# Implementation of Stochastic Modelling in Enhanced Cadastral Database for Multi-Classes Dataset

**Zulkifli, A.,[1*] Abbas, M. A.,[1] Hashim, N. M.,[1] Mustafar, M. A.,[1] Sulaiman, S. A.,[2] Razak, N. N. A.,[1] Yusop, M. K. M.[3] and Nordin, S.[4]**

[1]Surveying Science and Geomatics Studies, College of Built Environment, Universiti Teknologi Mara, Perlis Branch, Arau Campus, 02600 Arau, Perlis, Malaysia
E-mail: akramzulkifli96@gmail.com,* mohdazwan@uitm.edu.my, norshahrizan@uitm.edu.my, mohamadasrul@uitm.edu.my, nurnazuraabdrazak97@gmail.com

[2]College of Built Environment, Universiti Teknologi Mara Shah Alam, 40450 Shah Alam, Selangor, Malaysia
E-mail: saifulaman@uitm.edu.my

[3]Deparment Survey and Mapping Malaysia Perlis, Floor 6 Federal Building, Persiaran Jubli Emas, 01000 Kangar, Perlis Malaysia, E-mail: khairani@jupem.gov.my

[4]Geodetic Solutions Sdn. Bhd., No. 22 Floor 1 S2 D 39 Road, City Center Seremban 2, 70300 Seremban, Negeri Sembilan, Malaysia, E-mail: snsc.n9@ljt.org.my

**Corresponding Author*

**DOI: https://doi.org/10.52939/ijg.v19i5.2657**

**Abstract**

*Stochastic Modelling (SM) was a crucial component of least squares adjustment (LSA), particularly when processing data from geodetic networks. The projected variances which generate using SM execute an important part in defining both the accurateness of the computed parameter vectors and the impact of the adjustment outcomes. As positional precision becomes the primary objective, there is still potential for improvement because there are multiple sources of datasets with varying levels of data quality. Concerning the assertion that the National Digital Cadastral Database (NDCDB) is accurate, its development involved the use of historical datasets that were obtained from a number of different measurement classes, specifically the first, second, and third classes. In this study, researchers evaluated whether or not it is possible to employ stochastic modelling to maintain the position correctness of historical data that encompasses a wide range of data quality classes. In order to accomplish this, an approach known as a Least Squares Variance Component Estimator (LS-VCE) was utilised to generate reliable estimates of variances. Two (2) certified plans (CPs) that is CP93387 and CP33758 was selected as measurements for the first and second classes CP, respectively. The experiment showed that the variance that has been estimated by LS-VCE could produce realistic adjustment results, as shown by an analysis of the corrected results obtained by allocating the variance into different data classes. In light of these findings, the investigations showed and demonstrated conclusively that separate variance is necessary for each data classes with the aim of preserving positional accuracy. In conclusion, it is crucial to incorporate a realistic variance component inside a coordinated cadastral database in order to fulfil the objective of ensuring the accurateness of survey data for future time periods.*

**Keywords:** Cadastral Database**,** Least Square Adjustments, Least Square Variance Component Estimation, Multi-Classes Data, Stochastic Modelling

## 1. Introduction

For conservation and environmental management, historical cadastral maps are a vital resource for re-creating past land-use patterns. Old map records are an important source for reconstructing past land-use patterns because they provide spatial demarcations of land use [1]. Nowadays, the use of cadastral maps in digital raster format (scanned maps) is still used as a reference for measurement work. A lot of time will take to digitize these historical cadastral maps since many of the old cadastral plans were involved. Scientists and engineers are challenged to discover realistic techniques to transform this type of raster image into vector maps automatically [2]. In a database that has been built through time, it is important to remember that previous historic data will typically have a lesser level of accuracy rather than advanced data.

It also may have errors that have been identified or not during the creation of the maps [3]. Nowadays, some countries have come tackled the accuracy subjected to cadastral maps, such as the Department Survey of Jamaica, which produce standard digital cadastral maps, and Department Surveying and Mapping Malaysia (DSMM) which generates a coordinated digital-based representation of cadastral plans [4].

In Malaysia, the modernization of DSMM began in 1980 with computerized cadastral, followed in 1995 by the installation of the CALS system, which uses computer-aided land surveying. In addition to allowing cadastral surveys to be processed electronically, the CALS system also introduces the idea of a Digital Cadastral Database (DCDB) [5]. There is a growing demand for geographic data management, particularly for digital cadastral databases (DCDB). There's an issue with databases built to support a DCDB, which data can't be saved until it's passed rigorous validity checks first. This level of manual decluttering and rectification is required [3]. Then, it evolved when a coordinated database measurement, also known as the National Digital Cadastral Database (NDCDB) was used to store all the cadastral mapping data. This database was developed in 2010 and it is an improvement of the DCDB version [6]. An important goal of the NDCDB is to improve upon the Digital Cadastre Database (DCDB) flaws, such as its incompatibility with contemporary technology, lack of accuracy, and obstacles in terms of final coordinates where various types of projection along with georeference systems have been used without neglecting the need for data adjustment at this time [7].

In addition, the change from DCDB to NDCDB has also changed CP from old to new. The difference that can be seen between these two CPs is in terms of the type of coordinates used, the old CP uses the old cassini-soldner coordinates while the new CP uses the cassini-soldner geocentric coordinates. Old cassini coordinates are derive from datum MRT68 while new cassini coordinates are derive from datum GDM2000. Figures 1 and 2 shows the transformation map from DCDB to NDCDB [8].

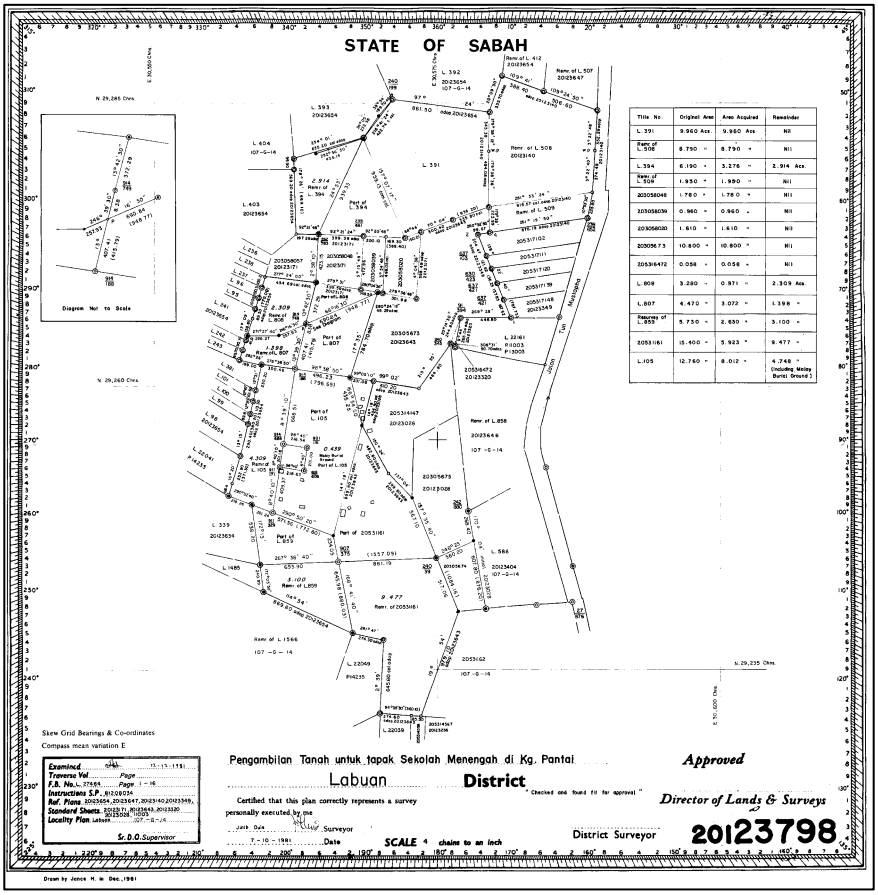



**Figure 1:** DCDB map before year 2010

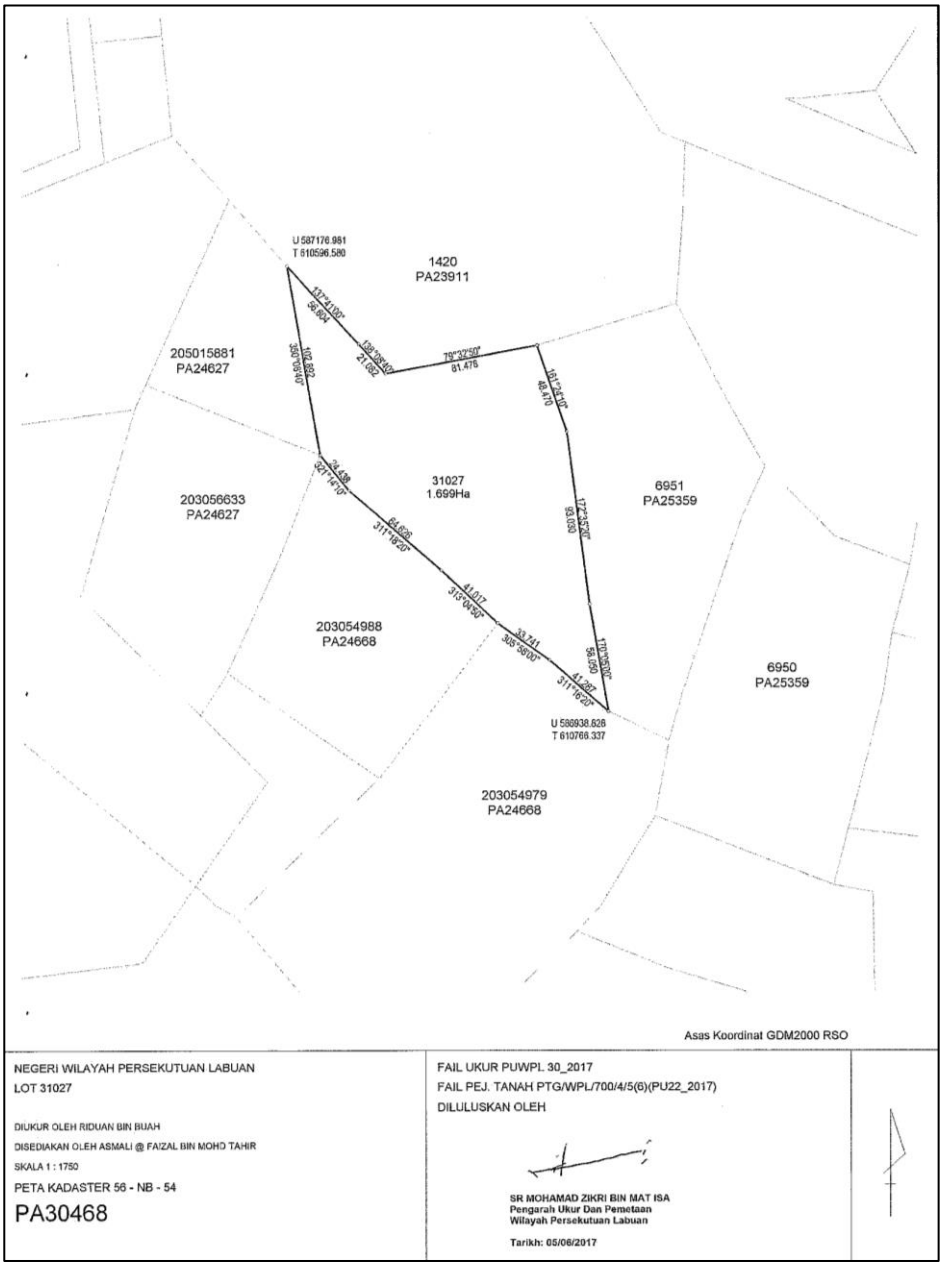



**Figure 2:** NDCDB map after year 2010

Malaysia has evolved from its initial old conventional survey systems to a contemporary design that taking into account the latest measurement equipment requirements approach that using Global National Satellite System (GNSS). In 2010 DSMM has launched a new system which has been named Ekadaster to meet the requirements of the latest measurement tools (GNSS) along with the GNSS measurement techniques that have been proposed [8]. With the establishment of a "coordinated cadastre" in 2010, the old system that based on DCDB databased was deemed outdated. The previous framework was unable to seize on the emergence of satellite-based technology (GNSS) and so delayed the practice of an absolute cadastral survey of real-time positioning. As a result, the incorporation of the organized cadastral survey idea into the recently created eKadaster technique necessitates a full overhaul of the system. The least square adjustment (LSA) approach is employed as the final evidence of boundary mark location in the distribution of propagation of measurement uncertainties, replacing the previous Bowditch method [9].

Survey data must be properly adjusted to increase survey accuracy. In mathematical geodesy, the stochastic model related on weight adjustments and calculations of survey field data has played a critical role in studying error magnitude and determining tolerance levels [10]. The weighting for matrix computation in the data adjustment has been formulated at the beginning by using the measurement data input as variance, where each observation of the measurement data will be utilised with the standard deviation value [11]. Grodecki (1997) states that the implementation of variance may have a direct effect on the LSA's outcomes. LSA solution is dependent on the variance of observations in data. Variance incorrectly can have a negative impact on LSA outcomes. A global LSA test (also known as a Chi-Square test) is computed using the weight matrix to see if the answer to LSA has passed or failed. If the global test fails, there can be no dispute about the LSA solution, and the implementation of the weight matrix utilised in the computation introduces a new factor that might have an impact on the outcome [11].

In conclusion, a detailed analysis of the stochastic model for calculating suitable weightage for each observation is necessary to enhance this PAI process on the basis of the evidence given in the problem description. Therefore, it is important to evaluate the variance component systematically. This shows that there is a need to estimate variance components in data adjustment by looking at the appropriate methods that can be used. Nowadays, knowledge and information about the covariance matrix observation is limited coupled with the fact that there are only a few studies done, especially in the application of geodetic and also cadastre. By knowing that knowledge and enlightenment about the covariance matrix observations is important for geodetic applications, since the variance component was the most common mathematical component that can evaluate a reasonable accuracy. Moreover, researchers can conduct more research into various error factors that contribute to observations [14].

## 2. Stochastic Modelling

The stochastic model provides a description of the precision underlying the measurements and the ways in which those are linked together to one another. The estimate of variance components might be improved using this modelling that is suitably described in light of the measurement used in the real world. Realistic stochastic modelling usually should be able to estimate the quality of data observables. Nowadays, there are several types of variance estimation techniques that are used to acquire realistic stochastic models for the accuracy of cadastral and geodetic data. These techniques include the Minimum Norm Quadratic Unbiased Estimator (MINQUE), the Maximum Likelihood Estimation (MLE), the Helmert Method, and the Least Square Variance Component Estimation (LS-VCE). Each of these methods generates the variance by employing standard time series techniques based on the data collected in previous periods [15]. According to Zangeneh-Nejad, the LS-VCE has the potential to yield the most effective way of best linear unbiased estimation (BLUE) since it takes into account the practical possibilities underlying on the stochastic model observables. Estimates' statistical features are often generated from a linear substitution issue, but selecting an unbiased one is critical since most functional interactions aren't linear [16].

Least Square Adjustment (LSA) technique or also known as L2-Norm is one of the appropriate adjustment procedures for the adjustment of cadastral measurement data if outliers in the raw observation data such as gross and systematic error can be eliminated. LSA calculates the most probable values via utilising into consideration the fundamentals of mathematical probability and then attempting to minimise the sum of squares that is the consequence of random error occurring in the weights that have been yield. LSA was developed for the purpose of numerically calculating the significance of variances (i.e. stochastic modelling). This was done in order to avoid an unbiased solution during the process of assigning weight on every component among the observations. The selection of an appropriate stochastic (weighting) model is essential to the adjustment since the weight of an observation determines how much correction it receives. As shown in Figure 3, the order of formulation used by LSA may put the adjusted results at risk due to the failure to choose the right variance, which would ultimately damage LSA's capacity to separate errors in observation sets. In Baarda's approach for identifying outliers, the criterion for deciding whether or not to reject errors is the calculated result of the standardised residuals, which is set at 3.29 [13].

## 3. Experiment

Based on Figure 3, it has shown that wrongly determining the value of weighting that has been derived from variance can be jeopardised by inaccuracy results from least square adjustment (LSA). Since that, this experiment will focus on two (2) intentions which are to analyse whether the homogenous variance is superior compared to Least Square Variance Component Estimation (LS-VCE) variance.

- Quality of observation represent by standard deviation ($\boldsymbol{\sigma_{obs}}$)
  Standard deviation used to yield weight matrix ($\boldsymbol{W = 1/\sigma_{obs}}^2$)
- Weight Matrix then employed to derive correction of unknown parameters
  [$\boldsymbol{X} = (A^T \boldsymbol{W} A)^{-1}(A^T \boldsymbol{W} L)$]
- **Adjusted Parameter** ($\boldsymbol{X_{adj}}$) = Approximate parameters + $\boldsymbol{X}$
- Adjusted observations ($\boldsymbol{L_{adj}}$) were derive from $\boldsymbol{X_{adj}}$
- Observations residual (**V**) = $\boldsymbol{AX - (L_{adj} - L_{raw})}$
- **Standardised residuals** = $\boldsymbol{V/\sigma_{obs}}$

**Figure 3:** The implications of selected variances when computing LSA [17]

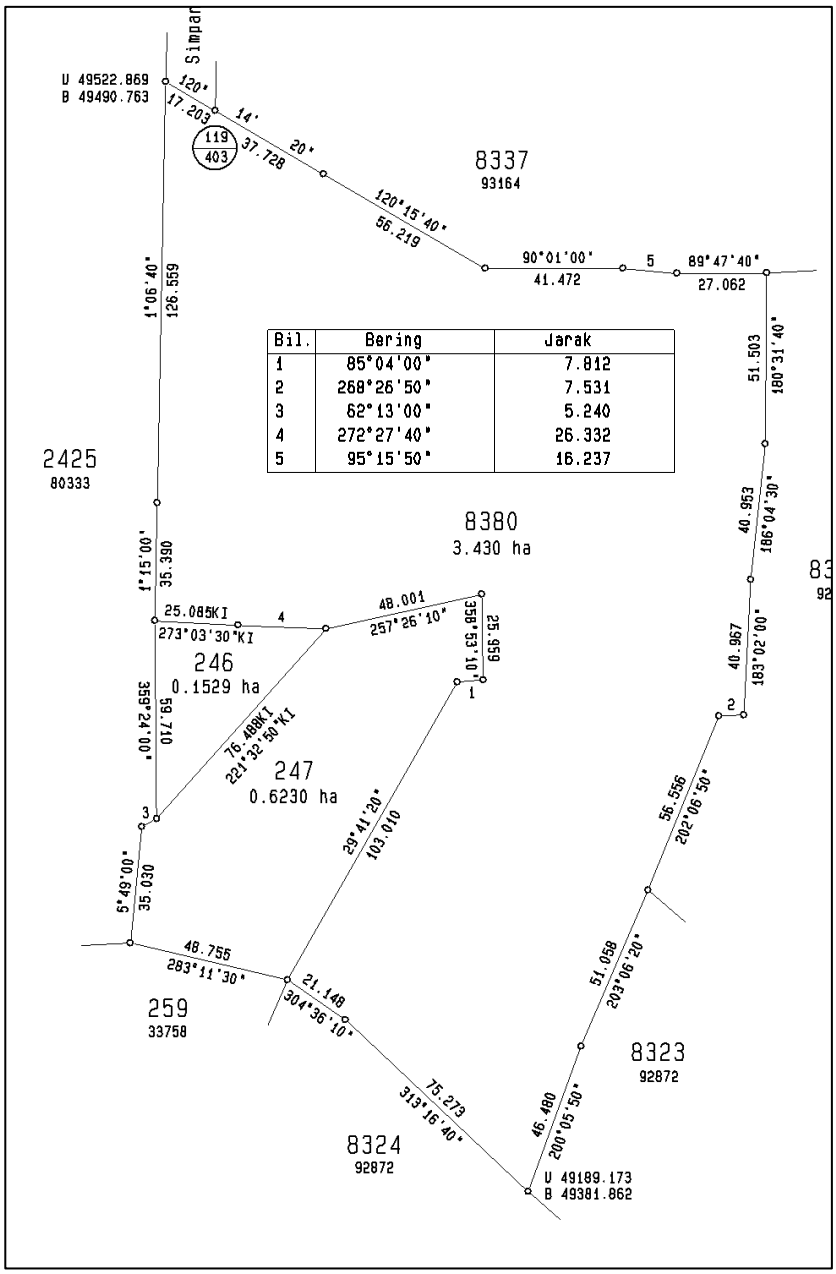



**Figure 4:** CP 93387 was employed to calculate the variance for the first classes

The second step is to determine whether or not the variance that has been isolated for the first classes and the second classes of CPs will perform admirably for the combined dataset if it is able to get a realistic variance. Figures 4 and 5 consists of first and second classes measurement certified plans (CP's), CP93387 (1st class) and CP33758 (2nd class), which were manufactured on July 21, 2009 and September 28, 2008, correspondingly. Mukim Seriab, Perlis is the location of the data has been collected. The results of the global test, standard deviations, residuals, and adjusted coordinates will all be included in the evaluation. These four (4) findings will be analysed to determine if the recommended variance from LS-VCE will be approved or rejected compared from homogenous variance for the combination classes of data.

Figure 6 from the LS-VCE was used to assist in the process of estimating two different sets of variances that is for the first class and another one for the second class of CP. The preliminary sigma in LS-VCE was defined as fifteen (15) seconds and 0.010m (CP 1st Class), whereas sixty (60) seconds and 0.010m (CP 2nd Class) regarding bearing and distance, correspondingly [18]. These values were determined by using the standard deviations that were defined in the circular that was issued by the Department of Surveying and Mapping Malaysia (DSMM). Two different configurations have been constructed in order to evaluate how well the LS-VCE system deals with multiple types of data quality that is specifically first and second classes datasets. The results of the first configuration have been evaluated using the homogeneous variance in accordance with the DSMM standard deviation. In contrast, the second setup has some of the variances determined according to the CPs class. As is seen once more in Figure 6, each and every one of these data will be placed through an LS-VCE computation for the purpose of getting an estimate of the realistic variance, and then it will be used once more in the LSA calculation. Table 1 displays the respective variances of the data collected from the first and second classes of CPs.

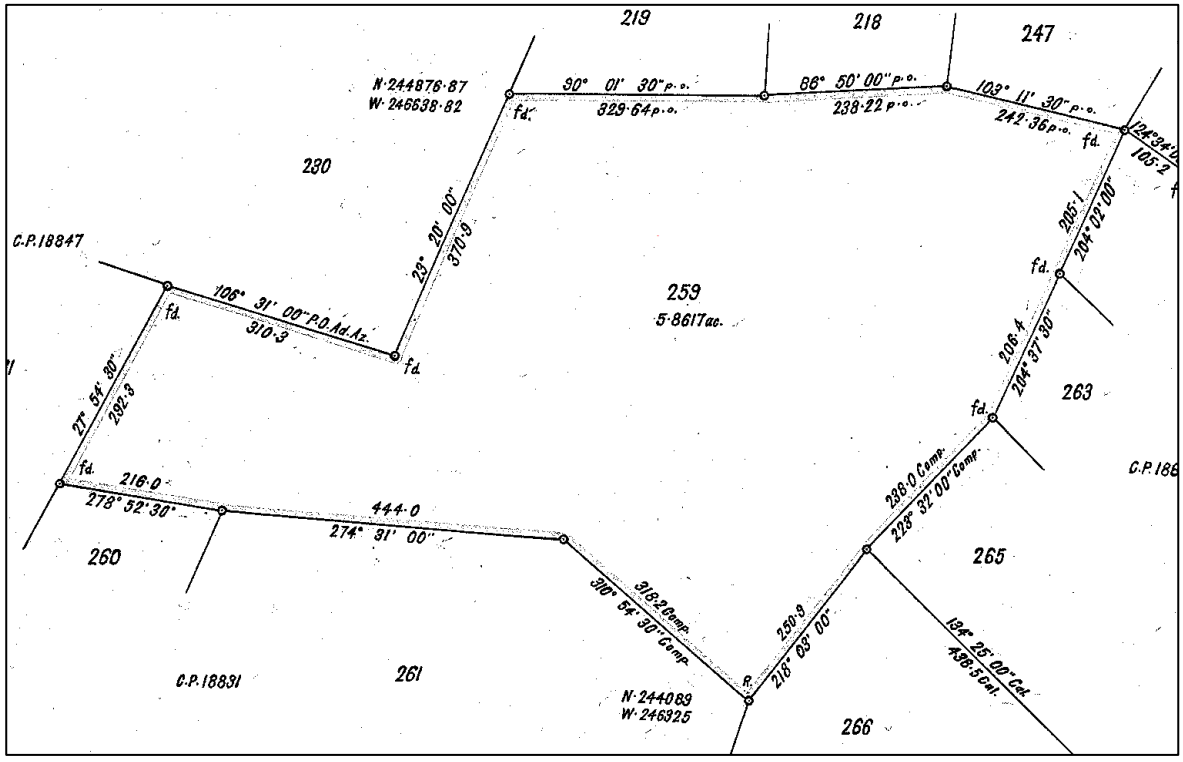



**Figure 5:** CP 33758 was employed to calculate the variance for CP second classes

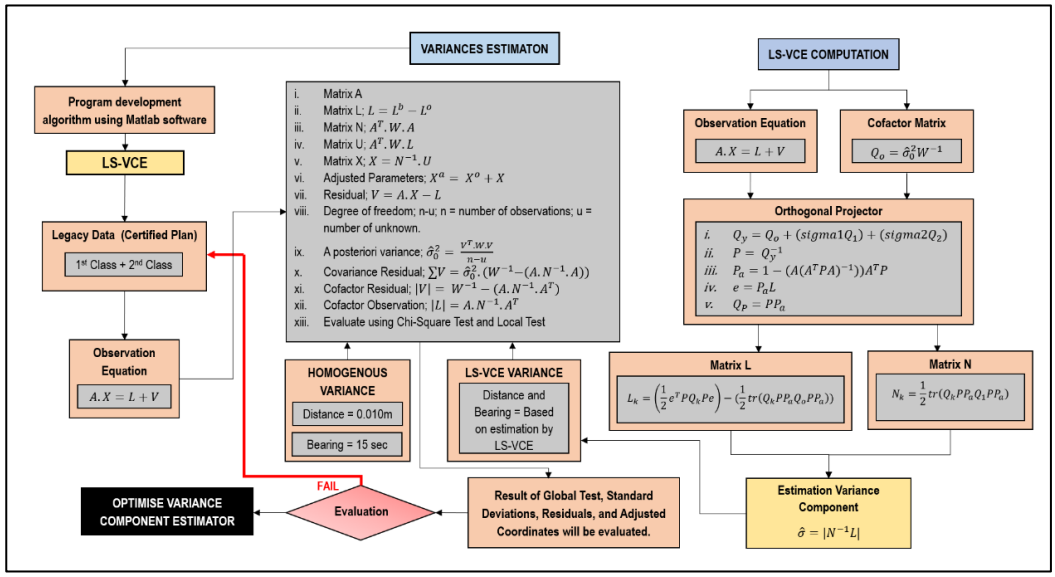
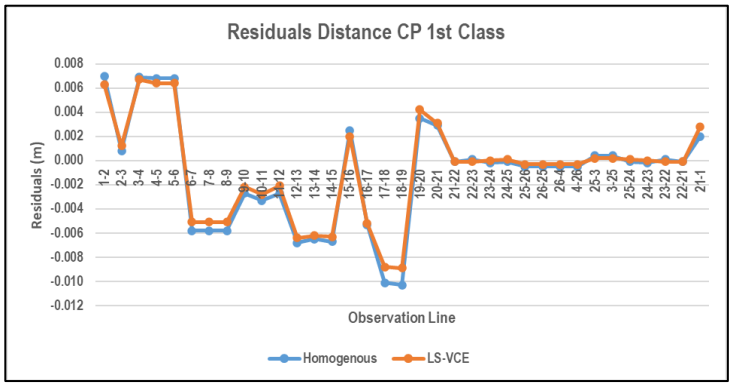


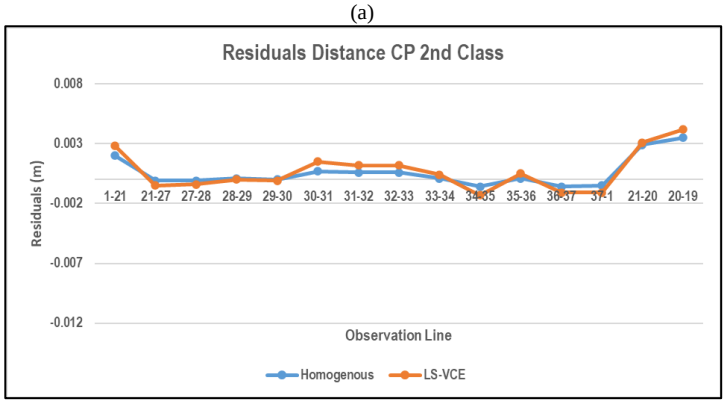

**Figure 6:** Flowchart of estimating variance using LS-VCE method

**Table 1:** Value two (2) sets of variances estimated by LS-VCE computation

| CP Classes | Standard Deviation Bearing | Standard Deviation Distance |
|---|---|---|
| 1st Class CP | 9” | 0.004m |
| 2nd Class CP | 35” | 0.010m |

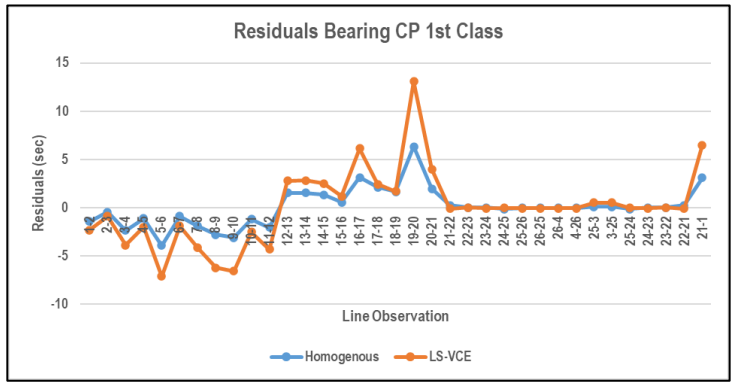



(a)

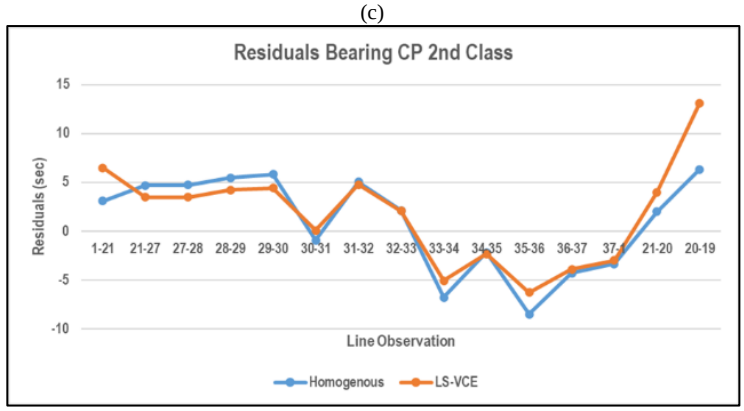



(b)

**Figure 7:** Residual results (a) Residual distance CP 1st class; (b) Residual distance CP 2nd class




(c)




(d)

**Figure 7:** Residual results (c) Residual bearing CP 1st class; (d) Residual bearing CP 2nd class

The LSA computation was carried out for two different CPs in line with their classes by the first configuration, which made use of the initial sigmas that were provided by DSMM. The use of the Chi-Square test is to identify the beginning of speculation and the final result that has been evaluated from the distribution pattern of the residuals and also the marked of error ellipse with the intention of evaluating the validity of the findings (which have been determined from standard deviations). In the second configuration, the variances of the first and second classes of CPs were estimated separately. This was performed in order that the range of data quality could be reduced. The implementation of a single LSA computing strategy enabled the modification of a particular set of variances on each data class that was carried out successfully.

The dependability of LS-VCE has been systematically examined using the results from both configurations. It is anticipated, based on the planned research, that the residuals of both classes will be fairly assessed in accordance with the measurement classes. Regarding parameters, the accuracy of the adjustment results or graphics may be a good indication of the data's quality.

## 4. Result and Analysis

During the experiment, the least square adjustment, often known as LSA, was carried out using the Star*Net software on the dataset that contained synthetic uncertainty. The global test showed various outcomes when homogenous variance implementations achieved a lower bound but different classes CP (first class and second class CP) of LS-VCE variances achieved a level passed with a 95% confidence level. Additional evaluations were carried out in order to provide a quantitative confirmation of the significance regarding the realistic variance with the experiment's data. This was conducted so that the cadastral network adjustments could be performed. The residual for distance and bearing is the main focus of the initial study. Following that, both the standard deviation as well as the error ellipse of the adjustment points for the multiclass CPs were examined. Finally, it is important to carefully examine coordinate changes for CPs to see if positional accuracy can be maintained.

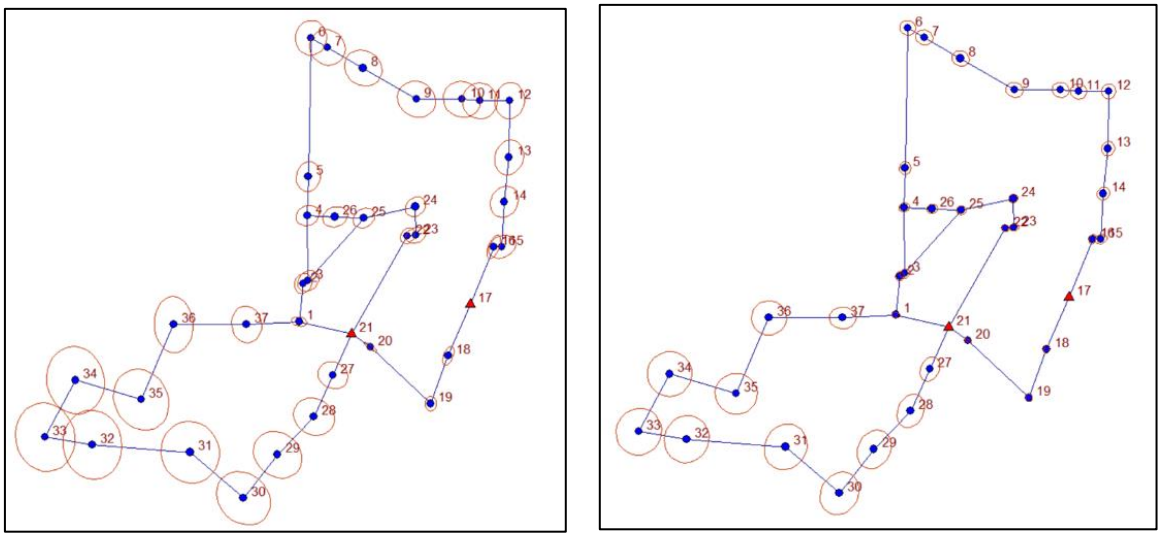
As can be seen, the Star*Net programme carried out an evaluation of the significance on residuals with the intent to determine the existence of outliers within the dataset by using the Star*Net outliers detection technique. The residuals of the

observations' distances and bearing for the CP 1st class and CP 2nd class are shown in Figure 7.

The results for residuals of distance were not significantly different after using the LS-VCE approach. Yet, there is a fluctuation value at line observation 19–20 for CP 1st Class that converges from 6.33 sec to 13.1 sec as well as the same for CP 2nd Class at the same line and value for residual bearing. Since there is a considerable difference between the homogeneous and LS-VCE variance for bearings, it can be inferred that the outcome of residuals distance and bearings may have a significant impact. Based on the results residuals shown, LS-VCE can be accepted rather than homogeneous variance due to the relationship between residuals and the result Chi-Square test. Since improvement from Lower Bound to 5% Level Passed, LS-VCE variance gives more realistic error in residuals and there is effective synchronisation on Chi-Square Result.

Following the application of LSA calculation, the standard deviation along with result of the error ellipse at point coordinates are depicted in Figure 8. Following the use of the LS-VCE approach, according to the results of the variance study, the standard deviation at stations 9 and 10 regarding homogenous variance with the LS-VCE variance increased dramatically from 0.0086m to 0.0197m for the CP 1st classes. In contrast, the stations at standard deviation with CP 2nd class has decreased from 0.0344m to 0.0243m throughout the course of this research. The error ellipse that was plotted illustrates that there is a considerable difference between separated and homogeneous variances. The calculated error ellipse demonstrates that separated variance shows a more realistic error of CP 2nd class. It occurs when the original homogeneous value that was utilised demonstrates the whole merging data classes correctness for second class data instead of opposed to the split variance which indicates according to classes data. To ensure that measurements are accurate and that they correlate with one another, the implementation of realistic variance is important for multi-classes data.

Figure 9 display the various coordinates for homogeneous variance together with LS-VCE variance that are related regarding CP 1st class as well as CP 2nd class, respectively. Considering Figure 7 demonstrated a considerable difference between the residual observations along line 19-20 in CP 1st class with line 20-19 in CP 2nd class as a result of classes, this resulted in revisions to the difference adjusted coordinates which utilised homogeneous and LS-VCE variances. Adjusting the coordinates at point 19 for both CP 1st class as well CP 2nd class brings an additional 0.0027 metres. The significant disparity between these two (2) points came about as a direct consequence of lines 19-20 for the distinct classes of CP 1st and CP 2nd Due to the estimated variance is calculated using LS-VCE represented by classes, adjusted coordinates on the LS-VCE variance have a better the potential of being accepted. This is due to the fact that LS-VCE variance produces a more accurate result for standard deviation.



(a) (b)

**Figure 8:** (a) error ellipses that were plotted for the homogeneous values (b) error ellipses for the adjusted values utilising LS-VCE CPs of the first and second classes

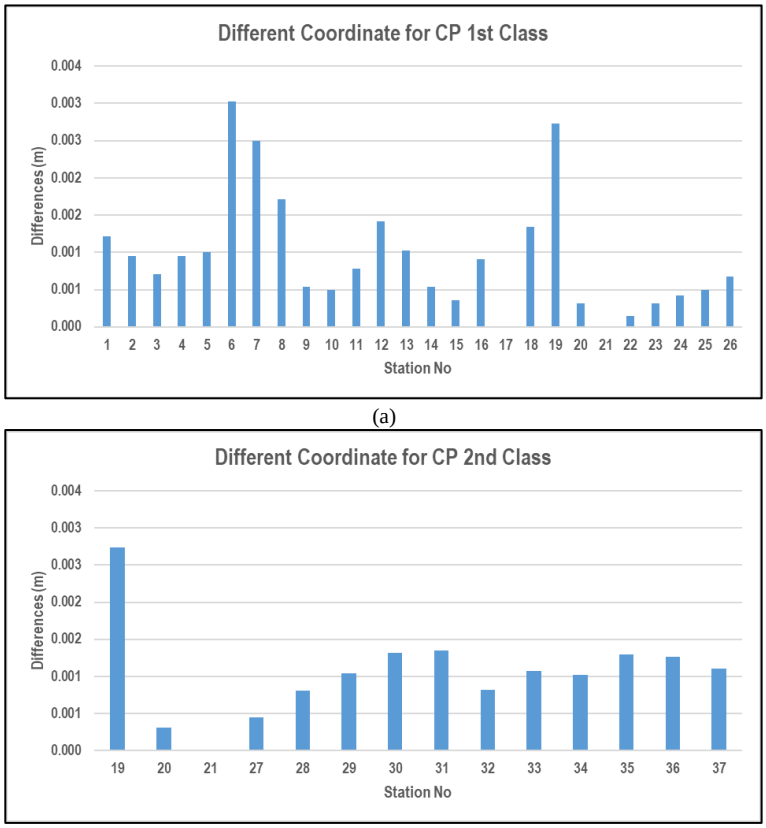



**Figure 9:** (a) different coordinates on CP 1st class (b) different coordinates on CP 2nd class between homogenous and LS-VCE variance

## 5. Conclusion

Throughout the confines of this research, the approach known as the least squares variance component estimation (LS-VCE) is utilised to produce an analysis of the predicted variance component of a cadastral network in multi-classes cadastral datasets. In Malaysia, homogenous variance has been used during adjustment for multi-classes cadastral data by the least square adjustment (LSA) method. According to the findings, LS-VCE is able to ascertain the realistic stochastic modelling value that should be utilised when estimating the variance for datasets that contain many classes of data. This research has the potential to additionally investigate whether or not separated variance is capable of producing genuine variance parallel with correction outcomes. This research suggests estimation variance by classes is essential in order to produce realistic adjustment results by data combination. It is anticipated that the propagation of errors would demonstrate significant changes in positioning uncertainty given the varying quality data in various CPs. According to the fundamental theory, the intentions of this research may be accomplished by estimating variances by using the LS-VCE method rather than using the homogenous variance in order to create accurate positional values.

## Acknowledgement

We wanted to take this opportunity to express our sincere gratitude to Universiti Teknologi Mara (UiTM) for the generous UiTM research funding grant [600-RMC/GIP 5/3 (038/2022)] where was given to this specific research project. We hope that UiTM will continue to support research in the future. In addition, we would like to extend our gratitude to our teammates from Universiti Teknologi MARA (UiTM), the Government sector from Department Survey and Mapping Malaysia (Perlis), and the Industrial sector from Geodetic Solutions Sdn. Bhd for offering information and insights that were of great assistance when carrying out this research.

**References**


[1] Forejt, M., Dolejš, M. and Raška, P., (2018). How Reliable is my Historical Land-Use Reconstruction? Assessing Uncertainties in Old Cadastral Maps. *Ecol. Indic.*, Vol. 94, 237–245. https://doi.org/10.1016/j.ecolind.2018.06. 053.

[2] Ignjatić, J., Nikolić, B., Rikalović, A. and Ćulibrk, D., (2018). Deep Learning For Historical Cadastral Maps Digitization: Overview, Challenges and Potential. *Comput. Sci. Res. Notes*, Vol. 2803, 42–47. http://dx.doi.org/10.24132/CSRN.2018.2803.6.

[3] Thompson, R. J., (2015). A Model for the Creation and Progressive Improvement of a Digital Cadastral Data Base. *Land use policy*, vol. 49, pp. 565–576, https://doi.org/10.1016/j.landusepol.2014.12.016.

[4] Kattan R.A. and Abdulrahman F.H. (2019). Accuracy Assessment of Duhok City Land use Official Maps. *Polytech. J.*, Vol. 9(2). 178–185. https://doi.org/10.25156/ptj.v9n2y2019.pp178-185.

[5] Jasmee, M. F., Rani, M. N. H. and Jaafar, J., (2017). Integration of Digital Cadastral Database Spatial Coordinates Towards Mapping in Google Earth. *J. Intelek*, Vol. 12(1), 50–54.

[6] Halim, N. Z. A., Sulaiman, S. A., Talib, K. and Ng, E. G., (2018). Identifying the Relevant Features of the National Digital Cadastral Database (NDCDB) for Spatial Analysis by Using the Delphi Technique. *IOP Conf. Ser. Earth Environ. Sci.*, Vol. 117(1). https://iopscience.iop.org/article/10.1088/1755-1315/117/1/012030.

[7] Jeffri, M., Hisham, O. and Joanes, J., (2017). Effectiveness of Localised Adjustment in Strengtening National Digital Cadastre Database. https://www.academia.edu/35949061/Effectiveness_of_Localised_Adjustment_in_Strengtening_National_Digital_Cadastre_Database.

[8] Yunus, M., Yusoff, M., Jamil, H., Zurairah, N. and Halim, A., (2013). Ekadaster : A Learning Experience for Malaysia. *The Importance of Geospatial Information*, 1–17.

[9] Yusof, M. Y. M. and Halim, N. Z. A., (2012). Unleashing the Full Potential of eKadaster on the Cadastral System of Malaysia. *Ninet. United Nations Reg. Cartogr. Conf. Asia Pacific*, 1–11.

[10] Yakubu, I., Ziggah, Y. Y. and Peprah, M. S., (2018). Adjustment of DGPS Data Using Artificial Intelligence and Classical Least Square Techniques. *J. Geomatics*, Vol. 12(1), 13–20.

[11] Bidi, N. K., Din, A. H. M., Som, Z. A. M. and Omar, A. H., (2019). Adjustment of Cadastral Network Using Least-Squares Variance Component Estimation. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Vol. XLII-4/W16, 161–167. https://doi.org/10.5194/isprs-archives-XLII-4-W16-161-2019

[12] Grodecki, J., (1997). *Estimation of Variance-Covariance Components for Geodetic Observations and Implications on Deformation Trend Analysis.* University of New Brunswick.

[13] Abbas, M. A., Hashim, N. M., Zulkifli, A., Sulaiman, S. A., Mustafar, M. A. and Ali, T. A. T., (2021). Variance Component Estimation Dilemma in Cadastral Network Adjustment. *2021 IEEE Int. Conf. Autom. Control Intell. Syst. I2CACIS 2021*. 414–417. http://dx.doi.org/10.1109/I2CACIS52118.2021.9495848.

[14] Amiri-Simkooei, A. R., (2007). *Least-Squares Variance Component Estimation: Theory and GPS Application*s. Ph.D Thesis, Delft University of Technolog.

[15] Bidi, N. K., (2019). *Least Squares Variance Component Estimation For Surveying Network Adjustment*. Masters Thesis, Divisions Built Environment, Universiti Teknologi Malaysia.

[16] Zangeneh-Nejad, F., Amiri-Simkooei, A. R., Sharifi, M. A. and Asgari, J., (2017). Recursive Least Squares with Real Time Stochastic Modeling: Application to GPS Relative Positioning. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, Vol. 42(4W4), 531–536. https://doi.org/10.5194/isprs-archives-XLII-4-W4-531-2017.

[17] Zulkifli, A., Abbas, M. A., Hashim, N. M., Mustafar, M. A. and Abdullah, A., (2022). Reliability of Using LS-VCE Computation in Deriving Variances for Multi-Classes Dataset. *IOP Conf. Ser. Earth Environ. Sci.*, Vol. 1051(1). http://dx.doi.org/10.1088/1755-1315/1051/1/012002.

[18] JUPEM, (2009). Guidelines for the Practice of Cadastral Survey Work in the eCadaster Environment. *Circ. Dir. Gen. Surv. Mapp.*, Vol. 6.