Reliability of using LS-VCE computation in Deriving Variances for Multi-Classes Dataset

Akram Zulkifli$^{1,1}$, Mohd Azwan Abbas$^{1,2}$, Norshahrizan Mohd Hashim$^{1,3}$, Mohamad Asrul Mustafar$^{1,4}$, Ashraf Abdullah$^{1,5}$

$^1$Faculty of Architecture, Planning & Surveying, Universiti Teknologi Mara, Perlis Branch, Arau Campus, 02600 Arau, Perlis, Malaysia
akramzulkifli96@gmail.com$^1$, mohdazwan@uitm.edu.my$^2$, norshahrizan@uitm.edu.my$^3$, mohamadasrul@uitm.edu.my$^4$, ashraf@uitm.edu.my$^5$

Abstract. Stochastic modelling (SM) plays an essential role in least-squares adjustment (LSA), especially for geodetic network data processing. Estimated variances derived from SM are vital factors in determining the reliability of the computed parameter vectors and ensuring the sensitivity of adjustment outcomes toward outliers. As there are multi-source of datasets consisting various of data quality, there is still room for improvement when positional accuracy becomes the main priority. Concerning the accuracy argument, legacy datasets that were exploited in establishing the National Digital Cadastral Database (NDCDB) were obtained from multi-classes of measurement (i.e., first, second and third classes). Taking into account this condition, this research has investigated the capability of stochastic modelling to preserve the positional accuracy of land records that comprehends from multi-classes data quality. To achieve that, the algorithm of Least Squares Variance Component Estimator (LSVCE) was employed in estimating realistic variances. First and second classes measurement were yielded from three (3) certified plans (CPs) which are CP93887, CP80333, and CP33758. Comparison between the adjusted results computed from the combined and separated variance according to data classes have demonstrated that combined variance can detect outliers while separated variance can give realistic adjustment results. From these outcomes, the experiments verified that a hybrid solution is needed for both data classes in order to preserve positional accuracy. In conclusion, to ensure the accuracy of survey data in the future, a proper variance component is needed to improve the coordinated cadastral database.

1. Introduction
A cadastral map is a parcel-based dataset that is used to identify the boundaries of a specific area. To keep up with the rapid advancement of spatial-based technologies, such as Geographic Information Systems (GIS), it is necessary to provide a coordinated digital representation of cadastral parcels in order to update and reform old cadastral databases [12]. A number of concerns with projection from conventional into a modern (digital) cadastral database, including A number of concerns with the Cadastre Database, including technologies, accuracy limitations, and challenges deriving from the use of multiple projections and georeference systems, are addressed by the digital cadastral database. Drastic changes have been made by the international’s cadastral authorities to improve the ability of adjustment procedures, from simplified (Bowditch) to comprehensive (LSA) approaches [2].
Since multi-classes of data come from various sources of the dataset, proper adjustment technique is inevitable. Previous implementation in Malaysia, the reduction of errors observation have been carried out by using the Bowditch method. This method is a simple algorithm which previously employ to reduce the total errors of latitude and departure [1]. Nowadays, the sophisticated instrument, as well as advanced tools of processing, had led to the implementation of the least square adjustment (LSA) method which eventually becomes an algorithm of calculation to reduce the data observation. As depicted in figure 1, parametric least square adjustment has been utilised in NDCDB implementation. Design matrix (A) and observation vectors (L) were extracted from the observation equation, while weight matrix (W) was derived from estimated variances. Based on simplified linear regression [2], those matrices were exploited in computing adjusted parameters with variance covariance matrices. The Chi-square test will be used to statistically measure the acceptance of the adjustment procedure by examining the similarity of computed variance (a posteriori) with respect to population variance (a priori).

![Diagram showing least square adjustment process](image)

**Figure 1.** The effect of variance selection in LSA computation.

The LSA method is an advanced adjustment technique that adapts observations based on the laws of probability, which predicts the existence of random errors [3]. Involved with complex computation procedures (i.e., matrices computation), LSA has not become an option in the previous implementation due to the limitation of technology and computer processors. These equations have already been made automatic by the availability of modern computers and the least square technique has thus gained popularity [4]. On top of that, for any form of cadastral data, the LSA is relevant and has the advantage that observations of different accuracies can be correctly weighted in the computations [5].

When processing geodetic findings, the least square theory introduced two types of models, functional and stochastic models [7]. The functional model represents the relation of observation and unknown parameters while the stochastic model describes the quality of the measurements and their correlation with each other. Through this modelling, the estimation of variance components could be
appropriately characterized according to the real-world measurement. Among the implementation, there is an approach that derives the variance based on the historical data for a certain period using normal time series techniques. According to Zangeneh-Nejad [12], the realistic option of the stochastic model observables can yield the best linear unbiased estimation (BLUE) using the Least Square Variance Component Estimator (LS-VCE). The choice of unbiased is critical since most functional relationships are non-linear, even though the statistical features of the estimates are usually derived from a linear substitution issue [13].

According to Grodecki, the observed weight can have a direct influence on the least square adjustment outcomes. The weight that should be assigned to the observations affects the least square adjustment computation solution [14]. To emphasise this issue, Amiri-Simkooei [11] has introduced the LS-VCE which can be considered as a robust approach to estimate unknown variance and covariance components. The reliability of LS-VCE has been verified by the bundle of research that has been carried out to modify the existing LS-VCE [9]-[10]. To establish more accurate weighting parameters for altering the geodetic network, a stochastic modelling approach is provided. The weight matrix is created in the beginning using the variance of observations, with input data given as the standard deviation for each multi-class dataset [5]. In Malaysia, cadastral databases consist of various data sources. The most prominent discrepancies contributed by multi-classes of data quality that have been used to develop the cadastral database. Taking into account the wide range of data quality exploited informing the database, the reliability of LS-VCE in deriving significant variance is questionable. Blindly accepting the variance estimated from LS-VCE without considering this fundamental uncertainty, it might jeopardise the final product land record. Thus, further investigation is necessary to examine the effect of multi-quality of cadastral data with regard to stochastic modelling. With determination to resolve the ambiguity this study has to evaluate the capability of LS-VCE either, it can preserve positional accuracy in cadastral network adjustment.

2. Experiment
To experiment with the above mentioned issue, there are three (3) certified plans (CP) consisting of 1st and 2nd classes data has been utilised. Those adjacent CP’s are 93887, 80333, and 33758 located at Kangar, Perlis. Since cadastral databases come from a variety of sources, it is a necessity to consider that the datasets are consisted of multiple types of errors to represent data quality. If this defect occurs, LSA must be able to identify the outliers or else the adjustment results may be harmed, and the positional correctness of the legacy data would be compromised. For summarization of this research, a sample of data that consists of three (3) datasets that is one 1st first-class CP and two 2nd class CP has been selected to accomplish the unbiased of the experiment (refer figure 2).
Figure 2. Three certified plans employed for this study (Lot 1, 2, and 3 represent 1st class CP (CP93887) and lot 4 (CP80333) and 5 (CP33758) represent 2nd class CP).

In order to extract adequate realistic variance for stochastic modelling, Amiri-least Simkooei's square variance component estimation (LSVCE) was used [11]. The formulation of LS-VCE to derive the realistic variance based on the findings of least square adjustment is shown in figure 3.

Figure 3. Three certified plans employed for this study (Lot 1, 2, and 3 represent 1st class CP and lot 4 and 5 represent 2nd class CP) [4].
With the aid of LS-VCE, three (3) sets of variances were estimated for those multi-classes CPs. Employing the standard deviations as stated in the circular of the Department of Surveying and Mapping Malaysia (DSMM), initial sigma in LS-VCE were set as fifteen (15) seconds and 0.010m for bearing and distance, respectively. To measure the reliability of LS-VCE in handling a wide range of data quality (i.e., 1st and 2nd classes data), two configurations have been designed. The first configuration has performed LS-VCE using all three (3) CPs that consist of both classes data in estimating the variance (tabulated in Table 1 for combined CPs). While the second configuration has partially derived the variances according to CPs class. Data from the 1st and 2nd classes CPs have their respective variances as presented in Table 1.

| Obtained From | Standard Deviation for Bearing | Standard Deviation for Distance |
|---------------|--------------------------------|---------------------------------|
| Combined CPs  | 22"                           | 0.014m                          |
| 1st Class CP  | 10"                           | 0.005m                          |
| 2nd Class CPs | 99"                           | 0.043m                          |

Utilising the single estimated sigmas, the first configuration has executed the LSA computation for all three (3) CPs. To examine the reliability of the outcomes, the Chi-square test is used to initially indicate the first assumption and final conclusion has been obtained from the distribution of residuals and plotted error ellipses (derived from standard deviations).

To reduce the range of data quality, second configuration has independently computed the variances for the 1st and 2nd classes CPs. An exclusive set of variances for each data class have been adjusted in a single LSA computation procedure. Similar to the previous configuration that utilises variances from combined CPs, this later configuration also has been evaluated based on the obtained global test, residuals and graphical standard deviations.

Results from both configurations have been used to systematically examine the reliability of LS-VCE. Based on the designed experiments, it is expected that residuals of both classes should be fairly evaluated according to the measurement classes. As for parameters, the numerical or graphical precision might indicate the quality of the data accordingly.

3. Result and Analysis

Star*Net software was utilised to perform least square adjustment (LSA) for the dataset with synthetic uncertainties. The global test demonstrated overwhelming results when both implementations of variance combine CPs and different classes (1st class and 2nd class CP) variances manage to get level passed at 95 % confidence level. Further assessments were carried out to mathematically verify the significance of realistic variance for the experiment data with respect to cadastral network adjustment. The first analysis focuses on residual for distance and bearing. Then, it was followed by examining the standard deviation and error ellipse of adjustment points of the multi-classes CPs. Lastly, coordinates changes for CPs must be scrutinized in order to determine whether it can preserve positional accuracy.

As been seen, Star*Net software has assessed the value of residuals to determine the presence of outliers in the dataset using Star*Net outliers detection. Figure 4 until 7 demonstrate the residual distance and bearing according to CP classes by using combined variance and separated variance by classes. 1 sigma will represent combined variance while 2 sigmas represent separated variance by classes. It is clear in figure 4 and 7 that there is a significant difference for residual for CP 1st class, especially at line 12 where the residual distance contribute 0.006m for combined variance while 0.039m for the separated variance. For the residual bearing in the same line, it contributes -34 sec for combined variance while -11 sec for the separated variance. Residuals for CP 2nd class also give a
slight difference where residual distance contribute -0.042m for combined variance while -0.075m for the separated variance at line 1. For the residual bearing, there is two line observation that shows the substantial difference which is line 1 and 11. For line 1 it contributes 36 sec for combine variance while 59 sec for the separated variance while for line 11 it contribute 10 sec for combine variance while 23 sec for the separated variance. The cause of the significant difference is due to data input using the original value from each of the classes CP. Ideally, for bearing values and adjacent distances between different CPs, it must follow the latest value for any recent CP. In this case, CP80333 must follow the value of bearing and distance from the recent CP that is CP93887. For this case, since input data for CP 2nd class does not follow the latest input of CP 1st class, it can be considered as outliers for the datasets. Outliers inside adjustment can lead to incorrect results during adjustment computation.

Figure 4. Residual distance for CP 1st class by using combine variance and separated variance by classes (left)

Figure 5. Residual distance for CP 2nd class by using combine variance and separated variance by classes (right)

Figure 6. Residual bearing for CP 1st class by using combine variance and separated variance by classes (left)

Figure 7. Residual bearing for CP 2nd class by using combine variance and separated variance by classes (right)

Figure 8 and 9 denote the error ellipse and also represent the standard deviation of station coordinate after going through for LSA. After applying the LS-VCE method, the variance showed the standard deviation of station 9541395328 for combined variance and separated variance changed from 0.043m to 0.085m. Plotted error ellipse show there is a major difference where for combined variance it showed an unrealistic error on CP 2nd class while separated variance for CP 2nd class it is more realistic. It happens when the estimated value by LS-VCE that has been used for combine variance represents the whole merging data classes accuracy for 2nd class data compared to the separated variance that directly represents according to classes data. Realistic variance is important in order to fulfil the quality of the measurements and their correlation with each other.
Figure 8. Plotted errors ellipses for the corrected values using LS-VCE combine CPs (left)

Figure 9. Plotted errors ellipses for the corrected values using LS-VCE 1st and 2nd class CPs (right)

Figure 10 and 11 represent the different coordinates for combined and separated variances for CP 1st class and CP 2nd class. Since Figure 4 until 7 has shown the significant difference for residual observation in line 12 for CP 1st class and line 1 and 11 for CP 2nd class due to outliers, it lead to difference adjusted coordinates changes using combine and separated variances. The adjusted coordinates in point 21 for CP 1st class and point 3 for CP 2nd class contribute 0.025m and 0.021m. The major difference of these two (2) points has resulted from line 12 for CP 1st class and lines 1 and 11 for CP 2nd class that contain outliers. Since separated variance gives a more realistic result for standard deviation, adjusted coordinates for separated variance are more likely will be accepted since the estimated variance using LS-VCE represent by classes.

Figure 10. Different coordinates for CP 1st class

Figure 11. Different coordinate for CP 2nd class

4. Conclusion
This study presents the estimated variance component of the cadastral network using the least squares variance component estimation (LS-VCE) method. Traditionally, the joint adjustment problem has been studied exclusively using least-squares theory, which considers only the random errors of each dataset's observation vector while disregarding any random error in the coefficient matrix. The results show that LS-VCE is capable of determining the realistic stochastic modelling value to employ in estimating the variance for multi-classes dataset. To produce the best linear unbiased estimation (BLUE), a realistic stochastic model of the observables must be used. This research also can examine that combined variance can detect outliers but cannot give realistic adjustment results while separated
variance able to produce realistic variance parallel with adjustment results. This research suggests and hybrid solution in order to run the LS-VCE method by using multi-classes dataset by detecting and cleaning outliers using combined variance and producing realistic adjustment results using separated variance by classes. Given the multiple quality data in different CP, it is expected that propagation of errors will reveal significant positioning uncertainty. In line with the fundamental concept, LS-VCE variances can produce realistic positional value and achieve the aim of this research.

Acknowledgement
We would like to express sincere appreciation to Universiti Teknologi Mara (UiTM) and the Ministry of Higher Education (MoHE) of Malaysia for the UiTM research grant FRGS (600-RMC/FRGS 5/3 (181/2021)) that has been given to this research project. We also would thank our colleagues from Universiti Teknologi MARA (UiTM) who provided insight and expertise that greatly assisted in this research.

References
[1] M. Y. Mohd Yusoff, H. Jamil, N. Z. Abdul Halim, N. A. Mohamed Yusof, and M. A. Mohd Zain, “Ekadaster: A Learning Experience for Malaysia” in FIG Pacific Small Island Developing States Symposium Policies and Practices for Responsible Governance Suva, Fiji, 18 – 20 September 2013.
[2] M. Jeffri, O. Hisham, J. Joanes, Effectiveness Of Localised Adjustment In Strengthening National Digital Cadastre Database in Jurnal Teknologi, 2017.
[3] E. Lambrou, N. Kanellopoulos, Check and Calibration of A Single GNSS Receiver by Using The VRS RTN Positioning Method in Measurement: Journal of the International Measurement Confederation, Vol. 117, Issue 21: 221-225, 2018.
[4] N. K. Bidi, A. H. M. Din, Z. A. M. Som, and A. H. Omar, Adjustment of Cadastral Network Using Least-Squares Variance Component Estimation in ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W16(October), 161–167, 2019.
[5] H. M. Taylor, S. Karlin, An Introduction to Stochastic Modeling, Third Edition, California: Academic Press, 1998.
[6] B. Li, Y. Shen, L. Lou, Efficient Estimation of Variance and Covariance Components: A Case Study for GPS Stochastic Model Evaluation in IEEE Transactions on Geoscience and Remote Sensing, Vol. 49, Issue 1: 203-210, 2011.
[7] Y. Q. Chen, A. Chrzanskowski, J. M. Secord, A Strategy for The Analysis of The Stability of Reference Points in Deformation Surveys in CISM Journal ACSGC, Vol. 44, Issue: 141-149, 1990.
[8] K. Moghtased-Azar, R. Tehranchi, A. R. Amiri-Simkooei, An Alternative Method for Non-Negative Estimation of Variance Components in Journal of Geodesy, Vol. 88, Issue 5: 427-439, 2014.
[9] P. J. G. Teunissen, A. R. Amiri-Simkooei, Least-Square Variance Component Estimation in Journal of Geodesy, Vol. 82, Issue 2: 65-82, 2008.
[10] S. Farzaneh, A. Safari, Kamal Parvazi, Improving Dam Deformation Analysis Using Least-squares Variance Component Estimation And Tikhonov Regularization in Journal of Surveying Engineering, Vol 147, Issue 1: 1-9, February 2021.
[11] F. Zangeneh-Nejad, A. R. Amiri-Simkooei, M. A. Sharifi, and J. Asgari, Recursive Least Squares With Real Time Stochastic Modeling: Application to GPS relative positioning in International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(4W4), 531–536, 2017.
[12] N. M. Hashim, A. H. Omar, K.M. Omar, N. M. Abdullah, M. H. M. Yatim, Cadastral Positioning Accuracy Improvement: A Case Study in Malaysia, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(4W1), 265-270, 2016.