 3D surface plot examples of Gaussian, Spherical, Exponential, EPO, EKO, and ETL models for the mountainous and flat areas are shown

in Figures 5.9 and 5.10. In particular, we showed the results of the 3D surface plot from SDVC algorithms for both areas in Figures 5.11 and 5.12. Our program supports different needs of algorithms for different applications. The DVC is suitable for applications that need high accuracy and has a variety of six kriging models to choose from. The SDVC is suitable for applications that require fast computation. It combined the strengths of the EKO, which works well on different datasets and terrain variations, and the Exponential model, which requires less computational time.
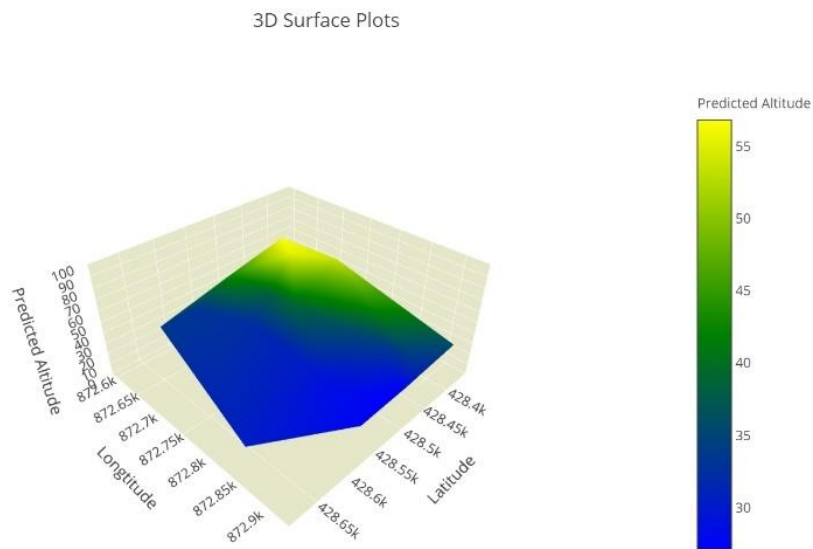
3D Surface Plots

Figure 5.3 The 3D surface plot of the study area.

3D Surface Plots
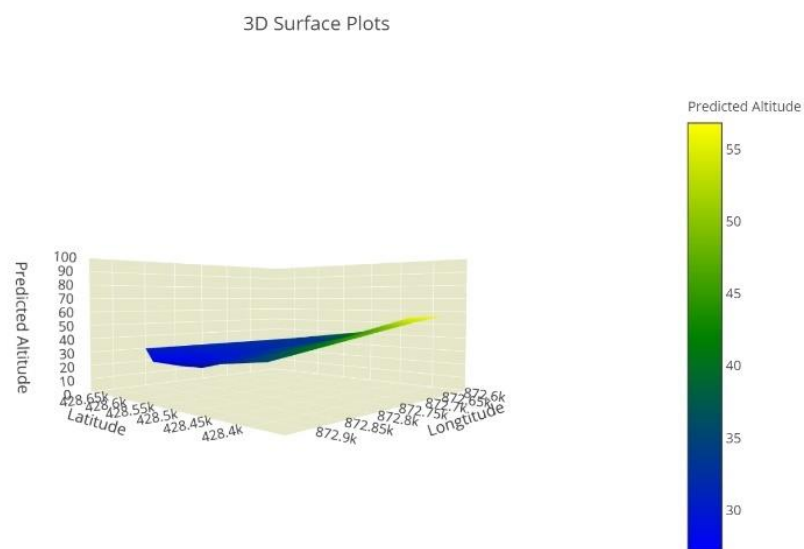
Figure 5.4 Different views for landscape visualization.
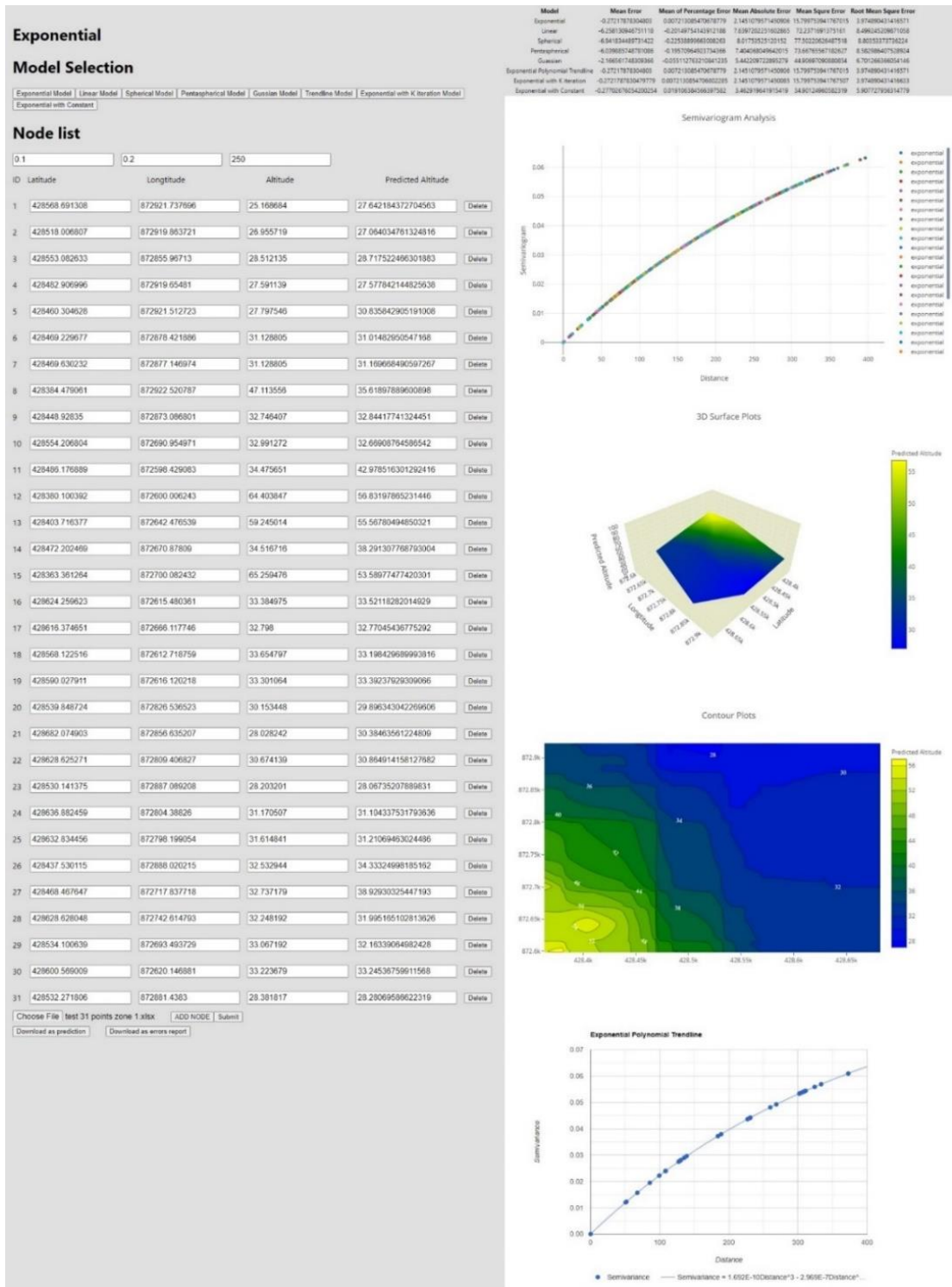
Figure 5.5 The main window for displaying outputs with three novel kriging algorithms.

64
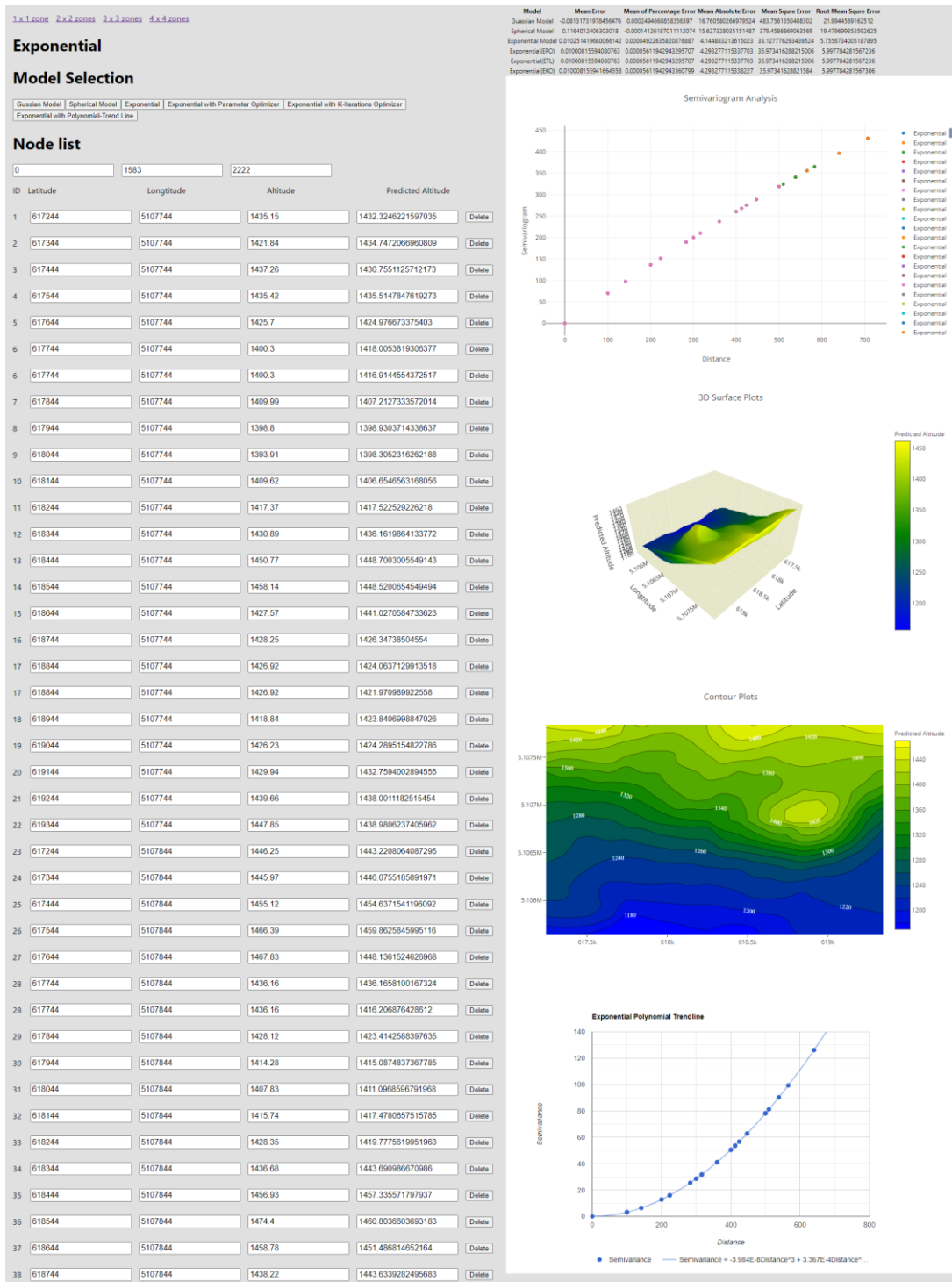
Figure 5.6 The main window for displaying outputs with the divide and conquer algorithms.

Gaussian	EPO
Spherical	EKO
Exponential	ETL
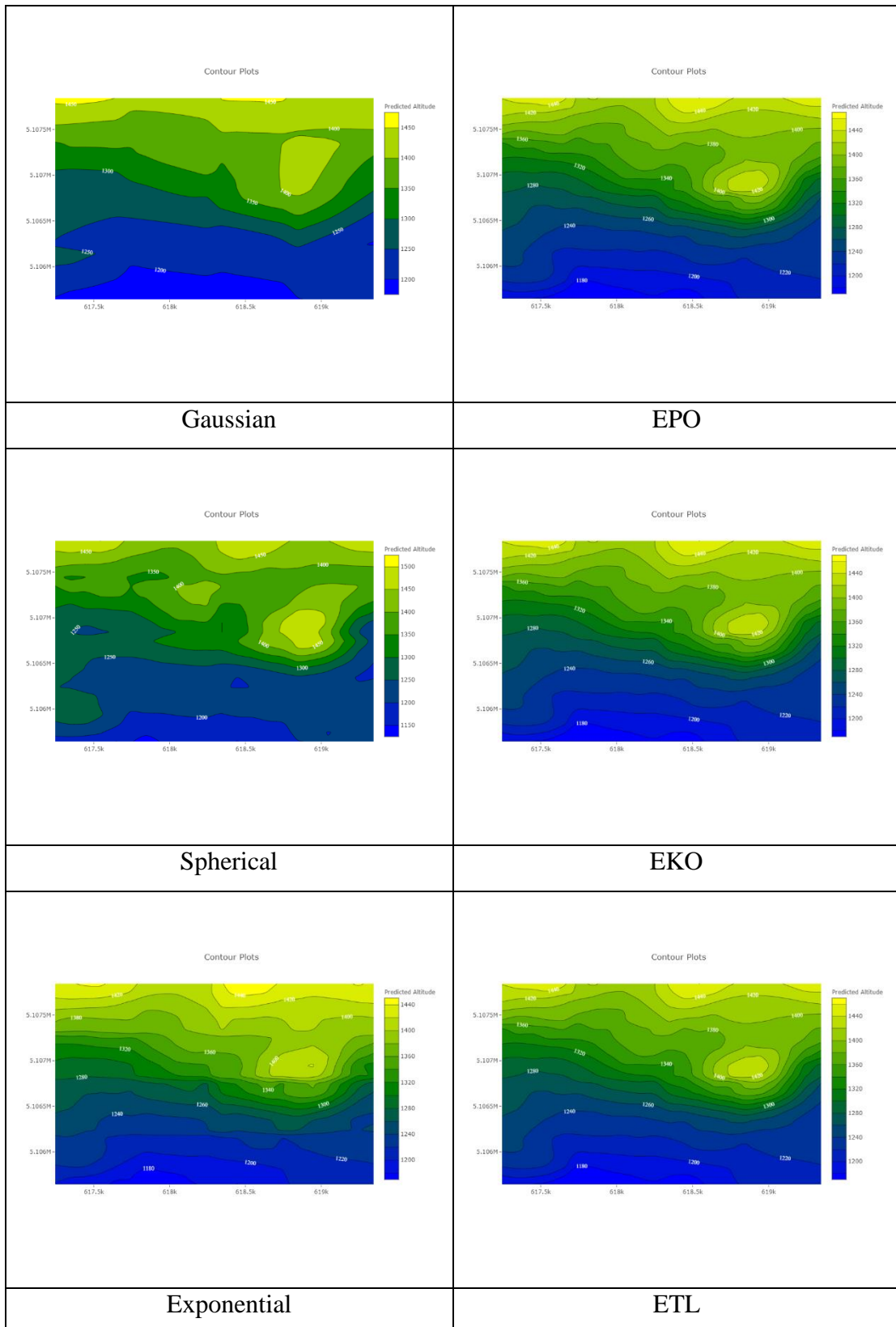
Figure 5.7 The contour from DVC with grid size of 4x4 in the mountainous area.
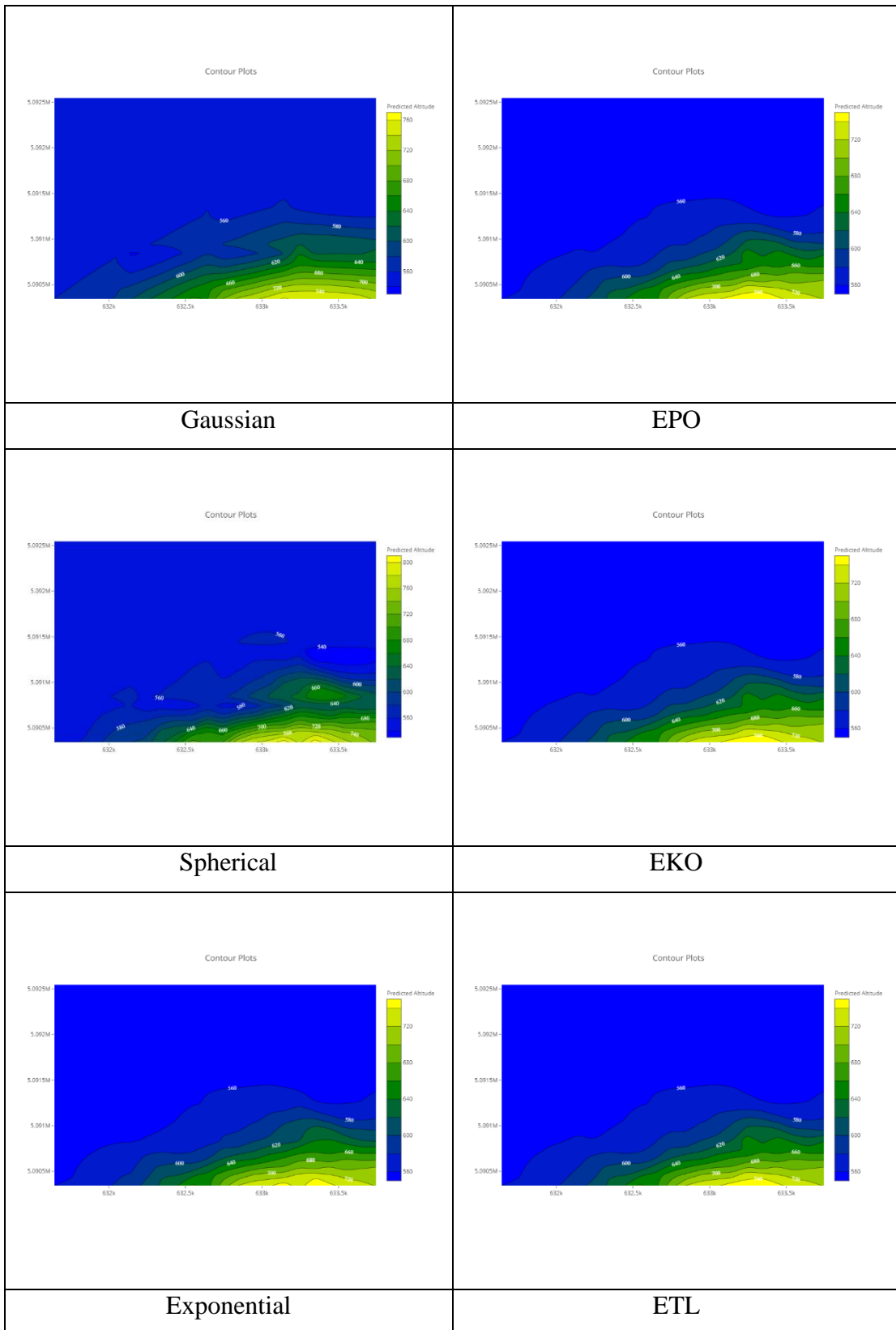
Figure 5.8 The contour from DVC with grid size of 4x4 in the flat area.

Gaussian	EPO
Spherical	EKO
Exponential	ETL
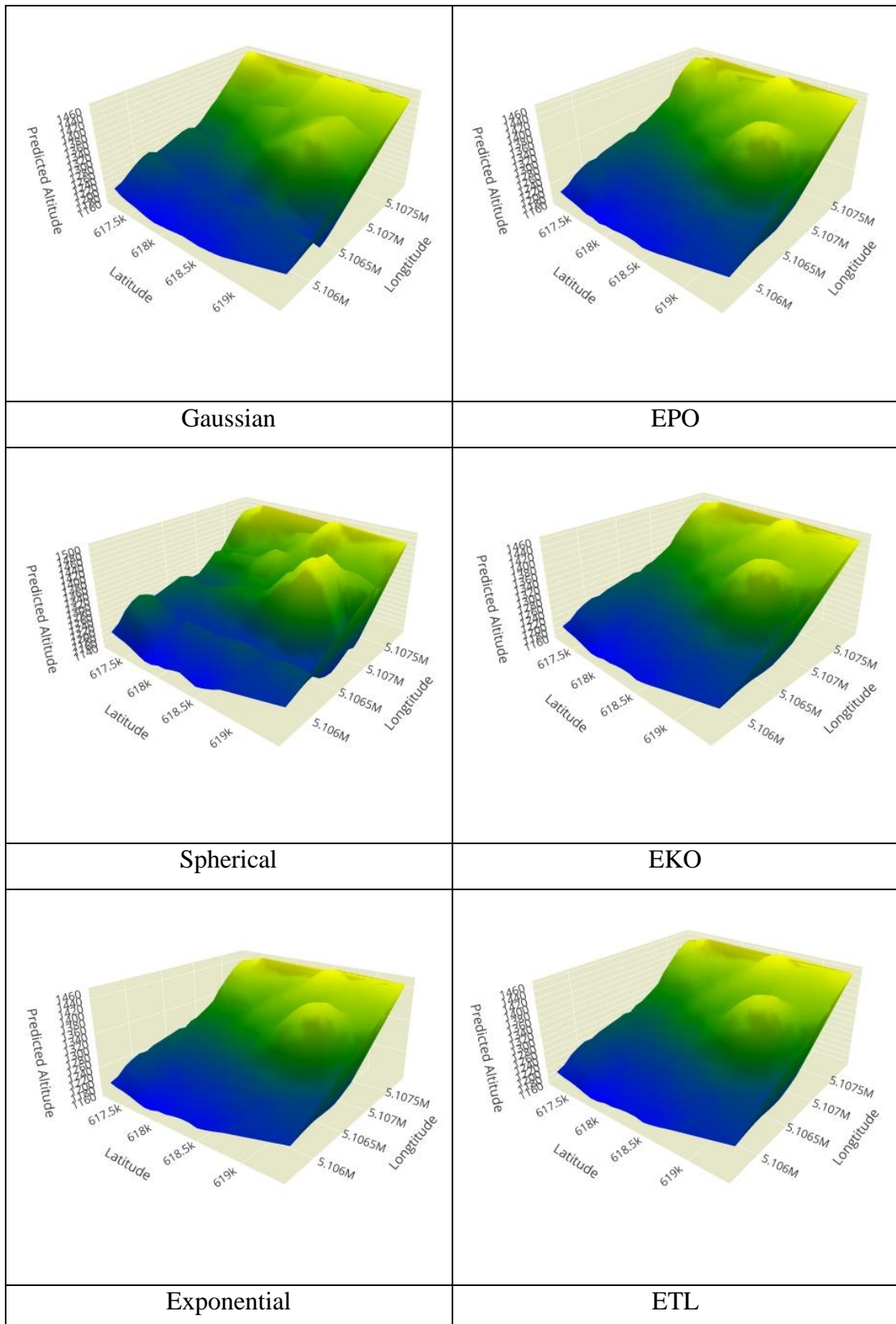
Figure 5.9 The 3D surface plot from DVC models with grid size of 4x4 in the mountainous area.
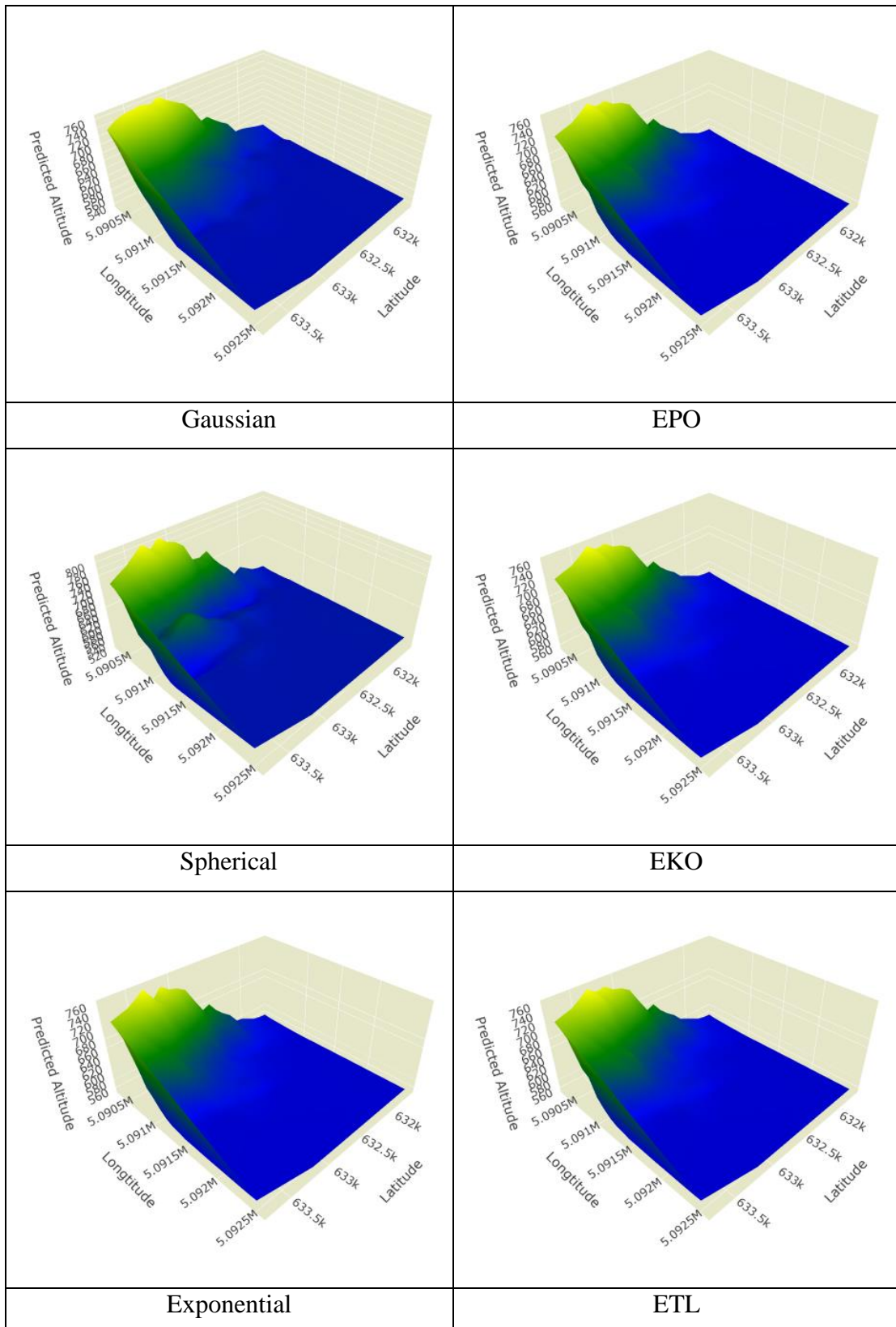
Figure 5.10 The 3D surface plot from DVC with grid size of 4x4 in the flat area.
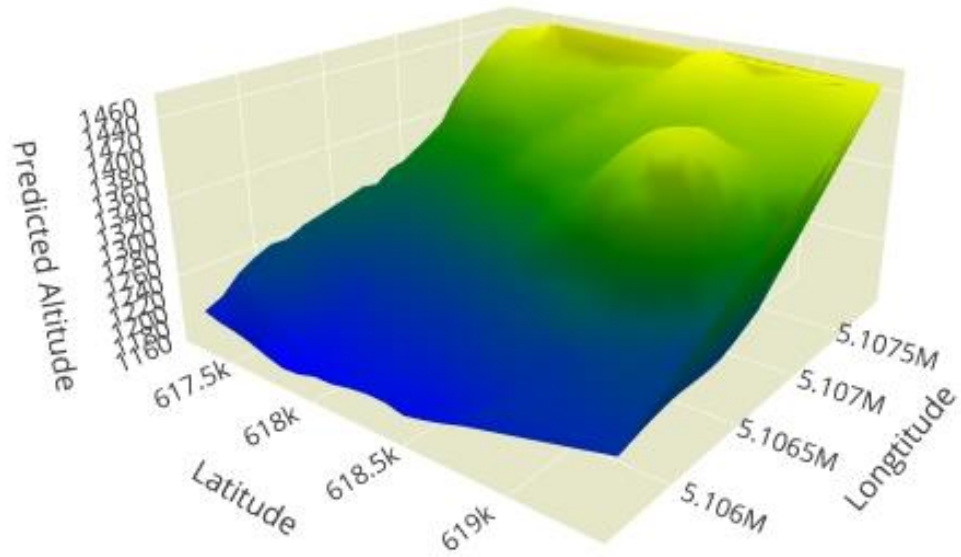
Figure 5.11 The 3D surface plot from SDVC with grid size of 4x4 in the mountainous area.
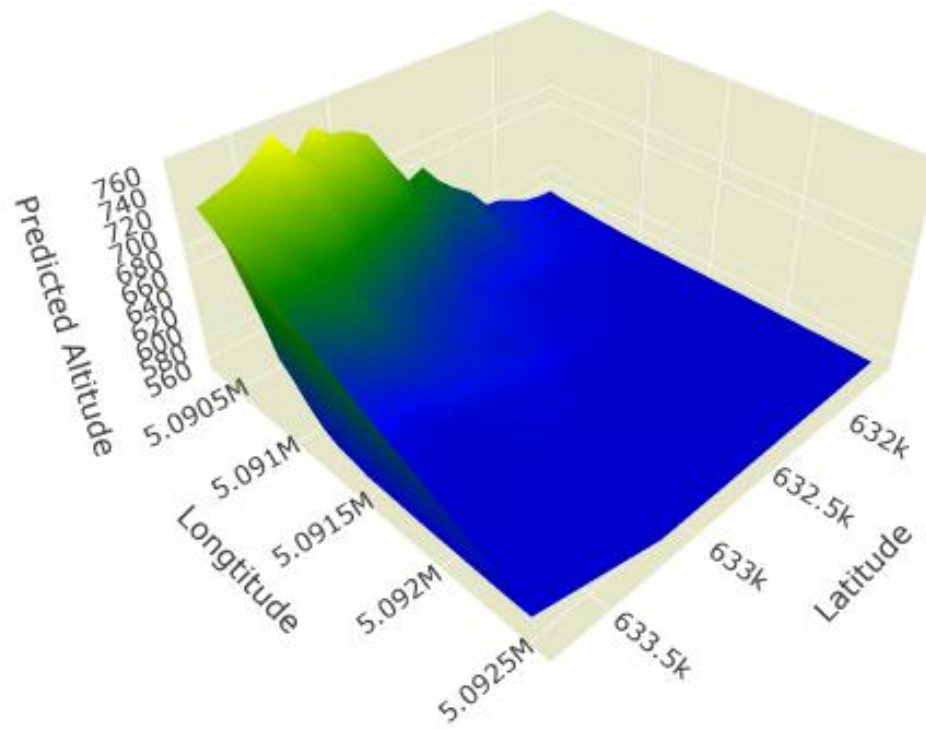
Figure 5.12 The 3D surface plot from SDVC with grid size of 4x4 in the flat area.

5.5 Additional Interpolated Points

For testing the effectiveness of our two divide and conquer algorithms in practice, we add 500 interpolated points into the existing data points. These interpolated points were from iteratively running our algorithms for randomly selected x and y coordinates in the study areas. We expect better accuracy in our 3D surface point with these additional points. Figures 5.13 and 5.14 provide 3D surface plot examples of Gaussian, Spherical, Exponential, EPO, EKO, and ETL models for mountainous and flat areas by adding 500 interpolated points. Specifically, we presented in Figures 5.15 and 5.16 the results of the 3D surface plot generated by SDVC algorithms for each region.

Our additional interpolated points have improved the accuracy of our 3D models. The resultant 3D surface plot has more details than the prior experiment because there are more points which give a better landscape visual assessment. The DVC with additional interpolated points is suitable for applications demanding a higher level of detail which supports six various kriging models. The SDVC with extra interpolated points is recommended for applications that require fast calculations where the models are fixed to EKO and Exponential.
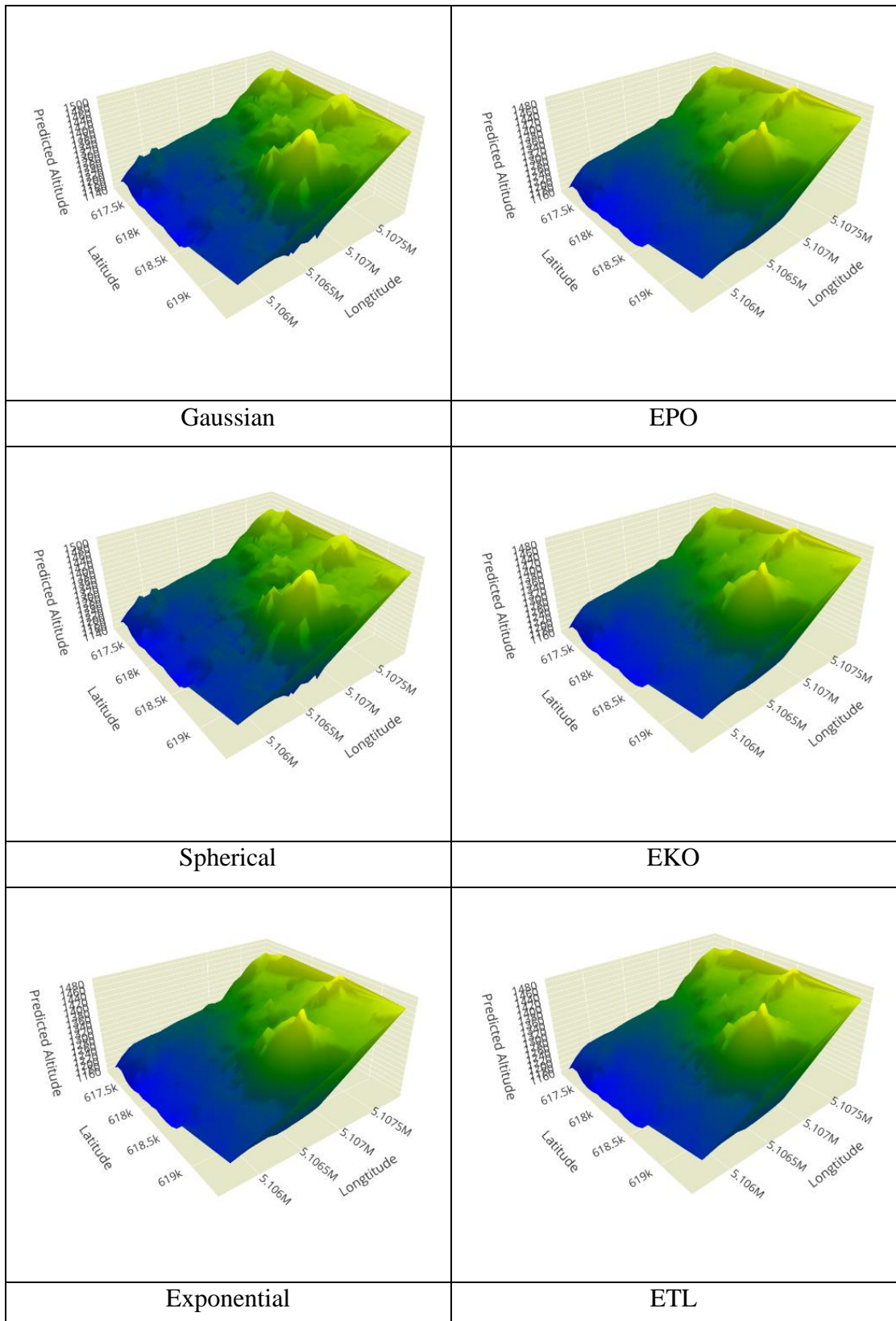
Figure 5.13 The 3D surface plot with additional interpolated points of DVC with grid size of 4x4 in the mountainous area.
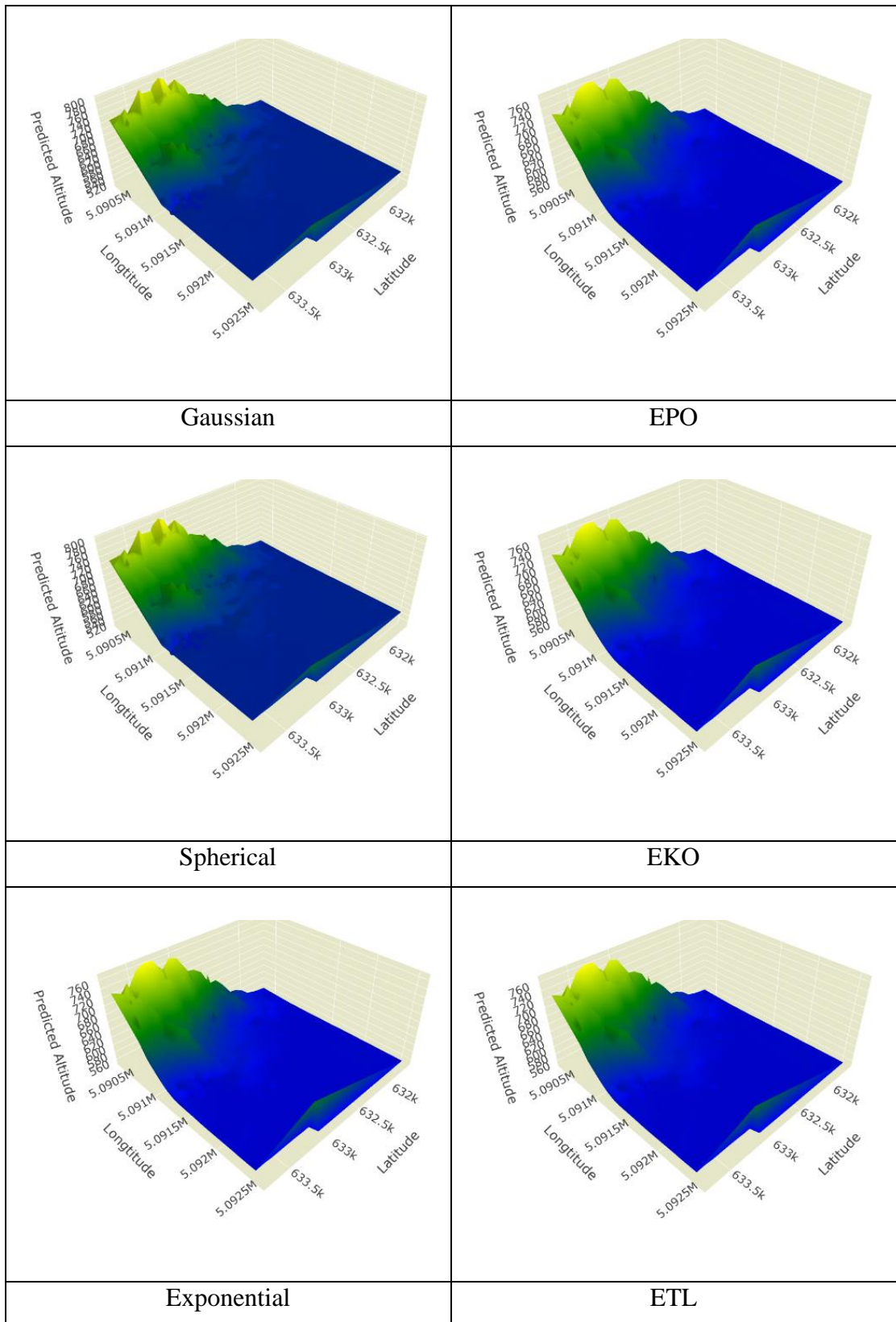
Figure 5.14 The 3D surface plot with additional interpolated points of DVC with grid size of 4x4 in the flat area.
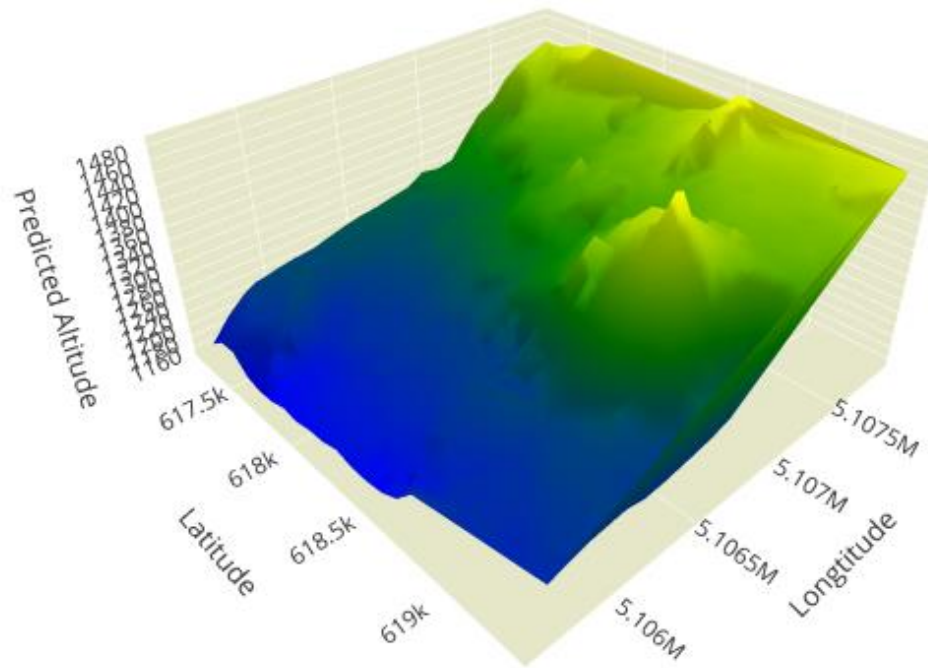
Figure 5.15 The 3D surface plot with additional interpolated points of SDVC with grid size of 4x4 in the mountainous area.
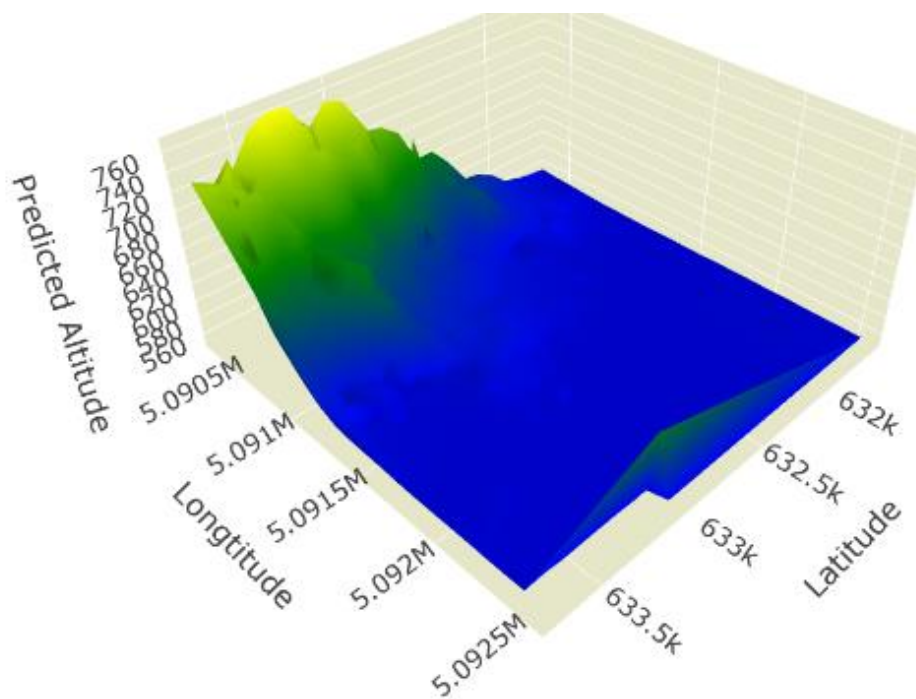

Figure 5.16 The 3D surface plot with additional interpolated points of SDVC with grid size of 4x4 in the flat area.

CHAPTER 6
CONCLUSION

This chapter describes the summary and discussion of the proposed algorithms studied in this research. Then, we conclude all contributions, limitations and future works.

6.1 Overall Discussion

We will discuss whether this research has met all of the objectives mentioned in Chapter 1. Referring to the first objective, it is to design novel algorithms for spatial interpolation techniques based on the kriging model. Our main contributions are the proposal of three novel kriging algorithms based on exponential semivariograms. We proposed three alternative kriging algorithms. The proposed kriging models are Exponential with Parameter Optimizer (EPO) model, Exponential with k-Iterations Optimizer (EKO) model, and Exponential with Polynomial-Trend Line (ETL) model. Our algorithms are easy to use because there is no need to specify nugget, sill, and range values. The three variants compute the optimal values of kriging parameters automatically and predict the unknown value at the unsampled location.

The second objective is to improve the accuracy of the kriging interpolation with our technique. The performance is evaluated by error reduction that eight models can perform. The strengths of each model are analyzed based on a different set of sample sizes coming from two zones of study areas. The resulting errors of our proposed methods are relatively small. The lower bounds of the 95% confidence interval of our models are mostly lower than all five contemporary models. However, in general, the result shows not many significant differences among models. The best result of our three methods appeared in Chapter 3.

The third objective is to implement the proposed algorithm and compare the performance with existing methods. Our result showed that the errors of our three methods are exceptionally small when comparing with the five conventional models, namely, Linear, Spherical, Pentaspherical, Gaussian, and classical Exponential models.

Statistically speaking, when the sample size is sufficiently large, it can be 95% certain that the EPO, EKO, and ETL models are more accurate than the Spherical, Pentaspherical, and Gaussian models.

The fourth objective is to improve the computational time of our proposed method. We presented two divide and conquer algorithms for the prediction of spatial heights. The first algorithm is called DVC. It contains six different kriging models to choose from. The evaluation revealed that the four exponential-based models (Exponential, EPO, EKO, and ETL) outperform the Gaussian and Spherical models in terms of accuracy. Among the exponential-based models, the two best performances are the EKO and Exponential models which are later used in the second algorithm called SDVC. In terms of the computational time, the DVC algorithm substantially reduced the waiting time from 7,705 to 33 minutes. Moreover, the SDVC algorithm can reduce the waiting time further down to 4 minutes.

The final objective is to apply the proposed algorithm to create a better 3D surface plot. The practical benefit of our results is in the area of improving the 3D surface plot as the number of unknown points can be predicted, so the plot can be filled with more points resulting in a more accurate plot. In addition, we found that the divide and conquer design paradigm truly helps improve both the accuracy (in the case of the Gaussian and Spherical models) and the running time (in the case of the EKO and Exponential models), resulting in a practical application for landscape 3D visual assessment.

6.2 Research Contribution

The research has proposed three novel algorithms for spatial interpolation methods using kriging models and presents two more algorithms with the aim of reducing the computational time of our proposed algorithms. All contributions in different aspects are summarized below.

6.2.1 Three alternative kriging algorithms

Three kriging algorithms were present in Chapter 3. These three variants compute the optimal values of kriging parameters automatically and predict the unknown value at the unsampled location. The proposed kriging models are Exponential with Parameter Optimizer (EPO) model, Exponential with k-Iterations Optimizer (EKO) model, and Exponential with Polynomial-Trend Line (ETL) model.

6.2.2 Divide and conquer design paradigm

The algorithms were present in Chapter 4. We proposed the divide and conquer design paradigm that can significantly reduce the computational time by mathematically decreasing the size of the distance matrix in our proposed models. Two algorithms are presented. The first algorithm divides the area into the smaller grid cell size and interpolates the points in each cell with our proposed methods, and the second algorithm improves the first one by introducing a slope parameter in the given cell.

6.2.3 New parameter of the terrain slope

We proposed a new parameter called β, a slope of terrain in the given cell which is calculated based on Equation (4.1) in Chapter 4. The slope is used to control the accuracy of the model in the SDVC algorithm. If the β value in a cell exceeds the specified slope (α), the algorithm will select the EKO model that works well in areas with high terrain variations.

6.2.4 Application

The program was presented in Chapter 5. The application is available as a web-based program where users can run and test their data online. It is an alternative choice for users who prefer a lightweight application that does not require installing additional packages or libraries.

6.3 Algorithms Limitations

Although the proposed algorithms for spatial interpolation methods using kriging models have a lot of advantages. There are some boundaries to the proposed algorithms. The limitations of the proposed algorithms are as follows.

1) The study area with large-scale data required a high-performance computer and waiting time to finish the model surface plot.

2) The original program is not supported or on the extension of the Geographic Information System (GIS) software.

3) The program is supported the maximum grid size of 4x4; some problems require more grid size to reduce the computational time by using our proposed algorithms.

6.4 Future Works

Novel algorithms for spatial interpolation techniques based on the kriging model still provide much potential to develop our area of research. The proposed future works are as follows.

1) To apply other kriging methods such as simple kriging and universal kriging for spatial interpolation techniques.

2) To combine the parameter estimation such as the maximum likelihood estimation technique into our algorithms.

3) To develop a system to create a better 3D surface plot for both large-scale and small-scale data.

4) To consider different parameter estimations for each grid cell.

REFERENCES

Ackere, S.V., Glas, H., Beullens, J., Deruyter, G., Wulf, A.D., and Maeyer, P.D., 2016. Development of a 3D dynamic flood WebGIS visualization tool. International journal of safety and security engineering, 6 (3), 560–569.

Aguilar, F.J., Agüera, F., Aguilar, M.A., and Carvajal, F., 2005. Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy. Photogrammetric Engineering and Remote Sensing, 71 (7), 805–816.

Al-Mamoori, S.K., Al-Maliki, L.A., Al-Sulttani, A.H., El-Tawil, K., and Al-Ansari, N., 2021. Statistical analysis of the best GIS interpolation method for bearing capacity estimation in An-Najaf City, Iraq. Environmental Earth Sciences, 80, 683.

Antal, A., Guerreiro, P.M.P., and Cheval, S., 2021. Comparison of spatial interpolation methods for estimating the precipitation distribution in Portugal. Theoretical and Applied Climatology, 145, 1193–1206.

Arundel, S.T., Archuleta, C.M., Phillips, L.A., Roche, B.L., and Constance, E.W., 2015. 1-Meter Digital Elevation Model Specification: U.S. Geological Survey Techniques and Methods, Book 11, chap. B7, 25 p. with appendixes.

Ajvazi, B., amd Czimber, K., 2019. A comparative analysis of different DEM interpolation methods in GIS: Case study of Rahovec Kosovo. Geodesy and Cartography, 45 (1), 43–48.

Arun, P. V., 2013. A comparative analysis of different DEM interpolation methods. The Egyptian Journal of Remote Sensing and Space Sciences, 16 (2), 133–139.

Azimi, H., Bonakdari, H., and Ebtehaj, I., 2019. Gene expression programming-based approach for predicting the roller length of a hydraulic jump on a rough bed. ISH Journal of Hydraulic Engineering, 1–11.

Bales, J.D., Wagner, C.R., Tighe, K.C., and Terziotti, S., 2007. Lidar-derived flood-inundation maps for real-time flood-mapping applications, Tar River Basin, North Carolina.Geological Survey (US), Reston, VA, USA.

Banjo, O.B., Salami, W.A., Adewale, R.O., and Adegoke, A., 2021. Spot height digital elevation model of Yewa division, Ogun State Nigeria. FUW Trends in Science & Technology Journal, J. 6 (2), 436–440.

Behzadi, S., and Jalilzadeh, A., 2021. Introducing a novel digital elevation model using artificial neural network algorithm. Civil Engineering Dimension, 22 (2), 47–51.

Bello-Pineda, J., and Hernandez-Stefanoni, J.L., 2007. Comparing the performance of two spatial interpolation methods for creating a digital bathymetric model of the Yucatan submerged platform. Pan-American Journal of Aquatic Sciences, 2 (3), 247–254.

Bhattacharyya, B., 2022. Uncertainty quantification of dynamical systems by a POD–kriging surrogate model. Journal of Scientific Computing, 60 (101602).

Bărbulescu, A., Serban, C., and Indrecan, M. L., 2021. Computing the beta parameter in IDW interpolation by using a genetic algorithm. Water, 13, 863.

Chen X, Hu T, Ren F, Chen D, Li L, and Gao N., 2014. Landscape analysis of geographical names in Hubei province, China. Entropy, 16(12), 6313-6337

Chilès, J.P., and Delfiner, P., 1999. Geostatistics: modeling spatial uncertainty. Wiley, New York.

Chilès, J.P., and Desassis, N., 2018. In: Daya Sagar, B.S., Cheng, Q., Agterberg, F. (Eds.), Fifty Years of Kriging. Handbook of Mathematical Geosciences; Fifty Years of IAMG. Springer, pp. 589–612.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., and Stein, C., 2009. Introduction to Algorithms, 3rd ed. MIT Press.

Cressie, N.A.C., 1993. Statistics for Spatial Data, revised edition. Wiley, New York.

Carollo, F.G., Ferro, V., and Pampalone, V., 2007. Hydraulic jumps on rough beds. Journal of Hydraulic Engineering, 133(9), 989–999.

Carollo, F.G., Ferro, V., and Pampalone, V., 2009. New solution of classical hydraulic jump. Journal of Hydraulic Engineering, 135(6), 527–531.

Delcev, S. and Draskovic, D., 2018. Modern JavaScript frameworks: A survey study. In 2018 Zooming Innovation in Consumer Technologies Conference (ZINC) (pp. 106-109). IEEE.

Deng, M., Yang, W., Liu, Q., and Zhang, Y., 2017. A divide-and-conquer method for space-time series prediction. Journal of Geographical Systems, 19 (1), 1–19.

Du, S., Deng, Q., 2021. Unscented particle filter algorithm based on divide-and-conquer sampling for target tracking. Sensors. 21 (6), 2236.

Diggle, P. J., and Ribeiro, P. J., 2007. Model-based Geostatistics. New York: Springer.

Elewa, H.H., and Qaddah, A.A., 2011. Groundwater potentiality mapping in the Sinai peninsula, Egypt, using remote sensing and GIS-watershed-based modeling. Hydrogeology Journal, 19 (3), 613–628.

Guisande, C., Rueda-Quecho, A.J., Rangel-Silva, F.A., Heine, J., García-Rosell´o, E., Gonz´alez-Dacosta, J., Gonz´alez-Vilas, L., Pelayo-Villamil, P., 2019. SINENVAP: an algorithm that employs kriging to identify optimal spatial interpolation models in polygons. Ecological Informatics, 53, 100975.

Günlü, A., Bulut, S., Keles¸, S., and Ercanlı, ˙I., 2019. Evaluating different spatial interpolation methods and modeling techniques for estimating spatial forest site index in pure beech forests: a case study from Turkey. Environmental Monitoring and Assessment, 192 (1), 53.

Gao, Y., Shen, X., and Salam, M. A., 2002. Global differential GPS positioning without using a base station. Geographic Information Systems, 8 (1), 9–15.

Hao, P., Feng, S., Li, Y., Wang, B., Chen, H., 2020. Adaptive infill sampling criterion for multi-fidelity gradient-enhanced kriging model. Struct. Multidiscip. Optim. 62, 353–373.

Hao, P., Feng, S., Liu, H., Wang, Y., Wang, B., Wang, B., 2021. A novel nested stochastic kriging model for response noise quantification and reliability analysis. Comput.Methods Appl. Mech. Eng. 384 (113941).

Hasan, K., Paul, S., Chy, T.J., and Antipova, A., 2021. Analysis of groundwater table variability and trend using ordinary kriging: the case study of Sylhet, Bangladesh. Applied Water Science, 11, 120.

Hasanipanah, M., Meng, D., Keshtegar, B., Trung, N.T., and Thai, D.K., 2021. Nonlinear models based on enhanced kriging interpolation for prediction of rock joint shear strength. Neural Computing and Applications, 33 (9), 4205–4215.

Iwahashi, J., and Yamazaki, D., 2022. Global polygons for terrain classification divided into uniform slopes and basins. Progress in Earth and Planetary Science, 9, 33.

Jiang L, Ling D, Zhao M, Wang C, Liang Q, Liu K., 2018. Effective identification of terrain positions from gridded DEM data using multimodal classification integration. ISPRS International Journal of Geo-Information, 7(11), 443.

Kaya, E., Agca, M, Adiguzel, F., and Cetin, M., 2019 Spatial data analysis with R programming for environment, Human and Ecological Risk Assessment: An International Journal, 25 (6), 1521-1530

Keler, A., and Mazimpaka, J.D., 2016. Safety-aware routing for motorised tourists based on open data and VG. Journal of Location Based Services, 10 (1), 64–77.

Krajewski, W.F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilllla, R., Niemeier, J., Quintero, F., and Seo, B.C., 2017. Real-time flood forecasting and information system for the state of Iowa. Bulletin of the American Meteorological Society, 98, 539–554.

Kredpattanakul, K. and Limpiyakorn, Y., 2018, June. Transforming javascript-based web application to cross-platform desktop with electron. In International Conference on Information Science and Applications (pp. 571-579). Springer, Singapore.

Lam, N.S.N., 1983. Spatial interpolation methods: a review. The American Cartographer, 10 (2), 129–149.

Li, J., and Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecological Informatics, 6 (3–4), 228–241.

Liang, J., and Zhoub, M., 2010. A geospatial model of forest dynamics with controlled trend surface. Ecological Modelling, 221 (19), 2339-2352.

Ma, J., Li, X., Jia, B., Liu, X., Li, T., Zhang, W., and Liu, W., 2021. Spatial variation analysis of urban forest vegetation carbon storage and sequestration in built-up areas of Beijing based on i-tree eco and kriging. Urban Forestry and Urban Greening, 66 (19), 127413.

Meng, J., 2021. Raster data projection transformation based-on kriging interpolation approximate grid algorithm. Alexandria Engineering Journal, 60 (2), 2013–2019.

Mert, B.A., Dag, A., 2017. A computer program for practical semivariogram modeling and ordinary kriging: a case study of porosity distribution in an oil field. Open Geosciences, 9 (1), 663–674.

Mitas, L., and Mitasova, H., 2005. Spatial interpolation in geographical information systems: Principles, techniques, management and applications (2nd ed.). Wiley, New York: Abridged.

Mohd Aziz, M. K. B., Yusof, F., Mohd Daud, Z., Yusop, Z., and Kasno, M. A., 2019. Comparison of semivariogram models in rain gauge network design. MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics, 35(2), 157–170.

Natsagdorj, B., Dalantai, S., Sumiya, E., Bao, Y., Bayarsaikhan, S., Batsaikhan, B., and Ganbat, D., 2021. Assessment of some meteorology data of average monthly air temperature over Mongolia using digital elevation model (DEM) and GIS techniques. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4–2021, pp. 117–121.

Nie, S., Bian, J., and Zhou, Y., 2021. Estimating the spatial distribution of soil salinity with geographically weighted regression kriging and its relationship to groundwater in the western Jilin irrigation area, Northeast China. Polish Journal of Environmental Studies, 30 (1), 283–294.

O'Sullivan, D., and Unwin, D., 2010. Geographical Information Analysis (Second Edition). USA, John Wiley and Sons: Hoboken.

Pan, J. X., and Fang, K. T., 2002. Maximum likelihood estimation. In Growth curve models and statistical diagnostics (pp. 77-158). Springer, New York, NY.

Ranjan, A., Sinha, A., and Battewad, R., 2020. JavaScript for modern web development: building a web application using HTML, CSS, and JavaScript. BPB Publications.

Reznik, T., Lukas, V., Charvat, K., Charvat Jr., K., Krivanek, Z., Kepka, M., Herman, L., and Reznikova, H., 2017. Disaster risk reduction in agriculture through geospatial (big) data processing. International Journal of Geographical Information Science, 6 (8), 238.

Robinson, T. P., and Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and Electronics in Agriculture, 50(2), 97–108.

Setianto, A., and Triandini, T., 2013. Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. Journal of Southeast Asian Applied Geology, 5 (1), 21–29.

Sukkuea, A., Heednacram, A., 2022. Prediction on spatial elevation using improved kriging algorithms: an application in environmental management. Expert Syst. Appl. 207 (117971).

Shi, C., and Wang, Y., 2021. Non-parametric machine learning methods for interpolation of spatially varying non-stationary and non-Gaussian geotechnical properties. Geoscience Frontiers, 12 (1), 339–350.

Simpson, G., and Wu, Y.H., 2014. "Accuracy and effort of interpolation and sampling: can GIS help lower field costs? ISPRS International Journal of Geo-Information. 3, 1317-1333.

Tran, Q. B., and Nguyen, T. T., 2008. Assessment of the influence of interpolation techniques on the accuracy of digital elevation model. VNU Journal of Science, Earth Sciences, 24, 176-183.

Taylor, J.R., 1999. An introduction to error analysis: the study of uncertainties in physical measurements. University Science Books. p. 94.

Thanoon, S. R., 2018. Application trend surface models with estimation. Tikrit Journal of Pure Science, 23 (10), 118–122.

Voltz, M., and Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. European Journal of Soil Science, 41(3), 473–490.

Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Applications, 3rd ed. Springer, Berlin.

Wang, L., Zhang, J., Wang, O., Lin, Z., Lu, H., 2020. June. SDC-depth: Semantic divide and conquer network for monocular depth estimation. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 538–547.

Wang, J., Zhao, M.W., Jiang, L., Yang, C.C., Huang, X.L., Yan Xu, Y., and Lu, J., 2021. A new strategy combined HASM and classical interpolation methods for DEM construction in areas without sufficient terrain data. Journal of Mountain Science, 18 (10), 2761–2775.

Webster, R., and Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 2nd ed. John Wiley & Sons Ltd, England.

Yanto, Apriyono, A., Santoso, P.K., and Sumiyanto, 2022. Landslide susceptible areas identification using IDW and ordinary kriging interpolation techniques from hard soil depth at middle western Central Java, Indonesia. Natural Hazards, 110 (2), 1405–1416.

Yin, J., Yu, D., and Wilby, R., 2016. Modelling the impact of land subsidence on urban pluvial flooding: a case study of downtown Shanghai, China. Science of the Total Environment, 544, 744–753.

Yu, Z., Song, Y., Song, D., and Liu, Y., 2021. Spatial interpolation-based analysis method targeting visualization of the indoor thermal environment. Building and Environment, 18, 107484.

Zerdoumi, S., Hashem, I.A.T. and Jhanjhi, N.Z., 2022. A new spatial spherical pattern model into interactive cartography pattern: multi-dimensional data via geostrategic cluster. Multimedia Tools and Applications, 81, 22903–22952.

APPENDICES

APPENDICES A

The 3D surface plot of 8 models for 31 and points in zone 1 and 2
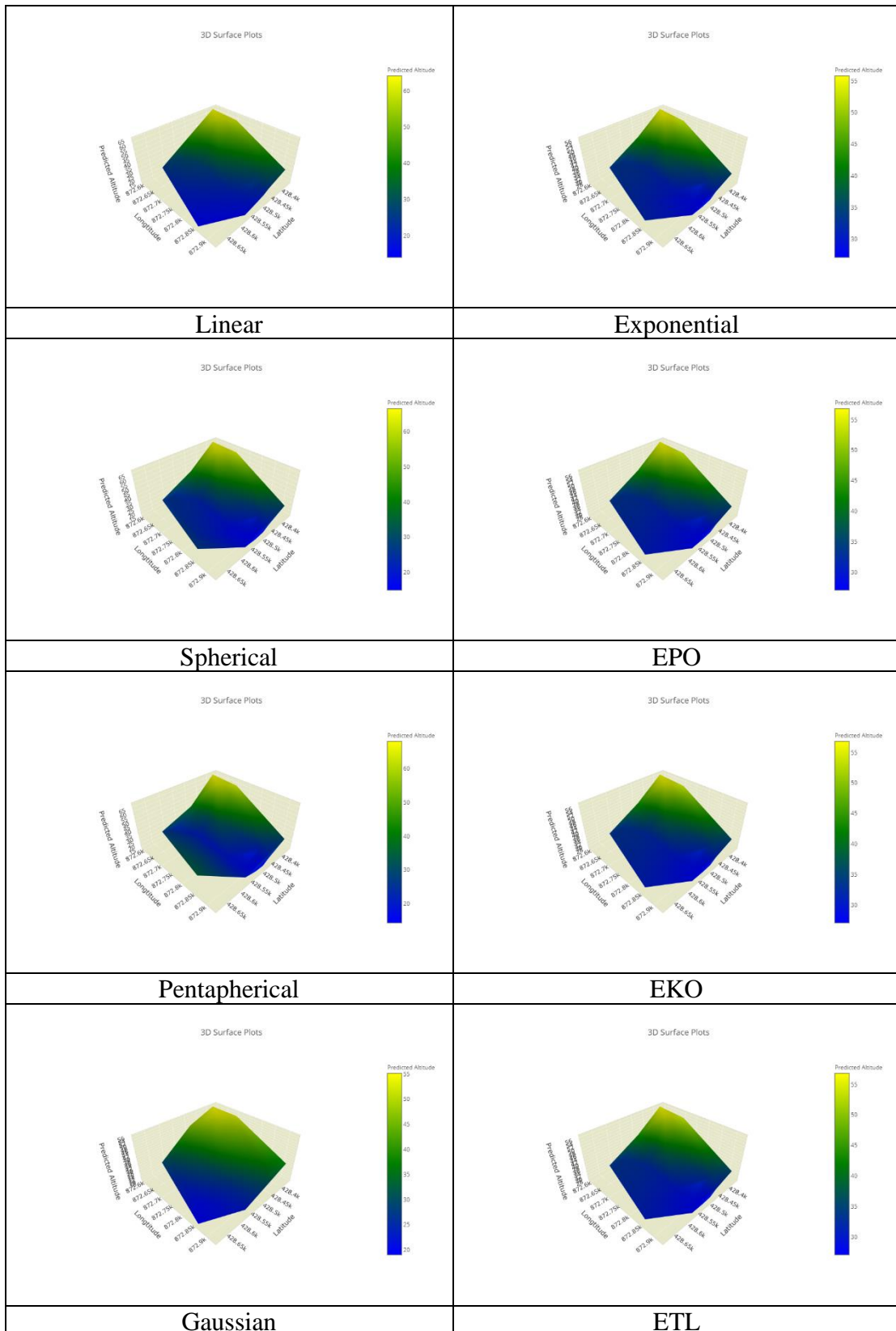
Figure A1 The 3D surface plot of 8 models for 31 points in zone 1.
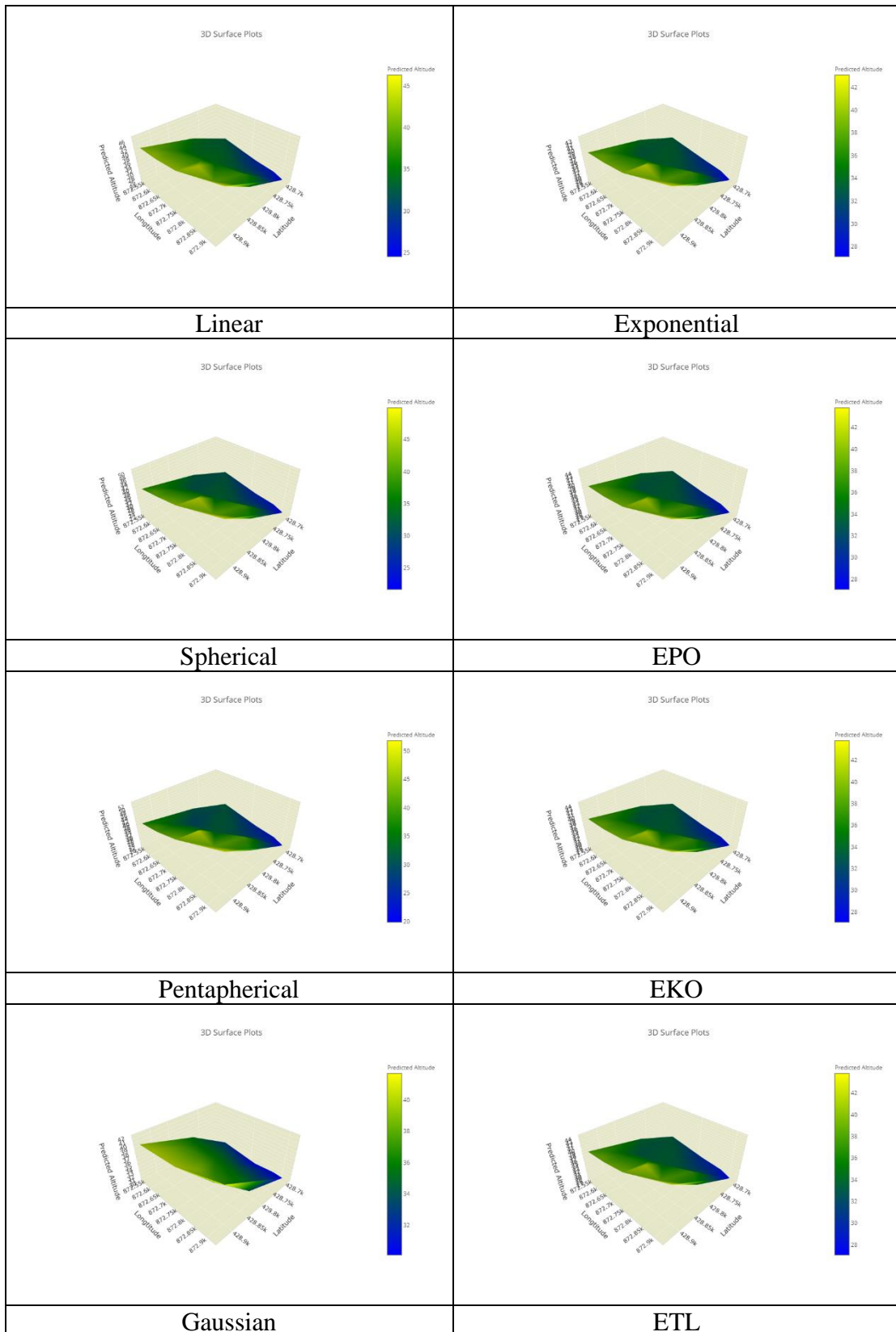
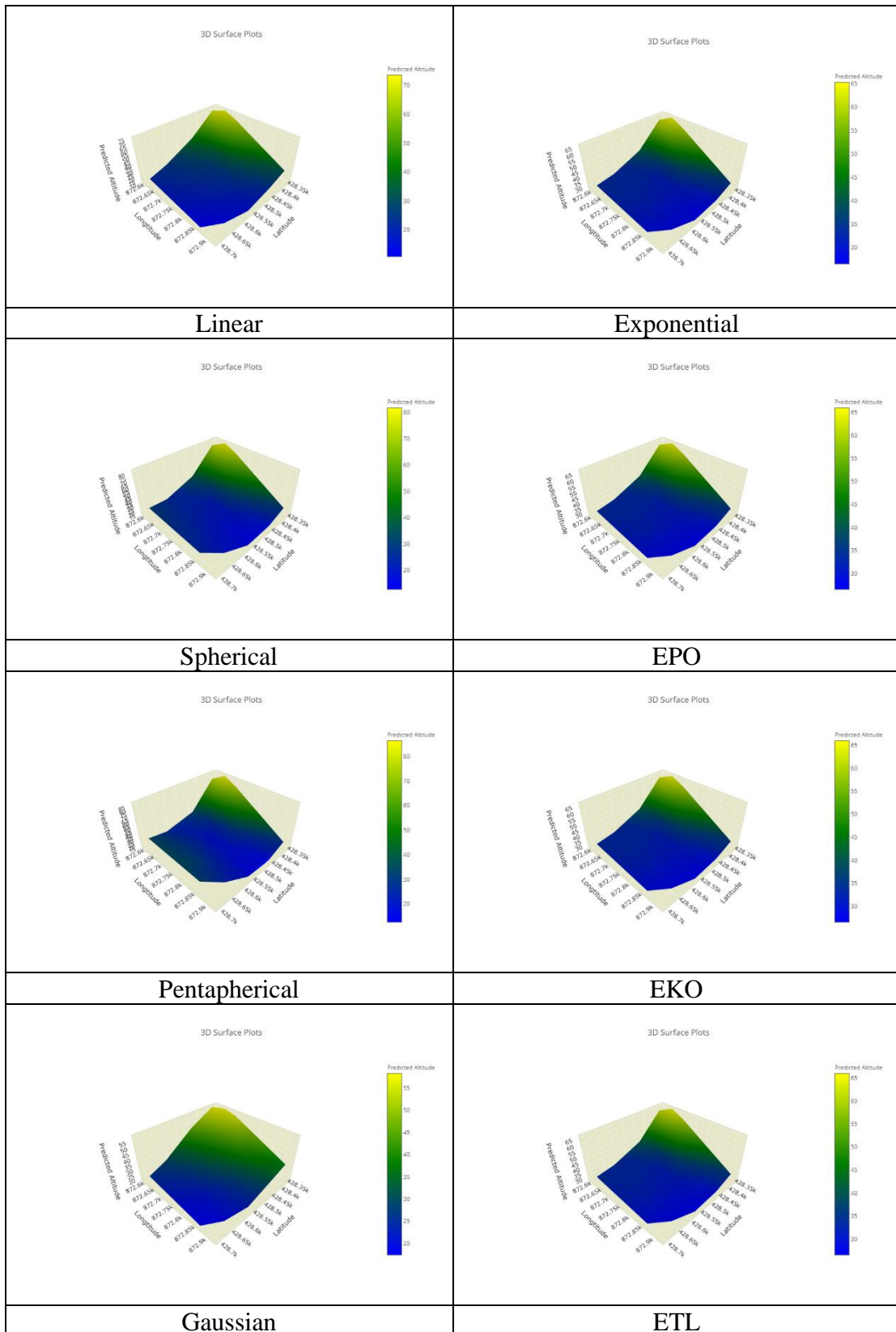**Figure A2** The 3D surface plot of 8 models for 31 points in zone 2.

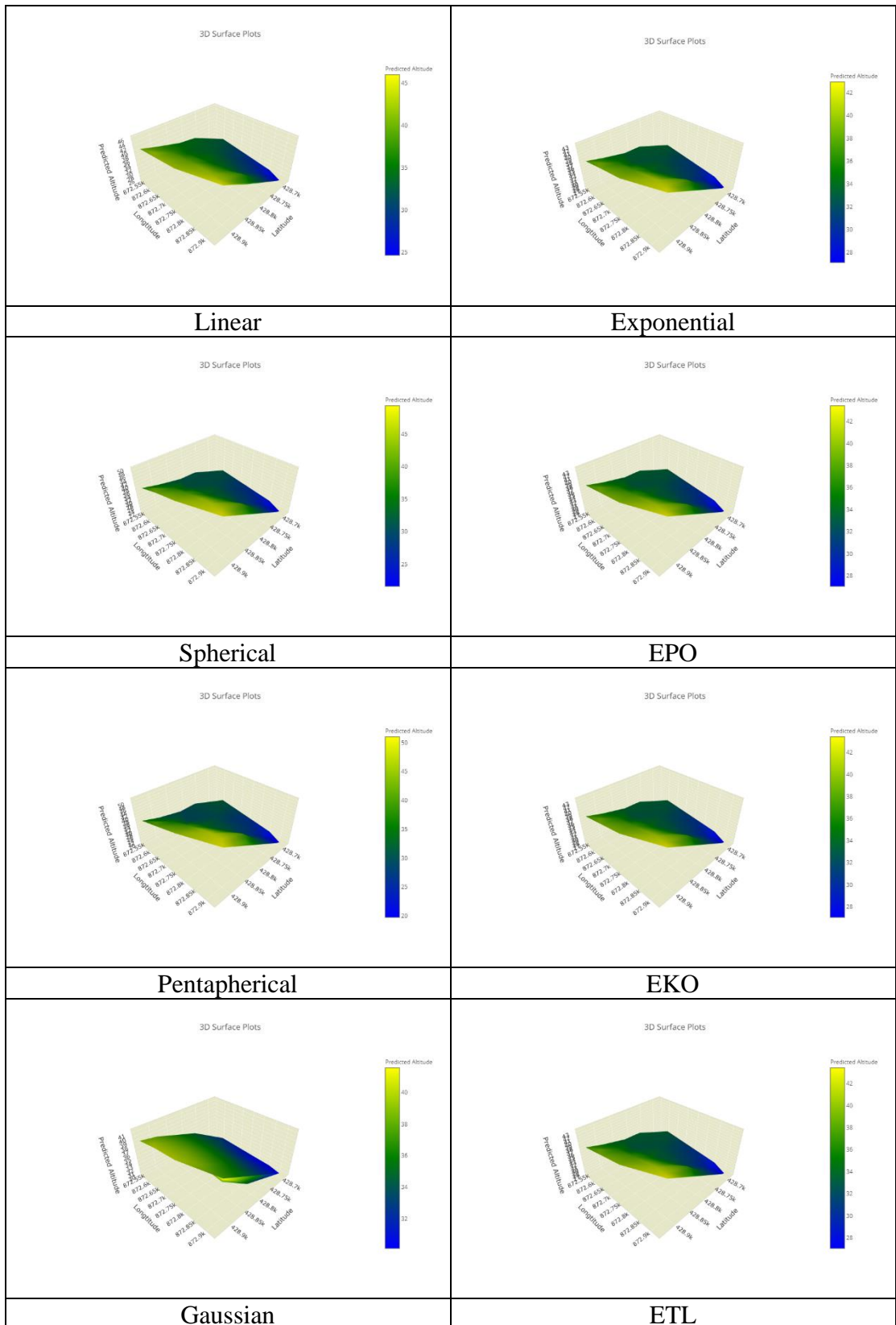**Figure A3** The 3D surface plot of 8 models for 51 points in zone 1.

Figure A4 The 3D surface plot of 8 models for 51 points in zone 2.

APPENDICES B

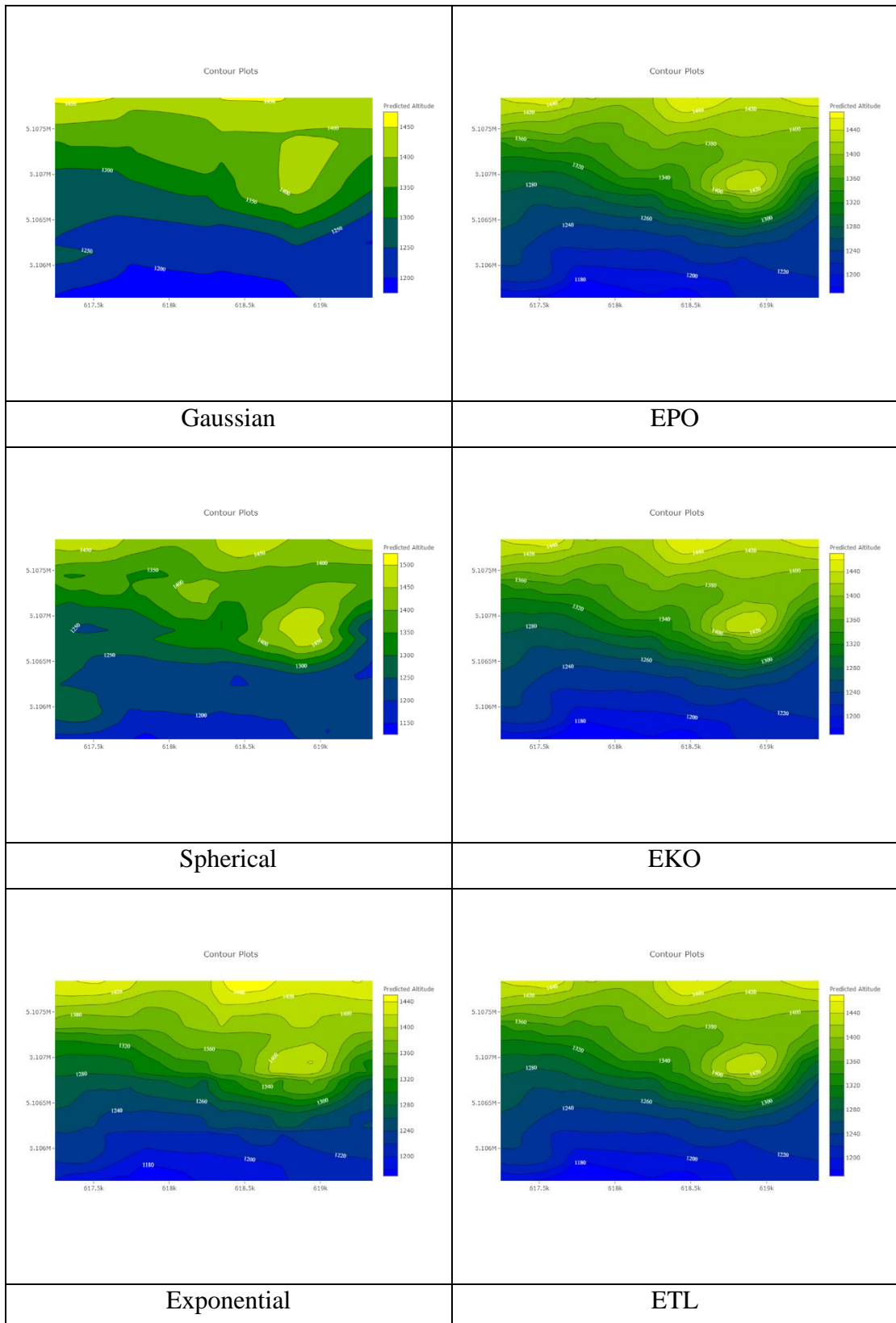**The contour from DVC with grid size of 1x1, 2x2, and 3x3 in the mountainous and flat area.**

Figure B1 The contour from DVC with grid size of 1x1 in the mountainous area.
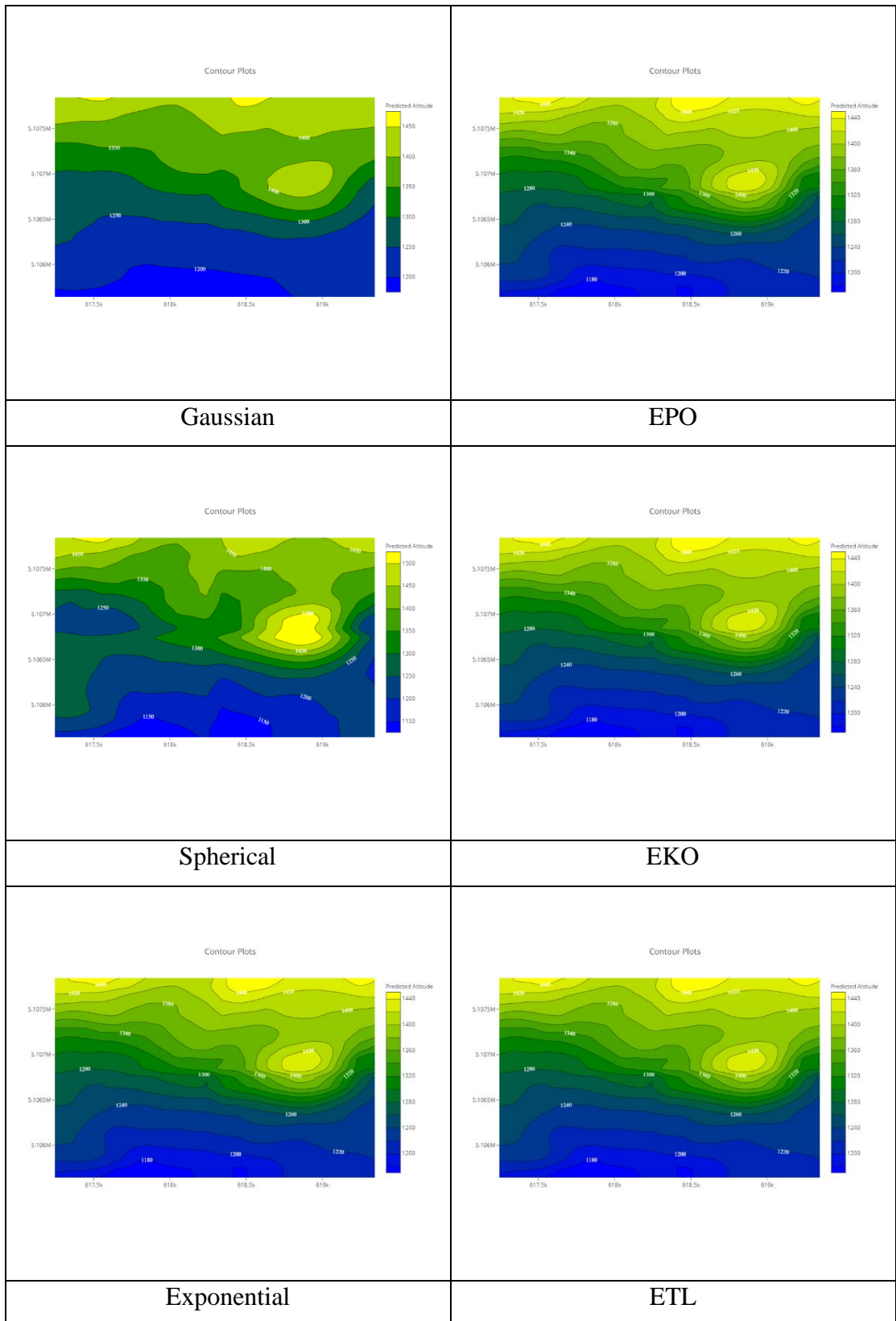
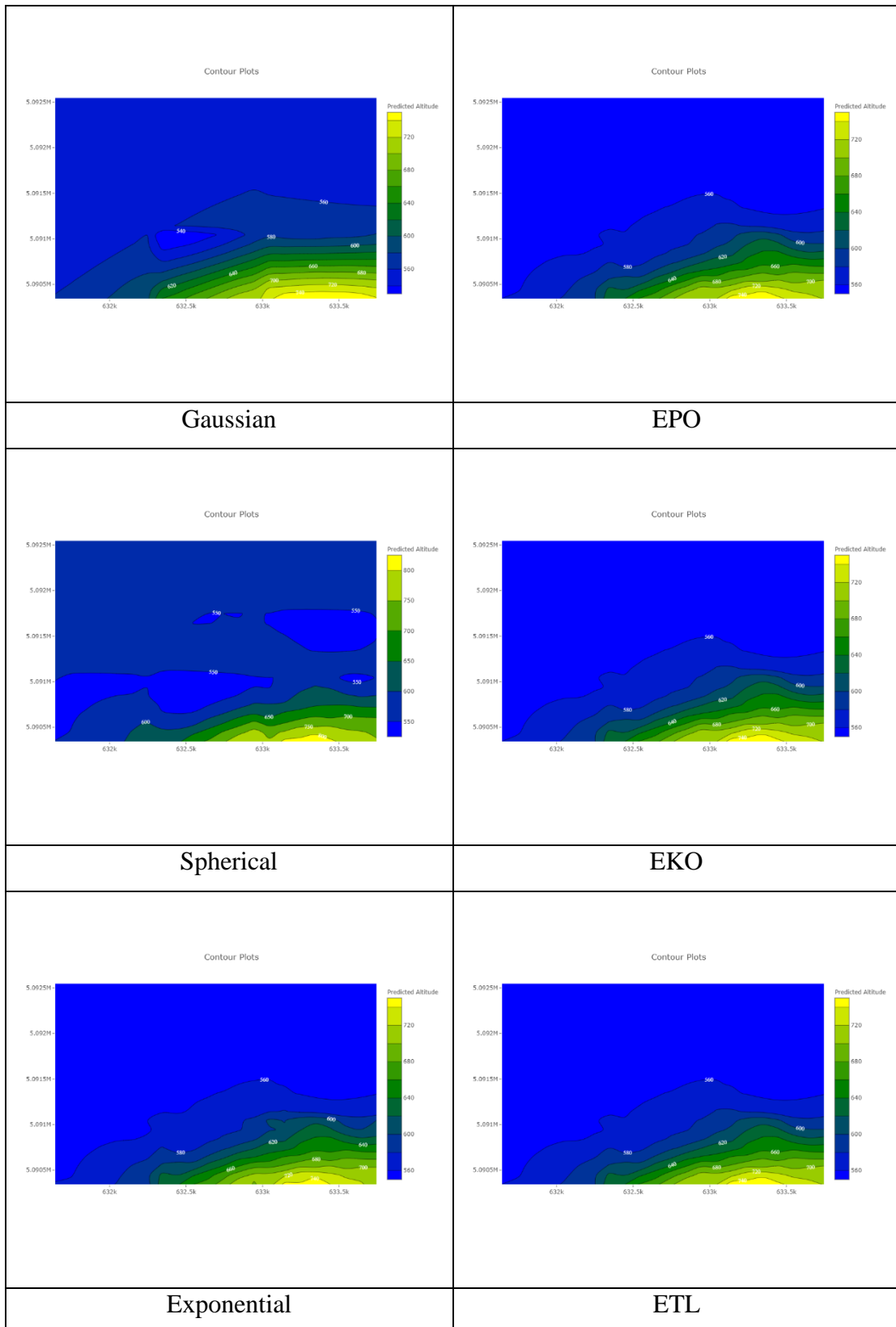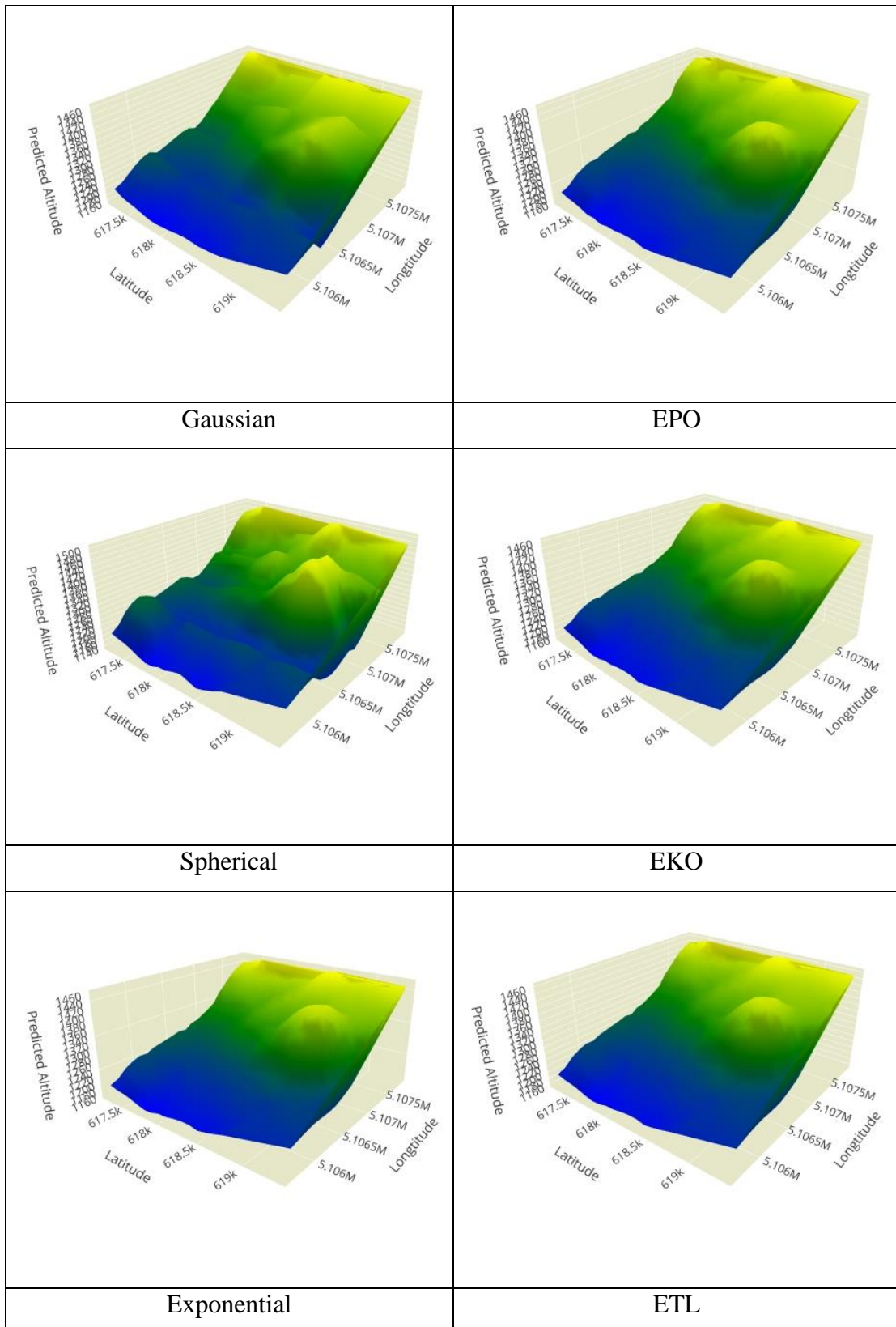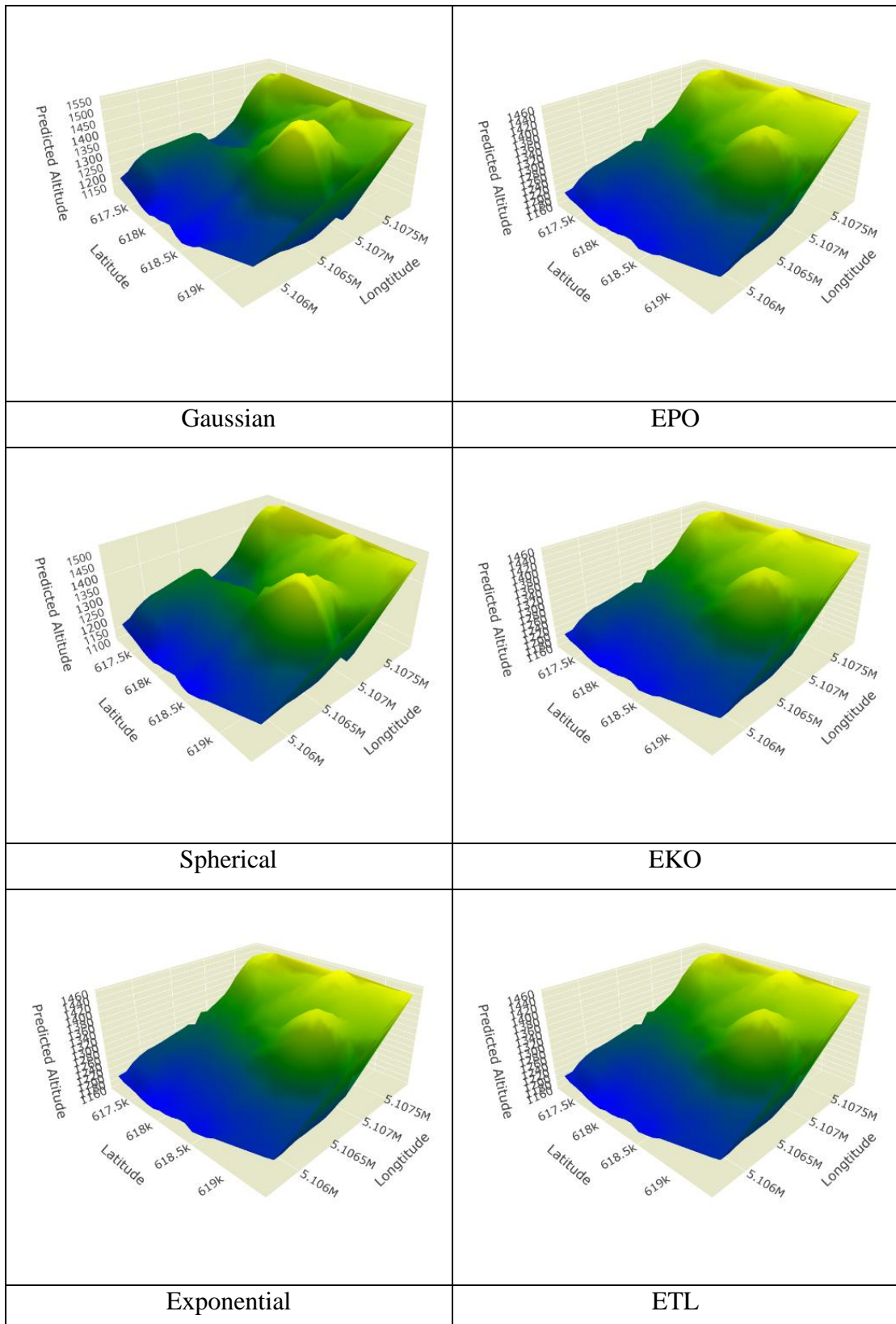**Figure B2** The contour from DVC with grid size of 2x2 in the mountainous area.

Gaussian	EPO
Spherical	EKO
Exponential	ETL
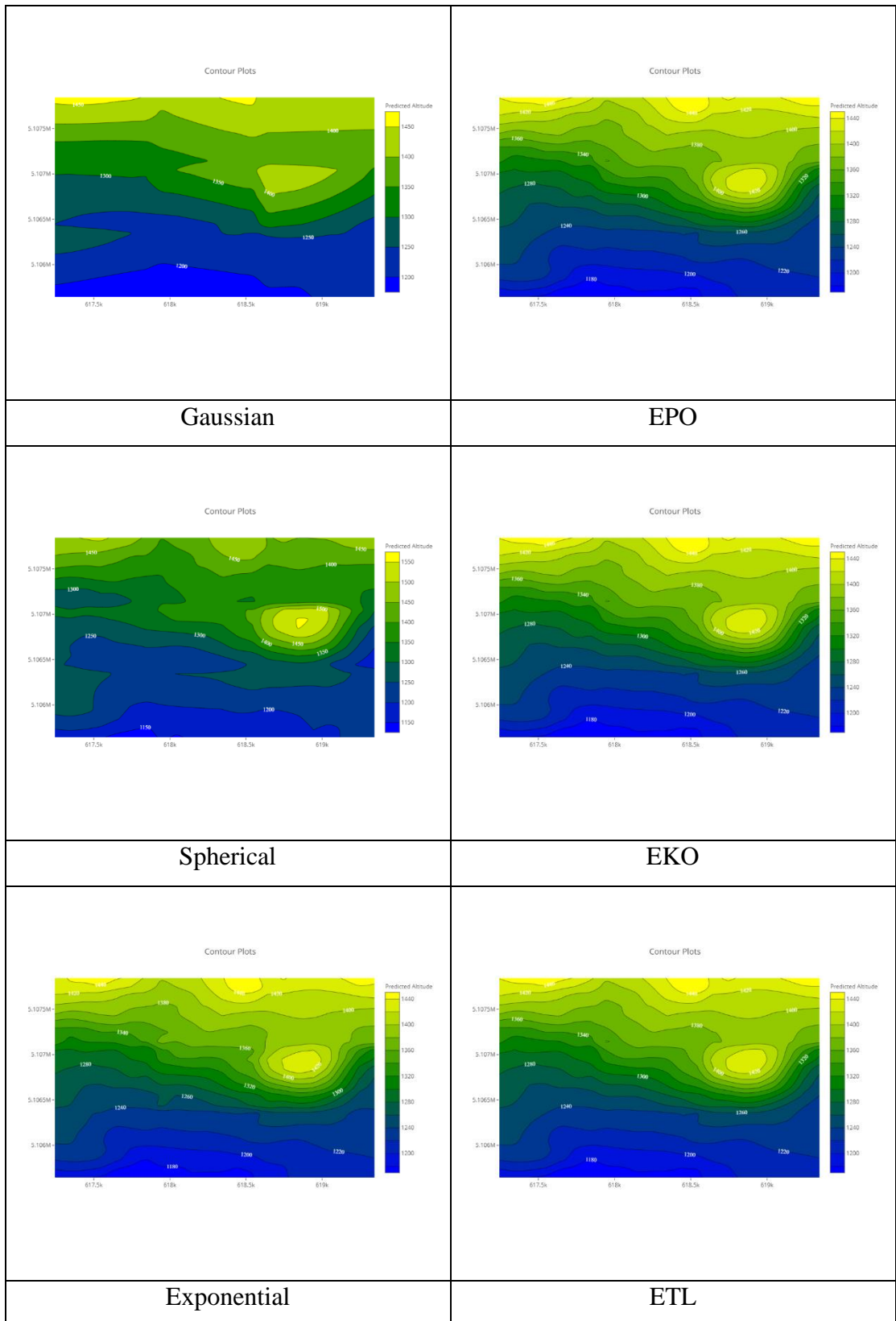
Figure B3 The contour from DVC with grid size of 3x3 in the mountainous area.
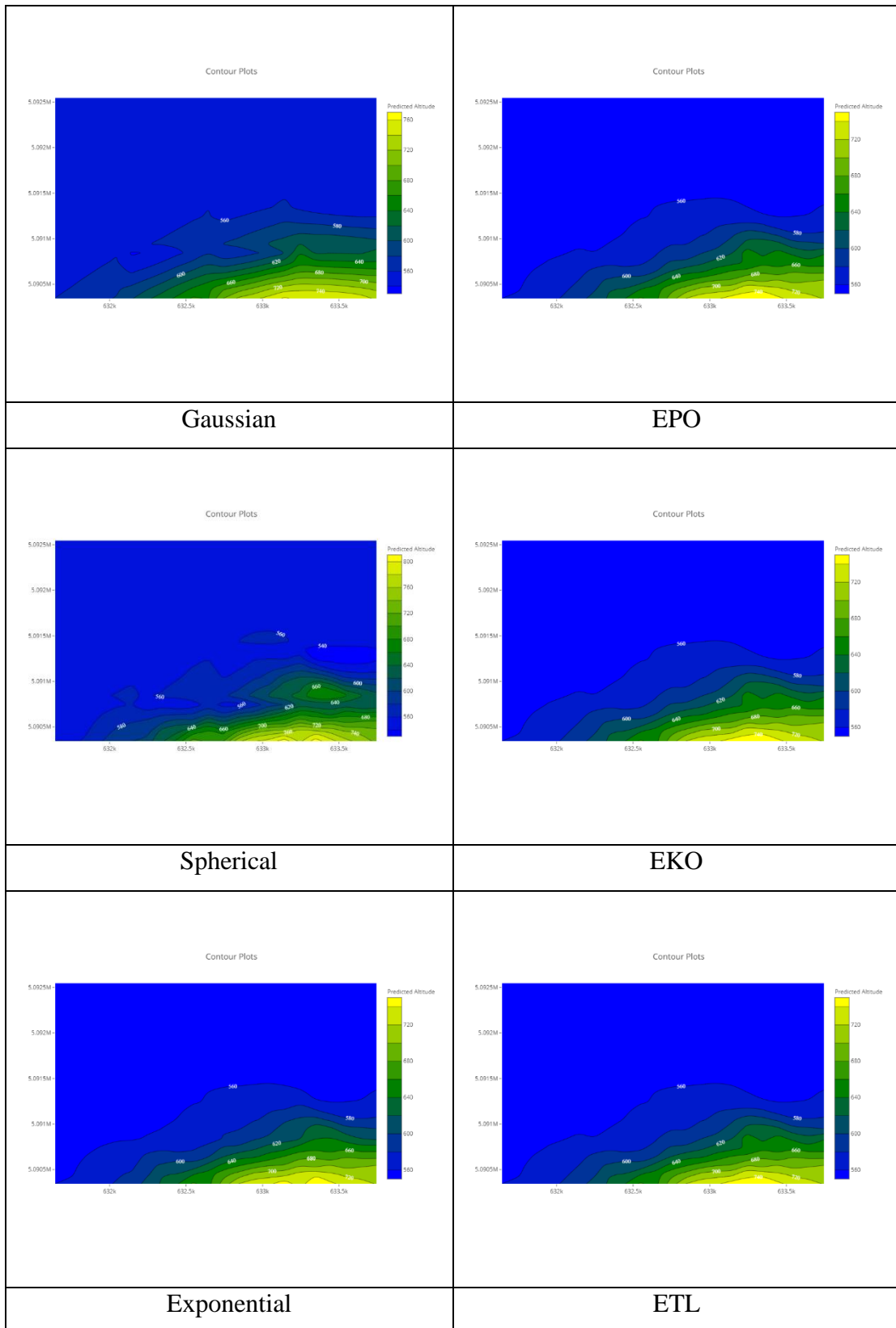
Figure B4 The contour from DVC with grid size of 1x1 in the flat area.

Gaussian	EPO
Spherical	EKO
Exponential	ETL
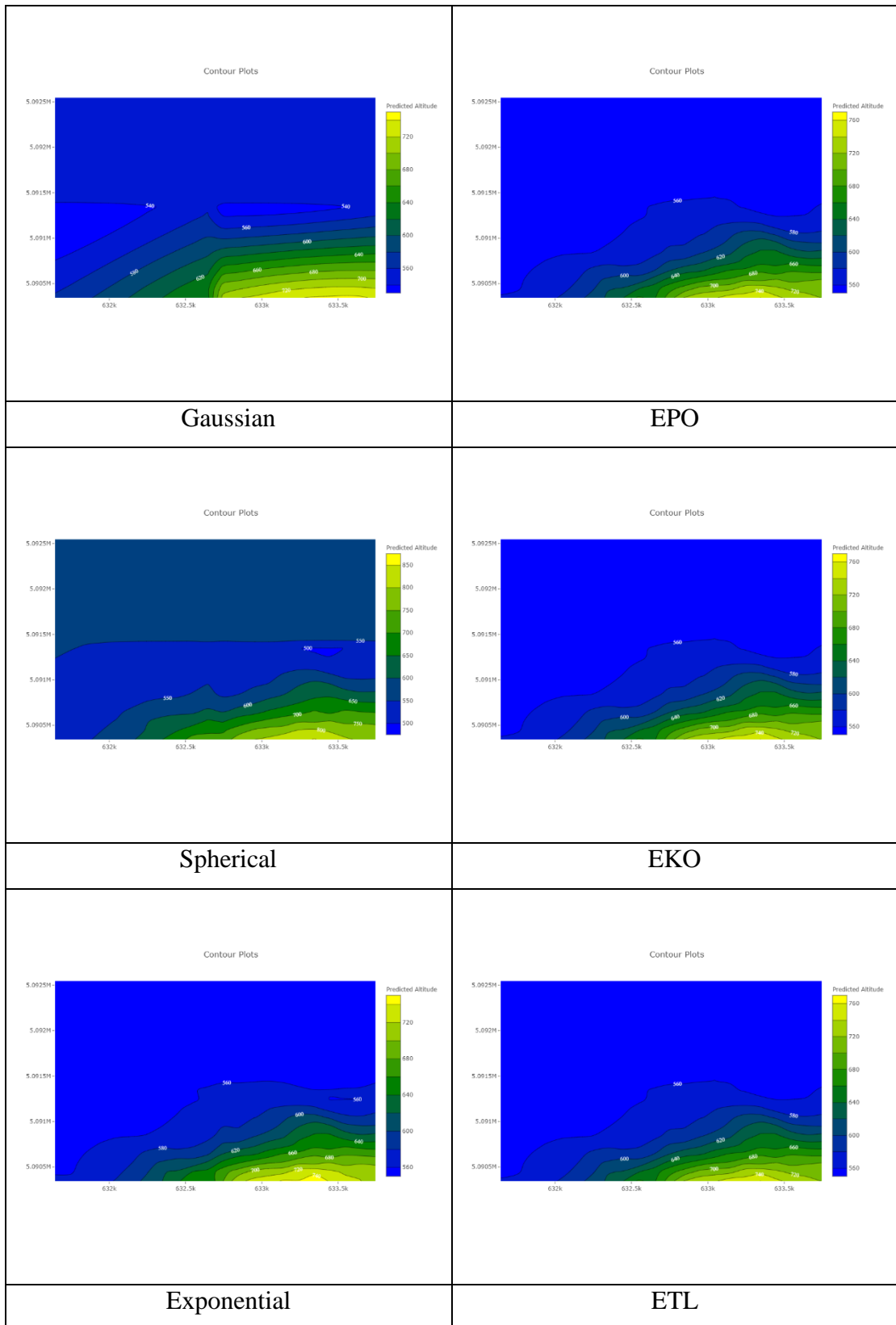
Figure B5 The contour from DVC with grid size of 2x2 in the flat area.

Figure B6 The contour from DVC with grid size of 3x3 in the flat area.

APPENDICES C

The 3D surface plot from DVC with grid size of 1x1, 2x2, and 3x3 in the mountainous and flat area

Figure C1 The 3D surface plot from DVC with grid size of 1x1 in the mountainous area.

Figure C2 The 3D surface plot from DVC with grid size of 2x2 in the mountainous area.
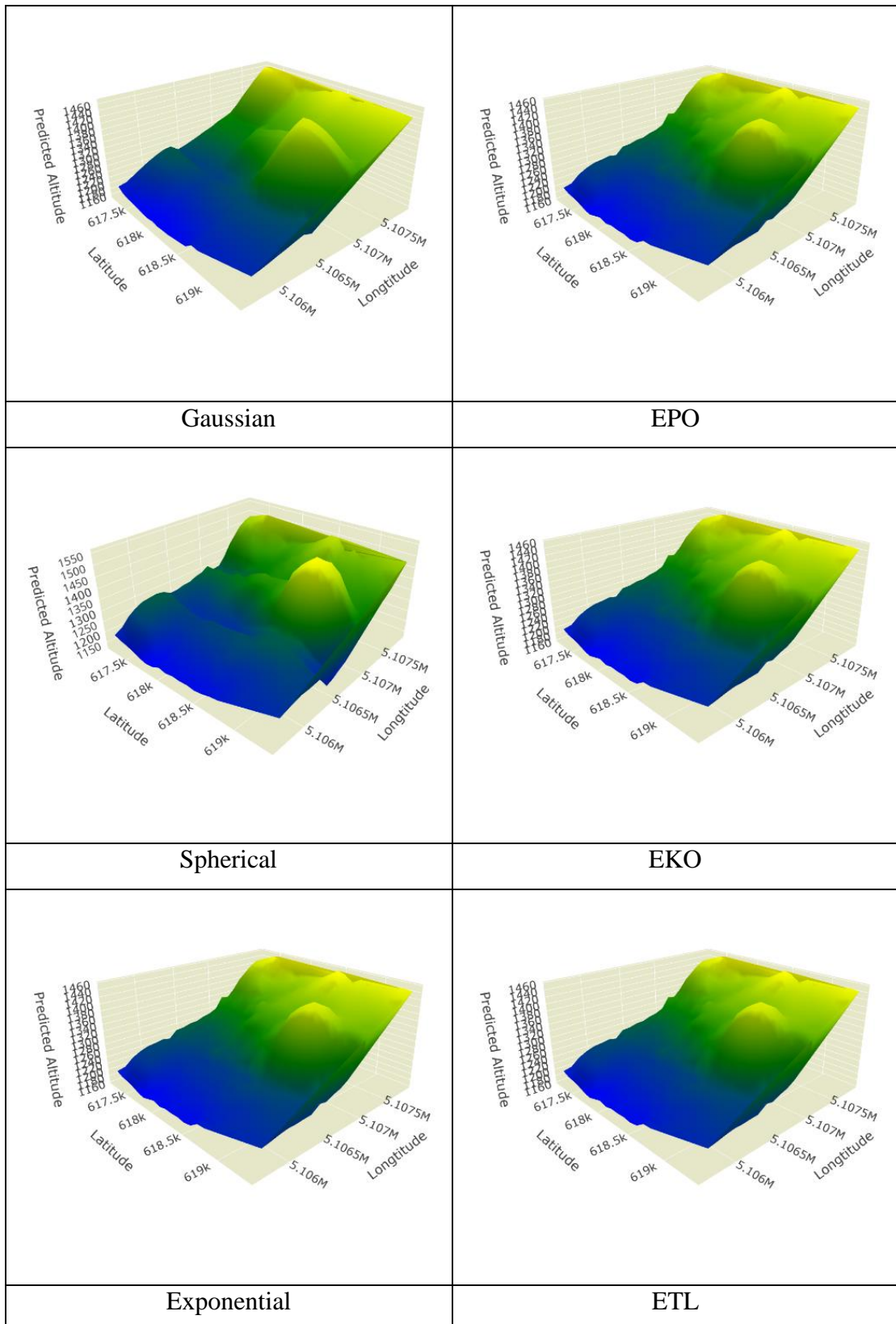
Figure C3 The 3D surface plot from DVC with grid size of 3x3 in the mountainous area.
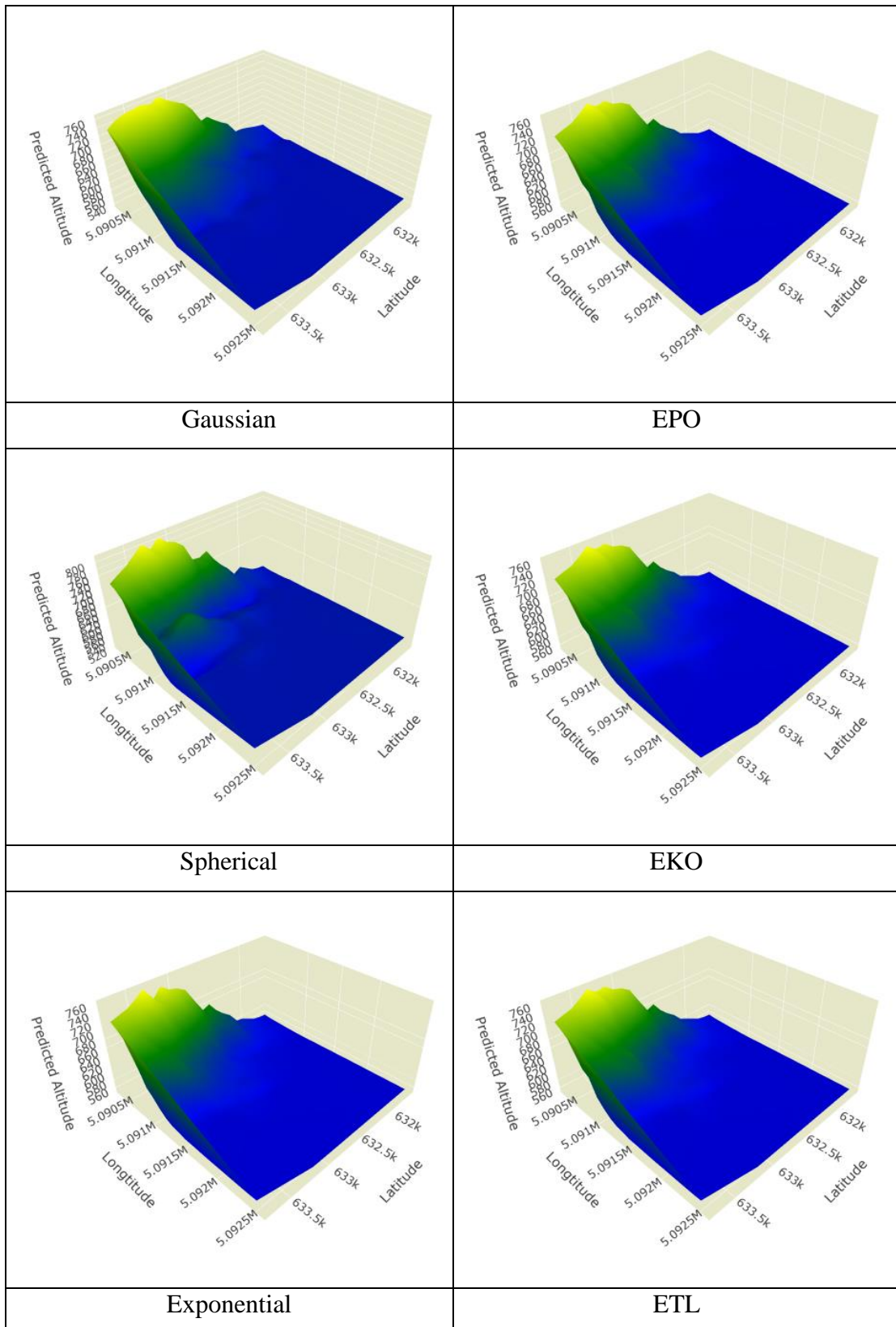
Gaussian

EPO

Spherical

EKO

Exponential

ETL

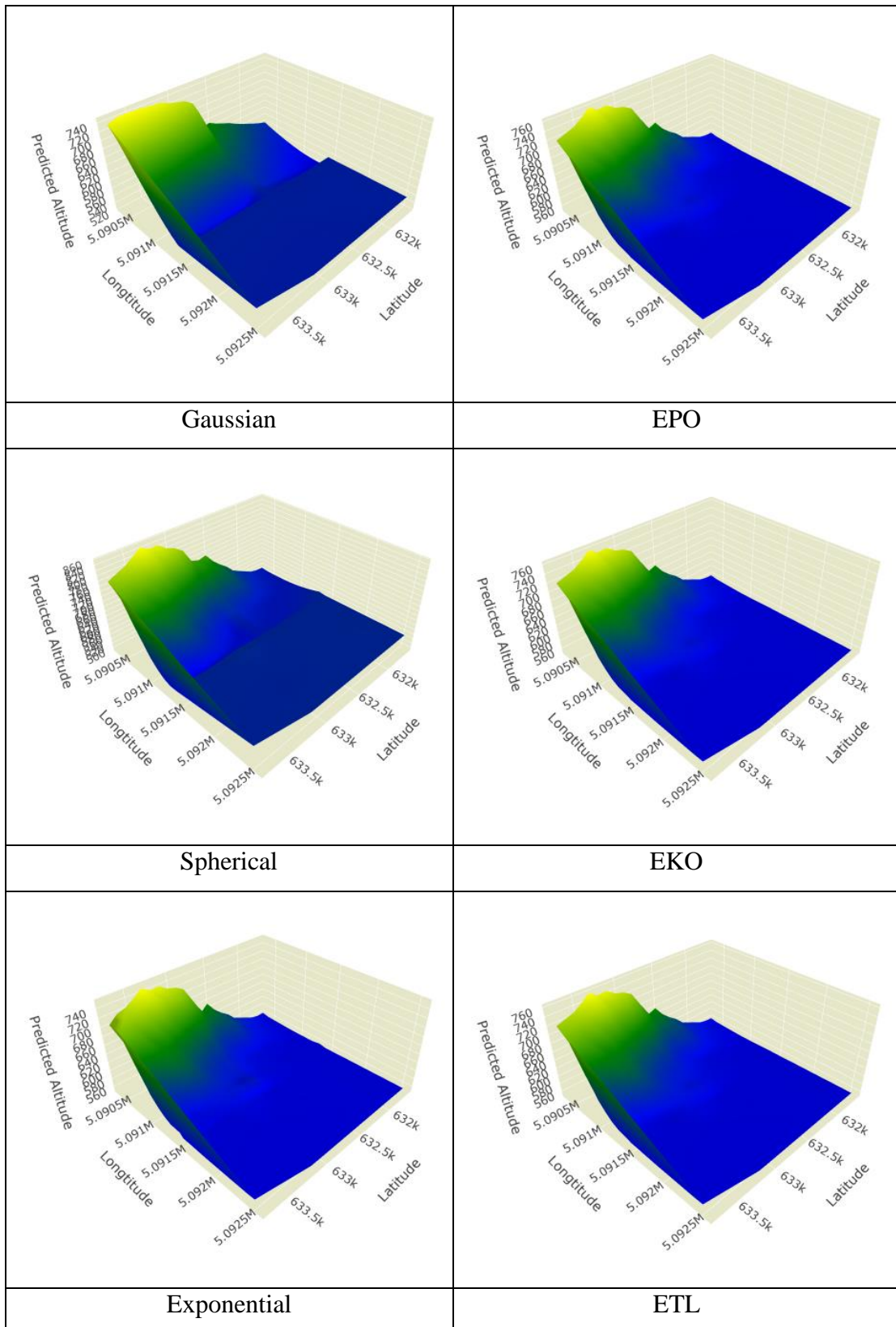**Figure C4** The 3D surface plot from DVC with grid size of 1x1 in the flat area.

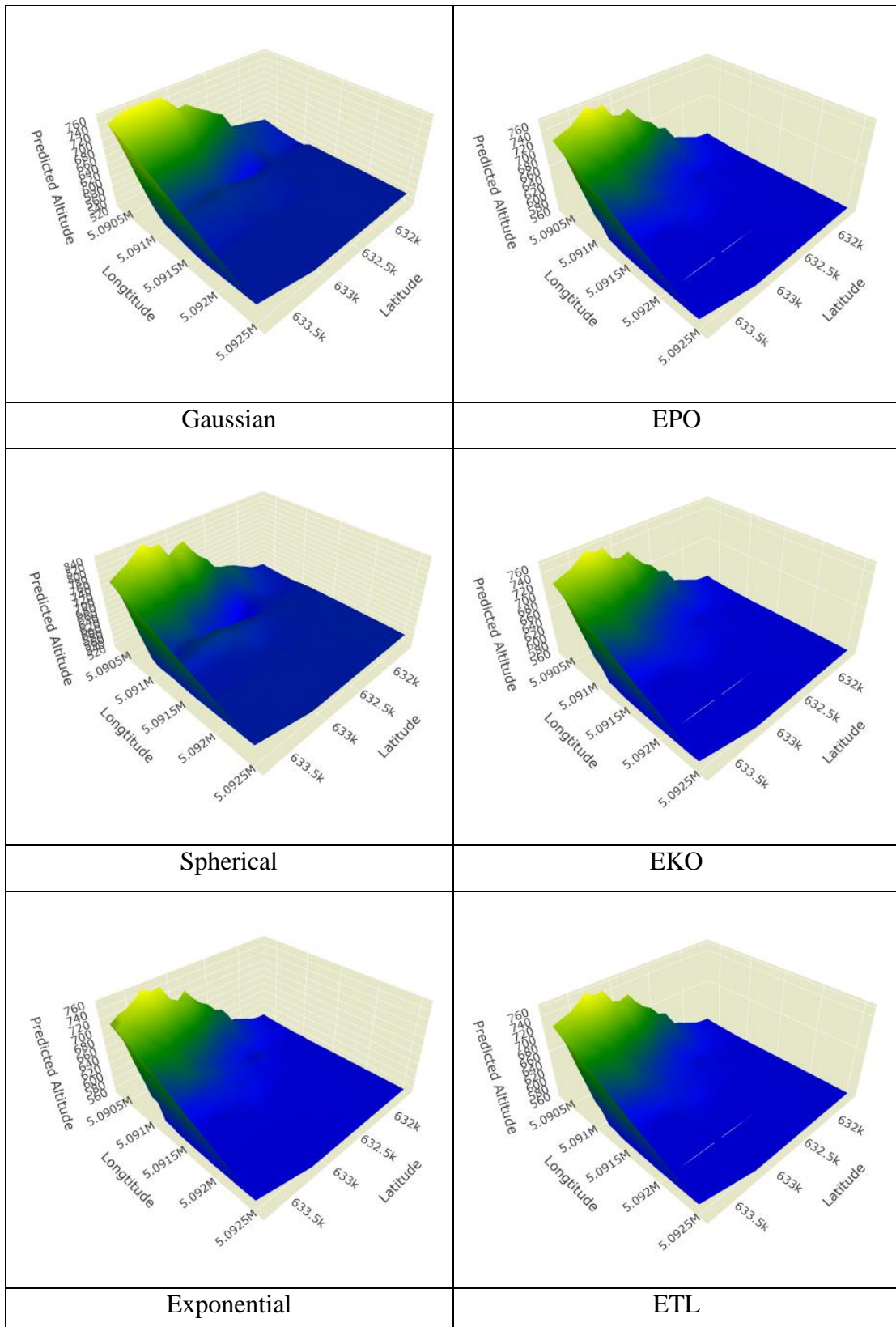**Figure C5** The 3D surface plot from DVC with grid size of 2x2 in the flat area.

104

Figure C6 The 3D surface plot from DVC with grid size of 3x3 in the flat area.

VITAE

Name   Arsanchai Sukkuea

Student ID   6130231001

Educational Attainment

Degree	Name of Institution	Year of Graduation
Bachelor of Engineering (Mechanical Engineering)	Kasetsart University	2006
Master of Engineering (Mechanical Engineering)	Chulalongkorn University	2009

List of Publication

Sukkuea, A., and Heednacram, A., 2022. Prediction on Spatial Elevation Using Improved Kriging Algorithms: An Application in Environmental Management. Expert Systems with Applications, 207(117971), 1-10. https://doi.org/10.1016/j.eswa.2022.117971.

Sukkuea, A., and Heednacram, A., 2022. Practical Kriging Models with Divide and Conquer Algorithms for Spatial Heights Forecast. Ecological Informatics, 70(101756), 1-11. https://doi.org/10.1016/j.ecoinf.2022.101756.