

## **DETERMINATION OF THE BEST SINGLE IMPUTATION ALGORITHM FOR MISSING RAINFALL DATA TREATMENT**

(Penentuan Al-Khwarizmi Imputasi Tunggal bagi Merawat Data Hujan Lenyap yang Terbaik)

GAMIL ABDULRAQEB ABDULLAH SAEED, ZUN LIANG CHUAN, ROSLINAZAIRIMAH  
ZAKARIA, WAN NUR SYAHIDAH WAN YUSOFF & MOHD ZUKI SALLEH

### *ABSTRACT*

The presence of missing rainfall data is inevitable due to error of recording, meteorological extremes and malfunction of instruments. Consequently, a competent imputation algorithm for missing data treatment algorithm is very much needed. There are several such efficient algorithms which have been introduced in earlier studies. However, the limitations of current algorithms are they are highly dependent on the information and homogeneity of adjoining rainfall stations. Therefore, this study is intended to introduce several single imputation algorithms for missing data treatment, which believed to be more competent in treating missing daily rainfall data without the need to depend on the information of adjoining rainfall stations. The proposed algorithms use descriptive measures of the data, including arithmetic means, geometric means, harmonic means, medians and midranges. These algorithms are tested on hourly rainfall data records from six selected rainfall stations located in the Kuantan River Basin. Based on the analysis, the proposed singular imputation algorithms, which treated missing data by geometric means, harmonic means and medians are more superior compared to the other imputation algorithms, irrespective of missing rates and rainfall stations.

*Keywords:* missing data; rainfall data; numerical descriptive measures; Kuantan River Basin

### *ABSTRAK*

Kehadiran data hujan lenyap tidak dapat dielakkan berpunca dari ralat perekodan, kejadian ekstrem meterologi dan kegagalan peralatan. Natiujahnya, suatu al-Khwarizmi imputasi bagi data lenyap yang cekap amat diperlukan. Terdapat beberapa al-Khwarizmi imputasi yang cekap telah diperkenalkan dalam kajian lepas. Walau bagaimanapun, kelemahan bagi al-Khwarizmi tersebut adalah mereka amat bersandar pada maklumat dan kehomogenan stesen-stesen hujan berhampiran. Oleh itu, kajian ini bertujuan untuk memperkenalkan beberapa al-Khwarizmi imputasi tunggal bagi data lenyap, yang mana ia mampu merawat data hujan harian lenyap tanpa bergantung pada maklumat dan kehomogenan stesen-stesen hujan yang berhampiran. Al-Khwarizmi yang diperkenalkan merawat data lenyap dengan menggunakan sukatan perihalan berangka, iaitu min aritmetik, min geometri, min harmonik, median dan julat tengah. Al-Khwarizmi yang dicadangkan ini dinilai menggunakan data hujan per jam dari enam stesen hujan terpilih yang terletak di Lembangan Sungai Kuantan. Hasil analisis mendapati al-Khwarizmi imputasi tunggal dengan menggunakan min geometri, min harmonik dan median merupakan al-Khwarizmi yang paling cekap berbanding dengan al-Khwarizmi imputasi lain tanpa mengira kadar data lenyap mahupun stesen hujan yang terlibat.

*Kata kunci:* data lenyap; data hujan; ukuran perihalan berangka; Lembangan Sungai Kuantan

### **References**

- Ahrens B. 2006. Distance in spatial interpolation of daily rain gauge data. *Hydrology and Earth System Sciences* **10**: 197-208.
- Alam M.M., Morshed G., Siwar C. & Murad M.W. 2012. Initiatives and challenges of agricultural crop sector in East Coast Economic Region (ECER) development projects in Malaysia. *American-Eurasian Journal Agriculture & Environmental Sciences* **12**(7): 922-931.

- Azman M.A., Zakaria R. & Radi N.F.A. 2015. Estimation of missing rainfall data in Pahang using modified spatial interpolation weighting methods. In *AIP Conference Proceedings*, **1643**, pp. 65-72.
- Chen F-W. & Liu C-W. 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment* **10**(3): 209-222.
- Dray S. & Josse J. 2015. Principal component analysis with missing values: a comparative survey of methods. *Plant Ecology* **216**(5): 657-667.
- Jamaludin S. & Suhaimi H. 2013. Spatial interpolation on rainfall data over Peninsular Malaysia using ordinary kriging. *Jurnal Teknologi* **63**(2): 51-58.
- Masseran N., Razali A.M., Ibrahim K., Zaharim A. & Sopian K. 2013. Application of the single imputation method to estimate missing wind speed data in Malaysia. *Research Journal of Applied Sciences, Engineering and Technology* **6**(10): 1780-1784.
- Nasir M.F.M., Zali M.A., Juahir H., Hussain H., Zain S.M. & Ramli N. 2012. Application of receptor models on water quality data in source apportionment in Kuantan River Basin. *Iranian Journal of Environmental Health Science & Engineering* **9**:18, doi:10.1186/1735-2746-9-18.
- Ozelkan E., Bagis S., Ozelkan E.C., Ustundag B.B., Yucel M. & Ormeci C. 2005. Spatial interpolation of climatic variables using land surface temperature and modified inverse distance weighting. *International Journal of Remote Sensing* **36**(4): 1000-1025.
- Radi N.F.A., Zakaria R. & Azman M.A. 2015. Estimation of missing rainfall data using spatial interpolation and imputation methods. In *AIP Conference Proceedings*, **1643**, pp. 42-48.
- Simanton J.R. & Osborn H.B. 1980. Reciprocal-distance estimate of point rainfall. *Journal of Hydraulic Engineering* **106**: 1242-1246.
- Suhaila J., Sayang M.D. & Jemain A.A. 2008. Revised spatial weighting methods for estimation of missing rainfall data. *Asia-Pacific Journal of Atmospheric Sciences* **44**(2): 93-104.
- Tabios G. & Salas J.D. 1985. A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resources Bulletin* **21**(3): 365-380.
- Tang W.Y., Kassim A.H.M. & Abubakar S.H. 1996. Comparative studies of various missing data treatment methods-Malaysia experience. *Atmospheric Research* **42**(1-4): 247-262.
- Teegavarapu R.S.V. & Chandramouli V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology* **312**(1-4): 191-206.
- Troyanskaya O., Cantor M., Sherlock G., Brown P., Hastie T., Tibshirani R., Botstein D. & Altman R.B. 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**(6): 520-525.
- Win N.L. & Win K.N. 2014. The probability distributions of daily rainfall for Kuantan River Basin in Malaysia. *International Journal of Science and Research* **3**(8): 977-983.
- Young K.C. 1992. A three-way model for interpolating for monthly precipitation values. *Monthly Weather Review* **120**(11): 2561-2569.
- Zainuri N.A., Jemain A.A. & Muda N. 2015. A comparison of various imputation methods for missing values in air quality data. *Sains Malaysiana* **44**(3): 449-456.

*Faculty of Industrial Sciences and Technology*

*Universiti Malaysia Pahang*

*Lebuhraya Tun Razak*

*26300 Gambang, Kuantan*

*Pahang DM*

*MALAYSIA*

*E-mail: hamadijamil2010@gmail.com, chuanzunliang@ump.edu.my\*, roslinazairimah@ump.edu.my, wnsyahidah@ump.edu.my*

---

\*Corresponding author