Semivariogram fitting based on SVM and GPR for DEM interpolation

To cite this article: A Setiyoko et al 2019 IOP Conf. Ser.: Earth Environ. Sci. 311 012076

View the article online for updates and enhancements.
Semivariogram fitting based on SVM and GPR for DEM interpolation

A Setiyoko¹², A M Arymurthy¹, T Basaruddin¹ and R Arief²
¹Faculty of Computer Science, University of Indonesia, Jl. Margonda Raya, Pondok Cina, Beji, Depok 16424, Indonesia
²Remote Sensing Data Centre-LAPAN, Jl. Lapan No.70 Jakarta Timur, Jakarta 13710, Indonesia
E-mail: andie.setiyoko@ui.ac.id

Abstract. DEM (Digital Elevation Model) as a digital model of the earth’s surface elevation could be generated from remote sensing technology such as stereo imaging for various applications. To generate DEM from stereo imagery, interpolation or approximation process stage is required. Stochastic interpolation e.g. ordinary kriging uses semivariogram fitting to calculate weights of interpolation values based on known points. This research is applying regression types of machine learning for semivariogram fitting to interpolate DEM. Previous research conducted was LS-SVM (Least Square-Support Vector Machine) and SVR (Support Vector Regression) for semivariogram fitting process. Types of SVM and GPR (Gaussian Process Regression) are adopted for semivariogram fitting for ordinary kriging interpolation in this experiment. The result showed that in general SVM types could predict accuracy better than other types of regression, and GPR types produce better DEM accuracy based on the experiment.

1. Introduction
DEM (Digital Elevation Model) is “an ordered array of numbers that represent the spatial distribution of elevations above some arbitrary datums in the landscape”, by Meijerink et al. [1]. DEM could be as “digital file consisting of terrain elevations for ground positions at regularly spaced horizontal intervals”, by USGS in [1]. DEM can be defined as an elevation representation of earth surface in a digital form. Other types of elevation representation are in the form of height points, contour, DTM (Digital Terrain Model), DSM (Digital Surface Model), etc. DEM could be applied for many applications such as civil engineering; earth sciences; planning and resource management; surveying and photogrammetry; military applications [1]. United Nations has proposed sustainable development goals to be achieved in 2030 which require elevation data as geospatial information will be applied to zero hunger; clean water and sanitation; affordable and clean energy; decent work and economic growth; industry, innovation and infrastructure; sustainable cities and communities; responsible consumption and production; climate action; life below water; life on land; partnership of the goals [2]. DEM can be generated from several sources. Contour is possible to be transformed into DEM type. Aerial photography which is based on photogrammetry processing activity will generate DEM
from overlapping stereo images. This technology is also adopted in optical satellite stereo image acquisition processing. Some examples of this type of satellite are ASTER, SPOT 5/6/7, IRS 1C, Cartosat 1, ALOS PRISM [3]. Another technology approach to producing DEM is based on active remote sensing sensor. SAR (Synthetic Aperture Radar) and LIDAR (Light Detection and Ranging) are for instances. SRTM (Shuttle Radar Topography Mission) is an example of a radar acquisition program to acquire a global DEM.

The quality of DEM is considered on two factors, which are in term of precision and accuracy. Precision can be defined as how detail the elevation is presented. This could be the spatial resolution of DEM. The accuracy of DEM is validated by performing error analysis using other data which could be the higher precision of measurement or higher scale of elevation data as a reference. Elmoustafa et al. stated that several factors play an important role on DEM accuracy, there are terrain roughness; sampling density (elevation data collection method); grid resolution or pixel size; interpolation algorithm; vertical resolution; terrain analysis algorithm [4]. Terrain roughness might be similar to topographic characteristic as also mentioned a factor that affects DEM accuracy [5]. Heritage et al. experimented that pointing accuracy also affects DEM accuracy, [6]. Then Setiyoko & Arymurthy added land cover type as an additional factor affecting DEM accuracy [3]. Some researchers have been focusing on sampling density in their experiment researches. Different sampling density might influence on predicting soil organic carbon estimation by using weighted regression kriging [7]. Sampling design also might affect the probability to detect soil carbon stock changes [8]. Sampling scale is also studied in order to increase accuracy in mapping soil carbon landscapes [9].

Concerning generating DEM from height points required a suitable method algorithm to perform interpolation process. Interpolation is the process of estimating the value of attributes at unsampled sites from measurements made at point locations within the same area or region [10]. Interpolation techniques are using the principles of spatial autocorrelation, which in the estimating process assumes that the value of closer points are more similar compared to farther one's values [11]. This process is related to approximation functions in mathematics and estimation based on sampling in statistics [12]. There two types of approximation functions can be applied for interpolation process: deterministic and stochastic interpolators. The deterministic method is using a mathematical formula to form weighted averages of nearby known values, and also provide no assessment of errors with predicted values. The stochastic method is using weighted averages and also probability models to make predictions, the assessment of prediction errors are also offered with estimated variances. Deterministic Interpolators: Inverse Distance Weighting, Global (Trend Surface), Local Polynomial, Radial Basis Functions. Stochastic Interpolators: Global (Regression), Simple Kriging, Ordinary Kriging, Universal Kriging, Indicator Kriging, Probabilistic Kriging, Disjunctive Kriging, Cokriging.

Kriging actually is a common method nowadays to perform the process in term of the geostatistical method. There are types of kriging interpolation: ordinary kriging, simple kriging, universal kriging, disjunctive kriging, indicator kriging, and cokriging [10]. Some modification of the kriging method was experienced in scientific research. Weighted regression kriging is conducted to predict soil organic carbon [7]. Cokriging which were using two variables to calculate unknown elevation values between known values was applied to analyze DEM accuracy [3]. Kriging is also applied for data fusion. MODIS data fusion by using kriging with an external drift [13]. Regression kriging is applied to fuse high-resolution satellite image [14]. Investigations that have been done by Arun, which proved that the kriging method performs better when compared to other contemporary methods [11]. Previous experiments were conducted using various types of stereo satellite image such as IRS-1C [15] and Cartosat-1 [16]. It’s proven that kriging was given the least residual error in the interpolation process. In order to strengthen the result analysis of previous research, this study was conducted to perform machine learning regression for semivariogram fitting in ordinary kriging process. Machine learning for semivariogram fitting process was experienced for some studies previously. Huang et al have applied machine learning method, Support Vector Machine (SVM) to contribute to kriging interpolation technique [17]. A type of SVM, called LS-SVM (Least Square-SVM) applied for fitting the semivariogram prior to kriging interpolation. SVM could avoid the subjectivity in choosing the
available function to perform semivariogram fitting. It is proven that the combination of both SVM and kriging interpolation is a good and adaptive method particularly for data which has complex spatial structure, for example, oceanic missing data. Then [18] continued to use a similar method, LS-SVM with kriging interpolation to predict coal seam thickness prediction. The method could adaptively fit according to the data structure, which significantly improves the interpolation precision. The Support Vector Regression (SVR) method as a modification of SVM and Multi-Gene Genetic Programming (MGGP) were initiated for semivariogram fitting [19]. Its proven that MGGP-based method as well as an SVR-based method for variogram modeling has the ability to fit more exactly experimental variogram without assuming the basic model shape and reflect the more objectively spatial variation of the real field comparing with traditional WLS method, and improve the kriging interpolation precision significantly. In this research we applied various types of machine learning regression such as SVM Linear (SVML), SVM Gaussian (SVMG), SVM Quadratic (SVMQ), SVM Cubic (SVMC), SVM Fine Gaussian (SVMGF), SVM Medium Gaussian (SVMGM), SVM Coarse Gaussian (SVMGC), Gaussian Process Regression, Rational Quadratic (GPRRQ), Gaussian Process Regression, Squared Exponential (GPRSE), Gaussian Process Regression, Matern 5/2 (GPRM), Gaussian Process Regression, Exponential (GPRE), Least-Square SVM (LSSVM). All mentioned types were applied for DEM accuracy prediction and post-process accuracy analysis using ordinary kriging.

2. Literature Review
2.1 Ordinary Kriging
Unknown point value estimation can be calculated by interpolation using ordinary kriging equations [20]:

$$\hat{z}(x_0) = \sum_{i=1}^{k} \lambda_i z(x_i)$$  \hspace{1cm} (1)

where, $\lambda_i$ is the weight indicating each of the sample sections for the calculation of the unknown point. If the weight of $\lambda_i$ is known, unknown point value estimation can be calculated.

The weight can be derived by solving a set of simultaneous equations, for example, to predict value at 0 is using 3 known points (1,2,3) [21]:

$$
\begin{align*}
W_1 y(h_{11}) + W_2 y(h_{12}) + W_3 y(h_{13}) + 0 = y(h_{10}) \\
W_1 y(h_{21}) + W_2 y(h_{22}) + W_3 y(h_{23}) + 0 = y(h_{20}) \\
W_1 y(h_{31}) + W_2 y(h_{32}) + W_3 y(h_{33}) + 0 = y(h_{30})
\end{align*}
$$  \hspace{1cm} (2)

where $y(h_{ij})$ is the semivariance between known point $i$ and $j$, $y(h_{10})$ is the semivariance between the ith known point and the point to be estimated, $0$ is a Langrange multiplier, which is added to ensure the minimum possible estimation error. The mentioned equations can be solved in the form of matrix:

$$
a = \begin{bmatrix}
\gamma(h_{11}) & \ldots & \gamma(h_{13}) \\
\vdots & \ddots & \vdots \\
1 & \ldots & 1
\end{bmatrix}
\begin{bmatrix}
W_1 \\
\vdots \\
W_n \\
\emptyset
\end{bmatrix} = \begin{bmatrix}
y(h_{10}) \\
\vdots \\
\emptyset
\end{bmatrix}
$$  \hspace{1cm} (3)

once the weights are solved then,

$$z_0 = z_1 W_1 + z_2 W_2 + z_3 W_3$$  \hspace{1cm} (4)

2.2 Support Vector Machine (SVM) Regression
SVM regression, instead of for classification this method is taking the least-squares error function that will keep the weights norm small [22].

$$\frac{1}{2} \sum_{i=1}^{N} (t_i - y_i)^2 + \frac{1}{2} \lambda \|w\|^2$$  \hspace{1cm} (5)

Then it is transformed using e-insensitive error function ($E_e$) that will be 0 if the delta between the target and the output is less than $e$. The form of this error is,

$$\sum_{i=1}^{N} E_e(t_i - y_i)^2 + \frac{1}{2} \lambda \|w\|^2$$  \hspace{1cm} (6)
2.3 Gaussian process regression (GPR)
Gaussian process regression (GPR) goal is to make the point prediction $y_{\text{guess}}$ which produces the least loss, but $y_{\text{true}}$ is yet known. Instead, minimizing the expected loss or risk, by averaging the models opinion [23], the truth might be

$$R_{L}(y_{\text{guess}}|X) = \int L(y_{*,y_{\text{guess}}})p(X,D)dy_{*}$$

Thus the best guess, in the sense that it minimizes the expected loss, is

$$y_{\text{optimal}}|X = \arg\min_{y_{\text{guess}}} R_{L}(y_{\text{guess}}|X)$$

In general the values of $y_{\text{guess}}$ that minimizes the risk of the loss function $|y_{\text{guess}} - y_{*}|$ is the median of $p(y_{*}|X,D)$, while for the squared loss $(y_{\text{guess}} - y_{*})^2$ it is the mean of this distribution.

3. Methodology
Data used in this experimental research is height point extracted from Cartosat-1 stereo imagery and validated using the high precision of height point measured using differential GPS, as seen in Figure 1.

The research flowchart is seen in Figure 2., which is started by calculating semivariance for each pair of height points in specific distance for all height point data. Then the semivariogram is fitted by conducting various types of SVM functions, various types of GPR functions, stable, exponential, and spherical. The results are accuracy prediction of each fitting model and post-processing accuracy using a validation point. Accuracy is measured using MAE (Mean Absolute Error).

---

**Figure 1.** Height point of Cartosat-1.

**Figure 2.** Research Flowchart.
4. Result and Analysis

The output of the fitting semivariogram is seen in Figure 3. Input data is Cartosat-1 semivariogram with 2207 semivariance values, and 5 folds of machine learning regression process. Output expected are fitting functions & predicted semivariogram values based on distance.

![Figure 3. Example of Semivariogram Fitting Functions: a. SVM Quadratic (SVMQ); b. Gaussian Process Regression, Squared Exponential (GPRSE).](image)

The research result showed that GPRE fitting model has the least error for ordinary kriging interpolation, which MAE is 0.46 m. LS-SVM is the second best fitting model where the MAE is 0.59 m. Overall SVM and GPR accuracies are better than traditional fitting functions, except for SVMC, SVMG, and SVMGF. The whole accuracy result is in Table 1.

| No | Function  | Accuracy Prediction (m) | Post-process Accuracy (m) | delta Accuracy (m) |
|----|-----------|-------------------------|---------------------------|-------------------|
| a  | Stable    | 0.88                    | 1.37                      | 0.48              |
| b  | Exponential | 0.88                   | 1.34                      | 0.46              |
| c  | Spherical | 0.89                    | 1.16                      | 0.27              |
| d  | SVML      | 1.02                    | 0.84                      | -0.18             |
| e  | SVMQ      | 0.91                    | 1.15                      | 0.24              |
| f  | SVMC      | 2.54                    | 7.65                      | 5.11              |
| g  | SVMG      | 2.67                    | 6.16                      | 3.49              |
| h  | SVMGM     | 0.94                    | 0.98                      | 0.04              |
| i  | SVMGF     | 1.25                    | 1.37                      | 0.12              |
| j  | SVMGC     | 0.89                    | 0.93                      | 0.05              |
| k  | GPRE      | 0.93                    | 0.46                      | -0.46             |
| l  | GPRM      | 1.00                    | 0.68                      | -0.32             |
| m  | GPRRQ     | 1.02                    | 0.60                      | -0.42             |
| n  | GPRSE     | 0.96                    | 0.79                      | -0.17             |
| o  | LS-SVM    | 0.91                    | 0.59                      | -0.32             |

As seen in Table1, SVMGM is the best in predicting accuracy, the difference between predicting and post-processing accuracies in term of ordinary kriging interpolation is the least compared to other fitting methods.

5. Conclusion

In general, the implementation of machine learning regression approach in semivariogram fitting process had shown contribution in both DEM accuracy prediction and it's post-process accuracy as it compared to the common traditional method used in the process such as stable, exponential, and
spherical. Overall GPRS is better than SVM in generating DEM based in term of accuracy, but SVM is more accurate in predicting the error. The further experiment might be using larger data which consists of more height points, in order to gain a better research contribution. Another thing to consider is a computational cost to be analyzed based on each type of machine learning regression.

References
[1] H Kiyoshi 1994 Digital Elevation Model (DEM) Processing Lecture AIT
[2] U Nations 2015 Transforming our World: Geospatial Information Key To Achieving The 2030 Agenda For Sustainable Development DigitalGlobe
[3] A Setiyoko and A M Arymurthy 2017 Accuracy Analysis of DEM Generated from Cokriging Interpolators Aeroearth 2017 no. 1
[4] A M Elmoustafa, H N Farwy and M M Elfuwy 2015 Effect of Elevation Data Accuracy on Storm Drainage Schemes, Lagos, Nigeria Nat. Resour vol. 6 pp. 433–441
[5] H Yang, C Wang, T Ma and W Guo 2015 Accuracy assessment of interpolation methods in grid DEMs based on a variance-scale relation Environ. Earth Sci pp. 6525–6539
[6] G L Heritage, D J Milan, A R G Large and I C Fuller 2009 Geomorphology Influence of survey strategy and interpolation model on DEM quality Geomorphology vol.112 no.3–4 pp.334–344
[7] H Ye, W Huang, S Huang, Y Huang, S Zhang, Y Dong and P Chen 2017 Effects of different sampling densities on geographically weighted regression kriging for predicting soil organic carbon Spat. Stat. vol. 20 pp. 76–91
[8] A Heim, L Wehrli, W Eugster and M W I Schmidt 2009 Effects of sampling design on the probability to detect soil carbon stock changes at the Swiss CarboEurope site Lageren Geoderma vol. 149 no. 3–4 pp. 347–354
[9] B A Miller, S Koszinski, W Hierold, H Rogasik, B Schröder, K Van Oost, M Wehrhan and M Sommer 2016 Towards mapping soil carbon landscapes: Issues of sampling scale and transferability Soil Tillage Res. vol. 156 pp. 194–208
[10] P A Burrough and R A McDonnel 1998 Principles of Geographical Information Systems no. January 1998 Oxford University Press
[11] P V Arun 2013 A comparative analysis of different DEM interpolation methods Egypt. J. Remote Sens. Sp. Sci. vol. 16 no. 2 pp. 133–139
[12] Z Li and Q Zhu 2005 Digital Terrain Modeling Principles and Methodology CRC Press
[13] M H R Sales, C M Souza, and P C Kyriakidis 2013 Fusion of MODIS Images Using Kriging With External Drift IEEE Trans. Geosci. Remote Sens. vol. 51 no. 4 pp. 2250–2259
[14] Q Meng, B Borders and M Madden 2010 High-resolution satellite image fusion using regression kriging Int. J. Remote Sens. vol. 31 no. April pp. 1857–1876
[15] A Setiyoko 2013 Analysis of spatial interpolation techniques for DEM generation using IRS-1C data 34th Asian Conference on Remote Sensing ACRS 2013 vol. 2 pp. 789–794
[16] A Setiyoko and A Kumar 2012 Comparison Analysis Of Interpolation Techniques For Dem Generation Using Cartosat-1 Stereo Data Int. J. Remote Sens. Earth Sci. vol.9 no.2 pp.78–87
[17] Z Huang, H Wang and R Zhang 2012 An Improved Kriging Interpolation Technique Based on SVM and Its Recovery Experiment in Oceanic Missing Data Am. J. Comput. Math. vol. 2 no. 4 pp. 56–60
[18] C Youkou, Y Yongguo and W Wangwen 2015 Coal Seam Thickness Prediction based on Least Squares Support Vector Machines and Kriging Method EIGE vol. 20 pp. 167–176
[19] C Han, J Wang, M Zheng, E Wang, J Xia, G Li and S Choe 2016 New variogram modeling method using MGGP and SVR Earth Sci. Informatics pp. 197–213
[20] P A Burrough and R A McDonnell 2009 Principles of Geographical Information Systems Oxford University Press
[21] K T Chang 2006 Introduction to geographic information systems no. v. 1. McGraw-Hill
[22] S Marsland 2015 MACHINE LEARNING An Algorithmic Perspective CRC Press
[23] C E Rasmussen and C K I Williams 2009 Gaussian Processes for Machine Learning MIT Press