ISSN 2278-3091
Volume 9, No.1.2, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse0391.22020.pdf
https://doi.org/10.30534/ijatcse/2020/0391.22020

A Statistical ModellingApproaches on Tidal Analysis and Forecasting

Firdaus Mohamad Hamzah[1,2*], Hazrina Tajudin[3], Mohd Khairul Amri Kamarudin[4], Norazman Arbin[5], Siti Hawa Mohd Yusof[6]

[1]Department of Engineering Education, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi Selangor Malaysia.
[2]Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 Bangi Selangor Malaysia.
[3]Smart and Sustainable Township Research Centre, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi Selangor Malaysia.
[4]Faculty of Applied Social Science and East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Terengganu, Malaysia.
[5]Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tanjung Malim, Perak, Malaysia.
[6]Department of Science and Biotechnology, Faculty of Engineering and Life Sciences, Universiti Selangor, Jalan Timur Tambahan, 45600 Bestari Jaya, Selangor Malaysia
[*]Corresponding Author:fir@ukm.edu.my

ABSTRACT

Increase in the number of population in the low-elevation coastal zone has increase the importance to reduce the risk of coastal and nuisance flooding, especially during high tide. This study attempts to generate forecast of high tide data using several statistical approaches such as seasonal naive model, Holt-Winter, Theta method and seasonal autoregressive and moving average method. Based on the decomposition plot using additive components of the time series, there are seasonal components in each data sets and increasing trend can be observed at Permatang Sedepa and Bagan Datuk, while decline follows by slowly increasing trend can be seen at Pelabuhan Klang station. Among all methods applied to the time series data, Theta method gives the lowest error for Pelabuhan Klang and Permatang Sedepa with accuracy of 0.0403 and 0.0457 respectively, while Holt-Winter method gives the lowest error for high tide data at Bagan Datuk with accuracy of 0.0456. Malacca straits serves various purposes including shipping, especially at Pelabuhan Klang, High number of activities in the area had caused unexpected outcome such as land subsidence, coastal erosion, deterioration of natural and man-made barrier, floods and inundation of land which indirectly influences the physical of the port.

Key words: Tidal station; Statistical Approaches; Forecasting

1. INTRODUCTION

Accurate streamflow forecasts are crucial to the decision-making concerning flood control, water supplies, navigation, drought mitigation, and hydropower generation[1]. Study done by Neumann et.al, [2] show a significant increases in coastal population living in the low-elevation coastal zone and these people being potentially exposed to coastal flood events. The estimates of future flood risk due to climate change will also affect the world's major cities, such as Guangzau, New York, New Orleans, and Mumbaiwhich are situated along low-lying coastal zone. Less developed countries, especially Arfica and Asia are more exposed to flooding than more developed regions.

However, rich cities are better protected than poorer ones [3]. In addition, the cumulative damage of nuisance flooding, which occurs during high tides with sea level rise, has a potential to exceed the storm floods as climate change escalate[4][5]. Therefore, it is importantto have a better understanding on infrastructure planning in the local estuarine with regards to reduce the flood risk[6].

Forecasting on sea level has been investigated by number of researchers [7][8], as it is very important for future planning and also flood hazards. Previous analyses of sea levels focus on examining global and local sea level rise. Srivet et.al [9], study on robust decision making on global sea level rise. Yang et.al, [10] examine changing sea levels due to different ocean dynamics. Velgara et al, [11] and Quilfen and Chapron, [12] uses global satellite altimeter data to access the wave heights. While Makris et.al [13] and Ezer, [14] studies on storm surges in gulfs and coastal areas.

Tides represents rise and fall of seas and oceans surface. Basically, it is due to the gravitational attraction (pull) of the moon and sun on the rotating earth. Tidal analysis also important since it is relate to the removal and passage of sediments andcontaminants, off-shore production, and oceanagraphy[15]. In previous studies, harmonic analysis was used extensively in tidal prediction [16].However, this method required a large number of tide records per site.

In this study, the conventional forecasting techniques such as Seasonal Naive, Holt-Winter, Theta model and and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) being applied to the high tide data. Since it is considered as univariate model, where the input and output has to depend on high tide series alone with a certain number of time lags. Despite the fact that the conventional techniques not widely used in tidal data, these methods extremely used in other time series data such as streamflow[17], rainfall[18], groundwater[19], drought [20] and, temperature [21].

Therefore, the objectives of this study are to generate forecast of high tide data using conventional statistical approches. Next, the high tide forecast are compared with forecast models and will be evaluated using statistical measurement and discussion will be presented in the last section.

2. METHODOLOGY

Before any model is fitted into the data series, the characteristics of the data need to be evaluated. There are three characteristics which are:

1. The data series are random.
2. The data series contains trend, whether it is positive or negative.
3. It has seasonality factor, which implies regular pattern in the data sets.

In order to generate forecast data, there are several statistical approaches considered in this study such as Seasonal Naive, Holt-Winter, Theta and Seasonal Auto-Regressive Integrated Moving Average (SARIMA)[22].

2.1 Seasonal Naive

Mathematically, this method is derived as follow:

$$\hat{Y}_{T+h|T} = Y_{T+h-m(k+1)} \quad (1)$$

where

2.2 Holt-Winter

This method is implemented on data series with trend and seasonality[22]. Mathematically, the HW method can be calculated using the following equation:

$$S_t = \propto \frac{y_t}{I_{t-L}} + (1-\propto)(S_{t-1} + b_{t-1}) \quad (2)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \quad (3)$$

$$I_t = \beta\frac{y_t}{S_t} + (1-\beta)I_{t-L} \quad (4)$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad (5)$$

where y represents the observation, S is the smoothed observation, b is the trend factor, I is the seasonal index, F is forecast at m periods ahead and t is the index denoting a time period. α, β and γ is the constant obtained when the error is minimized.

The initial value for each, overall, trend and seasonal smoothing is calculated mathematically using the following:

Initial trend factor,

$$b = \frac{1}{L}\left(\frac{y_{L+1}-y_1}{L} + \frac{y_{L+2}-y_2}{L} + \cdots + \frac{y_{L+L}-y_L}{L}\right) \quad (6)$$

Initial seasonal index, for example considering four quarters,

$$I_1 = \frac{\frac{y_1}{A_1}+\frac{y_5}{A_2}+\frac{y_9}{A_3}+\frac{y_{13}}{A_4}+\frac{y_{17}}{A_5}+\ldots+\frac{y_{21}}{A_N}}{N} \quad (7)$$

$$I_2 = \frac{\frac{y_2}{A_1}+\frac{y_6}{A_2}+\frac{y_{10}}{A_3}+\frac{y_{14}}{A_4}+\frac{y_{18}}{A_5}+\ldots+\frac{y_{22}}{A_N}}{N} \quad (8)$$

$$I_3 = \frac{\frac{y_3}{A_1}+\frac{y_6}{A_2}+\frac{y_{11}}{A_3}+\frac{y_{15}}{A_4}+\frac{y_{19}}{A_5}+\ldots+\frac{y_{23}}{A_N}}{N} \quad (9)$$

$$I_4 = \frac{\frac{y_4}{A_1}+\frac{y_6}{A_2}+\frac{y_{12}}{A_3}+\frac{y_{16}}{A_4}+\frac{y_{20}}{A_5}+\ldots+\frac{y_{24}}{A_N}}{N} \quad (10)$$

where $A_1, A_2, \ldots, A_N$represents average for each year computed in the study.

2.3Theta

Theta model is applied based on modification of local curvature of the time series. The modification is represented as Theta-coefficient, θ [23]. It is directly implemented on the second differencing of the time series. Mathematically, it is represented in equation 11.

$$X''_{new}(\theta) = \theta * X''_{data} \quad (11)$$

where

$$X''_{data} = X_t - 2X_{t-1} + X_{t-2} \quad (12)$$

Time series is deflated when the local curvature slowly decreasing. Small θ value indicates larger deflation rate. $\theta = 0$ indicates a linear regression line. In contrast, increase in local curvature implies dilation of time series.

2.4 Seasonal Auto-Regressive Integrated Moving Average

Seasonal implies a consistent repeatition of pattern over a particular period time. The seasonal AR and MA terms predict $x_t$using data values and errors at lags multiple of the seasonality span. By using monthly data, the first order seasonal autoregressive model, SAR (1) is use $x_{t-12}$to predict x_t. While for seasonal moving average, SMA (1), it uses past error,$w_{t-12}$to calculate forx_t[24]. Mathematically, the SARIMA model is formulated as follow:

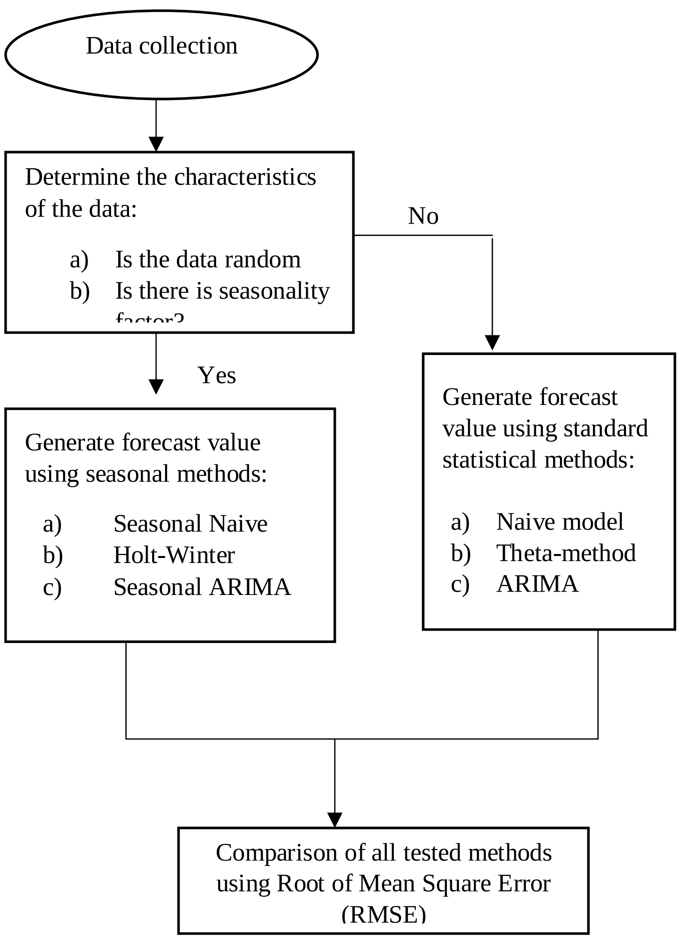
$$SARIMA\ (p, d, q) \times (P, D, Q)^s \quad (13)$$

$$\phi_p(L^s)\varphi_p(L)(1-L)^d(1-L^s)^D y_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t \quad (14)$$

The estimated value generated by each model is evaluated using Root Mean Square Error. Mathematically, it is calculated using equation 15.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(A_i - F_i)^2}{N}} \quad (15)$$

where $A_i$represents the actual observation, while $F_i$represents the forecasted value. Basically, the flow chart can be described in the following figure:



3. RESULTS AND DISCUSSION

3.1 Descriptive Analysis

This study evaluates daily high tide data between 2004 to 2017 at three tidal stations namely Pelabuhan Klang, Permatang Sedepa and Bagan Datuk station. The maximum, minimum and mean of daily data is shown in the boxplot in Figure 2.

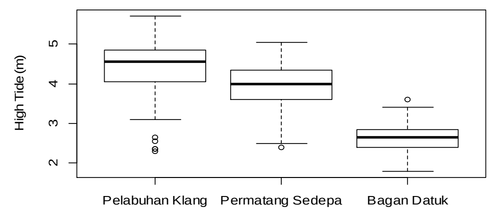


Figure 2: Boxplot of daily data at each tidal station

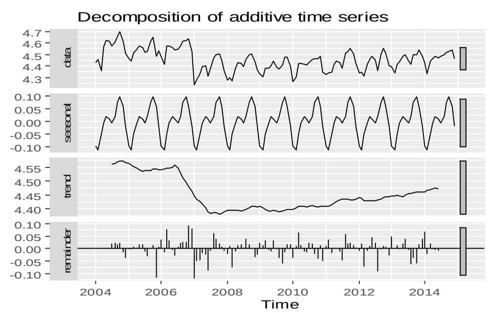


Figure 3: Additive components of time series at Pelabuhan Klang

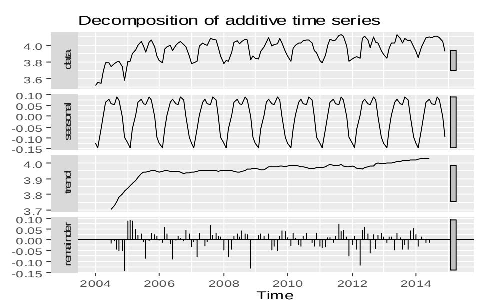


Figure 4: Additive components of time series at Permatang Sedepa station

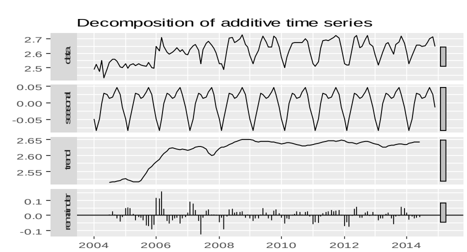


Figure 5: Additive components of time series at Bagan Datuk station

Based on the decomposition plot using additive components of the time series, there are seasonal components in each data sets and increasing trend can be observed at Permatang Sedepa and Bagan Datuk, while decline follows by slowly increasing trend can be seen at Pelabuhan Klang station.

3.2 Comparison between methods

The accuracy of each model is evaluated using root mean square error (RMSE) and the results are recorded in Table 1.

Table 1: Evaluation of accuracy for different statistical model

Station	Seasonal Naive	Holt-Winter	Theta	SARIMA
Pelabuhan Klang	0.075	0.044	**0.040**	0.046
Permatang Sedepa	0.098	0.049	**0.046**	0.049
Bagan Datuk	0.055	**0.046**	0.046	0.053

The performance of each model can be clearly seen in Figure 5, 6 and 7.

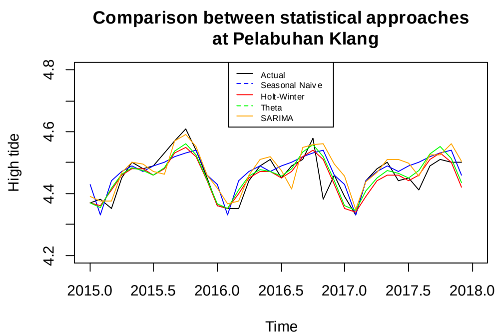


Figure 6: Comparison of several statistical approaches at Pelabuhan Klang

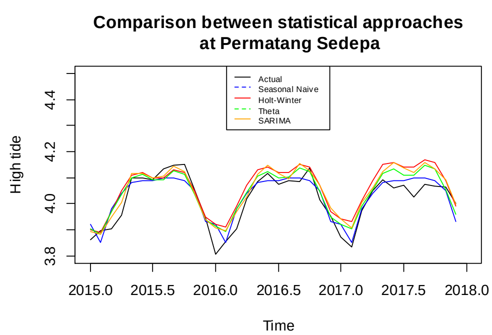


Figure 7: Comparison of several statistical approaches at Permatang Sedepa

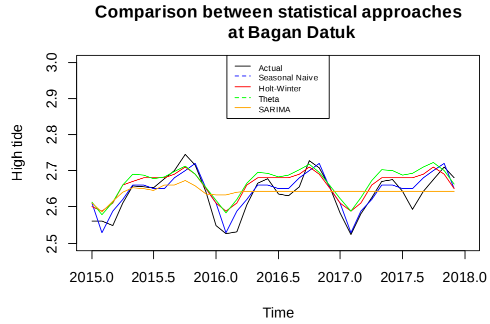


Figure 8: Comparison of several statistical approaches at Bagan Datuk

Malacca straits act as an important route for shipping and the number of vessel which past through the straits are reported to increase from year to year. Currently, the number of vessel which navigates along the straits are around 6000 per month. It is important to ensure minimal environmental degradation and also the demography along the straits so that these vessels could pass through the straits safely. Pelabuhan Klang also serves various purpose and greatly functional. Large volume of activities around the port had caused unexpected outcome such as land subsidence, coastal erosion, deterioration of natural and man-made barrier,floods and inundation of land which indirectly influences the physical of the port. Other environmental factor such as rise in sea level also triggers the changes.

Forecasting usually applies the usage of statistical approaches or machine learning[25]. The usage of statistical approaches involves assumptions and restrictions, however, it helps to identify unique pattern in the data series. All hydrological data series are unique, but using statistical methods, the location with similar physical character can be classified and evaluated as one. There are several advantages on conducting forecasting process via statistical methods. In comparison with machine learning, statistical method can projects short term and long term data series[5].Holt-Winter is one of the traditional statistical approach with great performance, compared to machine learning.

4.CONCLUSION

Most studies today involve the use of neural networks rather than statistical methods. The accuracy of the usage of neural network methods is limited to time series data that do not have extreme value, known as outliers. However, most hydrological data consist of extreme event, which in turn makes statistical approaches more reliable. Thus, it is important to consider the use of conventional statistical approaches to predict the tidal data. In this study, Holt-Winter and Theta method is best fitted to be applied to high tide data, with seasonal components. In addition, these conventional method can be compared with the application of appropriate neural network model.

ACKNOWLEDGEMENT

The authors would like to thank the Earth Observation Centre, UniversitiKebangsaan Malaysia and the Department of Survey and Mapping Malaysia (JUPEM) for providing the data for this research. The authors wish to express their gratitude to Yayasan Sime Darby (YSD) for supporting this research via research grant ZF-2017-008.

REFERENCES

[1] W. Fang *et al.*, "**Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models**," *J. Hydrol.*, vol. 568, pp. 534–550, 2019. https://doi.org/10.1016/j.jhydrol.2018.11.020

[2] B. Neumann, A. T. Vafeidis, J. Zimmermann, and R. J. Nicholls, "**Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment**," *PLoS One*, vol. 10, no. 3, 2015.

[3] S. Hallegatte, C. Green, R. J. Nicholls, and J. Corfee-Morlot, "**Future flood losses in major coastal cities**," *Nat. Clim. Chang.*, vol. 3, no. 9, pp. 802–806, 2013.

[4] H. R. Moftakhari, A. AghaKouchak, B. F. Sanders, and R. A. Matthew, "**Cumulative hazard: The case of nuisance flooding**," *Earth's Futur.*, vol. 5, no. 2, pp. 214–223, 2017.

[5] M. K. Singla, J. Gupta, and P. Nijhawan, "**Comparative Study on Backpropagation and Levenberg Marquardt Algorithm on Short Term Load Forecasting**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 2, pp. 194–202. https://doi.org/10.30534/ijatcse/2019/14822019

[6] R. Q. Wang, M. T. Stacey, L. M. M. Herdman, P. L. Barnard, and L. Erikson, "**The Influence of Sea Level Rise on the Regional Interdependence of Coastal Infrastructure**," *Earth's Futur.*, vol. 6, no. 5, pp. 677–688, 2018.

[7] M. El-Diasty, S. Al-Harbi, and S. Pagiatakis, "**Hybrid harmonic analysis andwavelet network model for sea water level prediction**," *Appl. Ocean Res.*, vol. 70, pp. 14–21, 2018. https://doi.org/10.1016/j.apor.2017.11.007

[8] C. Berrett *et al.*, "**Spatial Prediction of Sea Level Trends**," 2018.

[9] R. L. Sriver, R. J. Lempert, P. Wikman-Svahn, and K. Keller, **Characterizing uncertain sea-level rise projections to support investment decisions**, vol. 13, no. 2. 2018.

[10] J. Yang, D. S. Abbot, D. D. B. Koll, Y. Hu, and A. P. Showman, "**Ocean Dynamics and the Inner Edge of the Habitable Zone for Tidally Locked Terrestrial Planets**," *Astrophys. J.*, vol. 871, no. 1, p. 29, 2019.

[11] O. Vergara, R. Morrow, I. Pujol, G. Dibarboure, and C. Ubelmann, "**Revised Global Wave Number Spectra From Recent Altimeter Observations**," *J. Geophys. Res. Ocean.*, vol. 124, no. 6, pp. 3523–3537, 2019.

[12] Y. Quilfen and B. Chapron, "**Ocean Surface Wave-Current Signatures From Satellite Altimeter Measurements**," *Geophys. Res. Lett.*, vol. 46, no. 1, pp. 253–261, 2019.

[13] C. Makris, Y. Androulidakis, V. Baltikas, Y. Kontos, T. Karambas, and Y. Krestenitis, "**HiReSS: Storm surge simulation model for the operational forecasting of sea level elevation and currents in marine areas with harbor works**," *Proc. 1st Int. Sci. Conf. Des. Manag. Port Coast. Offshore Work.*, vol. 1, pp. 11–15, 2019.

[14] T. Ezer, "**On the interaction between a hurricane, the Gulf Stream and coastal sea level**," *Ocean Dyn.*, vol. 68, no. 10, pp. 1259–1272, 2018.

[15] Akhil Muhammad Salim, G. S. Dwarakish, Liju K. V., Justin Thomas, Gayathri Devi, and Rajeesh R., "**Weekly prediction of tides using Neural networks**," in *Procedia Engineering*, 2015, vol. 116, pp. 678–682.

[16] T. L. Lee, "**Back-propagation neural network for long-term tidal predictions**," *Ocean Eng.*, vol. 31, no. 2, pp. 225–238, 2004.

[17] Z. Zhang, Q. Zhang, and V. P. Singh, "**Univariate streamflow forecasting using commonly used data-driven models: literature review and case study**," *Hydrol. Sci. J.*, vol. 63, no. 7, pp. 1091–1111, 2018.

[18] L. Parviz and K. Rasouli, "**Development of Precipitation Forecast Model Based on Artificial Intelligence and Subseasonal Clustering**," *J. Hydrol. Eng. Eng.*, vol. 24, no. 12, pp. 1–13, 2019.

[19] S. Mohanasundaram, G. Suresh Kumar, and B. Narasimhan, "**A novel deseasonalized time series model with an improved seasonal estimate for groundwater level predictions**," *H2Open J.*, vol. 2, no. 1, pp. 25–44, 2019. https://doi.org/10.2166/h2oj.2019.022

[20] H.-F. Yeh and H.-L. Hsu, "**Stochastic Model for Drought Forecasting in the Southern Taiwan Basin**," *Water*, vol. 11, no. 10, p. 2041, 2019.

[21] D. T. Meshram, S. D. Gorantiwar, and N. Bake, "**Forecasting of Air Temperature of Western Part of Maharashtra, India,**" *Int. J. Sci. Environ. Technol.*, vol. 8, no. 1, pp. 201–217, 2019.

[22] P. M. MaCąira, F. L. C. Oliveria, and R. C. Souza, "**Forecasting natural inflow energy series with multi-channel singular spectrum analysis and bootstrap techniques**," *Int. J. Energy Stat.*, vol. 3, no. 1, pp. 1–17, 2015.

[23] V. Assimakopoulos and K. Nikolopoulos, "**The theta model: A decomposition approach to forecasting**," *Int. J. Forecast.*, vol. 16, no. 4, pp. 521–530, 2000.

[24] M. Valipour, "**Long-term runoff study using SARIMA and ARIMA models in the United States**," *Meteorol. Appl.*, 2015.

[25] Y. H. T. Louis, Kuok King Kuok, Monzur Imteaz, Wai Yan Lai, and Derrick Kuok Xiong Ling, "**Development of whale optimization neural network for daily water level forecasting**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 3, pp. 354–362, May 2019. https://doi.org/10.30534/ijatcse/2019/04832019