Machine learning model for detecting hydrogen leakage from hydrogen pipeline using physical modeling

Yuki Suzuki[1], Jo Nakayama[2], Tomoya Suzuki[2], Tomoya Soma[3], Yu-ichiro Izato[4], Atsumi Miyake[4]

[1]*Graduate School of Environment and Information Sciences, Yokohama National University,79-5, Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa, 240-8501, Japan*

suzuki-yuki-cs@ynu.jp

[2] *Center for Creation of Symbiosis Society with Risk, Institute for Multidisciplinary Sciences, Yokohama National University, 79-5, Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa, 240-8501, Japan*

nakayama-jo-sj@ynu.ac.jp
suzuki-tomoya-sk@ynu.ac.jp

[3]*NEC, 7-2, Shiba 5-chome Minato-ku, Tokyo, 108-8001, Japan*

tomoya@arogopilot.com

[4] *Faculty of Environment and Information Sciences, Yokohama National University,79-5, Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa, 240-8501, Japan*

izato-yuichiro-tk@ynu.ac.jp
miyake-atsumi-wp@ynu.ac.jp

ABSTRACT

Hydrogen pipelines (HPL) are one of the hydrogen transportation systems for realizing a hydrogen society. Hydrogen leakage from HPL is a challenge because hydrogen has a wide flammable range and low minimum ignition energy. Thus, hydrogen leakage from the HPL must be rapidly detected, and appropriate actions should be taken. Leakage detection is important for safe operation of HPL. The basic leakage detection method for HPL involves monitoring the pressure and flow rate values of the sensors. However, in some cases, it is difficult to distinguish between non-leakage and leakage conditions using this method. In this study, we focus on a leakage detection method using machine learning (ML) based on the relationship between pressure and flow rate data. There are two challenges in applying the ML-based leak detection method to an HPL. First, there are insufficient operational data for ML during the process-design stage. Secondly, it is difficult to obtain the pressure and flow rate behaviors during hydrogen leakage because leakage does not occur frequently. Consequently, this study employed an unsupervised ML method based on data simulated using a physical model of the HPL. First, a physical model of the HPL (HPL model) was constructed, and an ML model for leak detection was constructed based on the data simulated by the HPL model. The leak detection capability of the ML model was verified by comparing the anomaly scores of the non-leakage and leakage conditions. From the results, the ML model can distinguish between non-leakage and leakage behaviors and identify leakage points under certain conditions. This method can contribute to the optimization of the sensors required for leak detection during the process design stage.

1. INTRODUCTION

Hydrogen pipelines (HPL) have attracted significant attention as one of the hydrogen transportation methods. HPL can transport large amounts of hydrogen and therefore have low transportation costs (Faye, Szpunar, & Fduok, 2022). HPL have been used for industrial purposes in plants and are expected to supply hydrogen to households and public facilities in the future. This study targets HPL which transport hydrogen at low pressure and are designated to transport hydrogen in urban areas.

One of the challenges for the social implementation of HPL is to ensure their safety (Nakayama, Suzuki, Owada, Shiota, Izato, Noguchi, & Miyake, 2022). Hydrogen leaks from HPL are a challenge because hydrogen has a wide flammable range and low minimum ignition energy. Therefore, if hydrogen leaks from HPL, the hydrogen may be ignited easily by very small energy, such as electrostatic spark discharge, which may lead to serious fire and/or explosion accidents (Imamura, Hamada, Mogi, Wada, Horiguchi, Miyake, & Ogawa, 2008). HPL are expected to be laid in urban areas in the future, and there could be a significant impact on surrounding residents if fires and/or explosions occur. Thus, prevention and mitigation measures for hydrogen leakage are important for the safe operation of HPL.

The detection of hydrogen leakage from an HPL is an important safety measure. This enables the early detection of leakage and prevents accidents. The basic leakage detection method applied in the oil, gas, and water industries involves monitoring the pressure and flow rate values of sensors, which can also be used for HPL. This method is straightforward but sensitive to the instrument and the dynamics of the pipelines (Zhonglin, Li, Shan, Jing, Hao & Tong, 2022). In the future, HPL are expected to supply hydrogen to multiple consumers. Hence, the hydrogen pressure and flow rate behaviors in the HPL may be unstable and change dynamically owing to fluctuations in the hydrogen demand of consumers. Hence, it is difficult to easily distinguish between non-leakage and leakage conditions using the conventional method when the pressure and flow rate behaviors under non-leakage conditions are similar to those under leakage conditions.

To address this challenge, we focused on a leak detection method using machine learning (ML). In recent years, considerable research has been conducted on leak detection methods using ML, some of which have been applied to oil pipelines (Idachaba, & Tomomewo, 2023), and fuel cell vehicles (Mingbin, Teng, Chenhui, Mingjia, Shui, David, & Xuefang, 2021). ML can discover hidden relationships among data that humans cannot discover. The ML method can manage the aforementioned challenges using the relationships established between sensors as monitoring indicators. There are two challenges to applying the ML method to HPL. The first is insufficient operational data for ML. The ML method requires large amounts of operational data. However, operational data have not been accumulated because HPL is still in the social implementation stage, and there are few demonstration cases. Secondly, it is difficult to obtain the pressure and flow rate behaviors during hydrogen leakage because leakage does not occur frequently. We approached these challenges in the following manner. For the former, we acquired operational data via simulation. The simulation was performed using a physical model of the hydrogen pipeline (HPL model). For the latter, an unsupervised ML method was employed. This method can learn using only non-leakage data and does not require leakage data. In other words, we employed an unsupervised ML method based on data simulated using the HPL model to detect hydrogen leakage from the HPL.

This study aimed to construct an ML model for detecting hydrogen leakage from an HPL using physical modeling. First, an HPL model designed to supply hydrogen to multiple consumers was constructed, and an ML model for leak detection was developed based on the data simulated by the HPL model. The leak detection capability of the ML model was verified by comparing the anomaly scores of the non-leakage and leakage conditions. In addition, the ability of the ML model to identify leakage points was verified because it was very important to know where the leakage points were to take appropriate action after the leak was detected. Consequently, we developed a fundamental technology for leak detection based on ML, which can be considered in the design stage of the HPL.

2. METHODOLOGY

2.1. Physical model of hydrogen pipeline (HPL model)

2.1.1. Model construction

The HPL model was defined as follows: A physical model represents the hydrogen transportation processes from the hydrogen production site to the hydrogen consumption site. The HPL model is based on Modelica, an equation-based, object-oriented modeling language that allows the acausal modeling of complex cyber-physical systems (The Modelica Association). Modelica was used to model the complex coupled mechanical, electrical, thermal, and control system phenomena. Modelica is widely applied for multi-physics and system-level modeling language for model-based design and analysis in the hydrogen infrastructure (Suzuki, Kawatsu, Shiota, Izato, Komori, Sato, Takai, Ninomiya, & Miyake, 2021), and aerospace fields (Kawatsu, 2018). This modeling method has a low calculation load and can easily modify physical models. Thus, it can generate large amounts of dynamic data, such as pressure, flow, and temperature behavior. The Modelica-based modeling and simulations in this study were supported by the Modelica-based tool SimulationX (ESI ITI GmbH). The components provided in SimulationX were used for the model. SimulationX included the basic elements of pneumatics and hydrogen power, such as volume, pipelines, fuel cells, electrolyzers, and pressure sensors.

First, an HPL model designed to supply hydrogen to one consumer was constructed using physical equations, such as the ideal gas equation and the first law of thermodynamics. The model was validated by comparing the pressure and flow rate profiles obtained from a demonstration experiment. The input parameters of the model were selected based on the demonstration. By comparing the simulation and experimental data, the pressure and flow rate were validated because the experimental and model simulation values exhibited similar behavior. The HPL is a relatively simple system, and it was determined that it is feasible to extend the HPL model with validated physical equations and construct the HPL model as shown in Figure 1. This model was

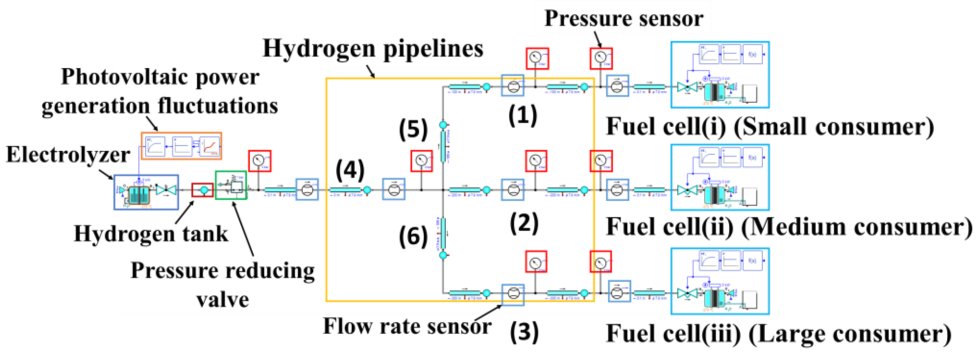

Figure 1. HPL model illustrated using SimulationX

designed to supply hydrogen to multiple consumers and was used to acquire the operational data for ML.

2.1.2. Data acquisition

Pressure and flow rate data were generated through simulations using the HPL model. Table 1 shows the main input parameters of the HPL model. The simulations were performed under the conditions listed in Table 1. A Monte Carlo simulation was used to obtain various simulation data because it requires operational data under many operating conditions. It can be used to obtain simulation data for various temperatures and consumer power demands. The fluctuation parameters are listed in Table 2. We performed 50 calculations for a simulation time of 1000 s. The simulation data were used as training data to construct the ML model. Figure 2 shows a representative example of the pressure and flow rate data.

2.2. ML model construction

The ML model was defined as the relationship established between sensors under non-leakage conditions. The ML model can predict the values of the sensors at a certain time. When the residual between the predicted value from the ML model and the observed value for a sensor exceeds the set threshold value, it can be considered a leakage condition. An ML model was constructed using Invariant analysis technology (NEC). Invariant analysis technology can extract and create invariant relationships between two sensors that represent the characteristics of facilities or systems based on massive quantities of sensor data. This technology is widely applied to many systems such as accelerator systems (Soma, Ishii, Fukuta, & Shiga, 2018).

2.3. Verification

The leakage detection capability of the ML model was verified to ensure that the ML model can distinguish between non-leakage and leakage conditions in certain leakage scenarios. The verification method is as follows.

Table 1. Main input parameters of the HPL model

Parameters	Value	
Hydrogen supply pressure	0.7 - 1.0 MPa	
Diameter of HPL	7.9 mm	
Length of HPL	1,2,3	360 m
	4,5,6	100 m
Rated power of fuel cell	Small consumer	0.7 kW
	Medium consumer	4.0 kW
	Large consumer	20 kW
Initial Atmospheric temperature	283 - 303 K	
Pressure / Flow rate sensor	8 / 8	

Table 2. Parameters given the fluctuations

Parameters	Value		
Power demand fluctuation of consumer	Time interval	Large fluctuation pattern	Small fluctuation pattern
	1 s	-0.01 - +0.01 %	
	10 s	-0.1 - +0.1 %	-1.0 - +1.0 %
	60 s	-0.5 - +0.5 %	-5.0 - +5.0 %
Temperature fluctuation	-0.2 - + 0.2 K		

1. Generating test data

The test data were acquired via simulations using the HPL model. Both non-leakage and leakage condition data were generated. The simulation conditions of the non-leakage

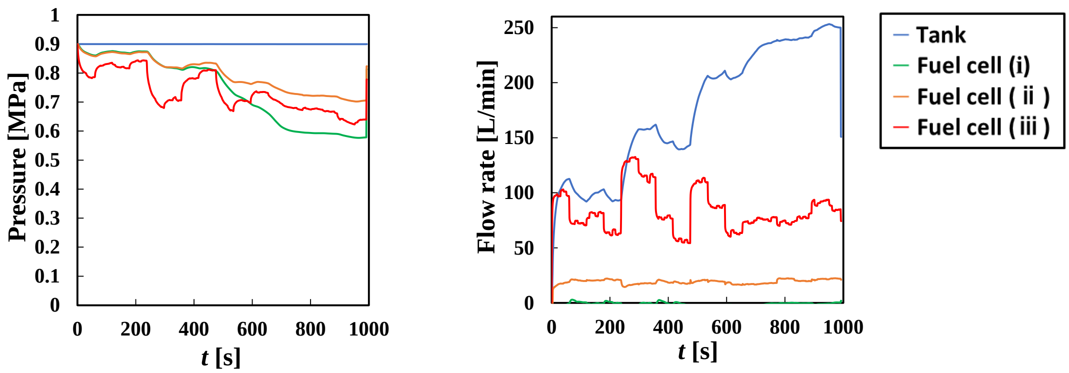

Figure 2. The simulation data acquired by the HPL model (left figure: pressure data, right figure: flow rate data)

condition data were similar to those of the training data, and there were four scenarios with different hydrogen supply pressures.

The leakage phenomenon was represented by the flow rate regulation using orifices. For the leakage condition data, five leak diameters (0.2, 0.4, 0.6, 0.8, 1.0 mm), three leak points (HPL (1), (2), (3)) in Figure 1, and four different hydrogen supply pressures (0.7, 0.8, 0.9, 1.0 MPa) were assumed. Thus, 60 scenarios were considered for the leakage conditions. We performed five calculations for every 64 cases (4 (non-leakage) + 60 (leakage) = 64). In other words, 320 data were generated as test data.

2. Calculating anomaly score

Anomaly score is based on the differences between the values predicted by the ML model and the test data. A higher anomaly score indicates a deviated condition from the non-leakage conditions and is likely to be under leakage conditions. The steps for calculating the anomaly score are as follows. (1) Calculate the residual between the predicted value from the ML model and the test data. (2) If the residual exceeds the set threshold, it was determined that the relationship has collapsed. (3) The fitness score of the collapsed relationship is added to the anomaly score. The fitness score indicates the predictive accuracy of the relational equation; the closer it is to 1, the higher the accuracy is. These steps are performed for all relationships.

3. Comparing anomaly score under non-leakage and leakage conditions

The anomaly scores under non-leakage and leakage conditions were compared. If the distinction between them could be confirmed, it was determined that the ML model could distinguish between non-leakage and leakage conditions.

3. Results and Discussion

3.1. ML model

Table 3 presents examples of the relationships established between the sensors. The equation in Table 3 can predict the pressure of the fuel cell (i) from the pressure of the HPL (1) and the equation. In total, 62 relationships were constructed, and we collectively named them ML models. All these relationships were used for leak detection.

3.2. Result of anomaly score

Figure 3 shows a representative example of an anomaly score. The anomaly score for the non-leakage condition data (blue line) was always zero. This implies that the residual between the predicted value from the ML model and the test data was small, and the deviation from the non-leakage conditions was small. However, the anomaly score for some leakage condition data (red dotted line) increased when leakage occurred (200 s). This means that the residual between the predicted value from the ML model and the test data was large, and the deviation from the non-leakage condition was large. There was a distinction between them; therefore, it was confirmed that the ML model could distinguish between non-leakage and leakage under certain

Table 3. Relationships established between sensors

Input parameter	Predicted parameter	Equation	Fitness score
Pressure of HPL (1)	Pressure of fuel cell(i)	$y(t) = 9.34\times10^{-1}\,x(t)+6.53\times10^{-2}\,x(t-1)+1.39\times10^{-4}$	0.9997
Pressure of HPL (2)	Flow rate of HPL (3)	$y(t) = 9.99\times10^{-1}\,y(t-1)-265\,x(t-2)+265\,x(t-3)-5.06\times10^{-2}$	0.8163

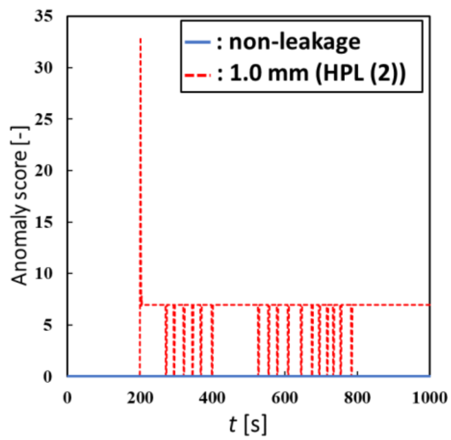

Figure 3. Result of anomaly score

conditions. However, the anomaly score for some leakage condition data did not increase when the leakage occurred. This was seen in the case of small leakage diameters, especially when the leakage diameter was 0.2 mm, which failed for all data. Tables 4, 5, and 6 present the verification results for the leakage condition data. The results were calculated as a percentage of the total number of test data points that were successfully judged. 100% indicated that the judgment was successful for all five sets of test data. It was thought that one of the causes of failure was that the scale of the leakage was too small to be detected using the ML model. Our method determines leakage based on the degree of deviation from non-leakage conditions. Thus, the deviation may be too small to be detected when the scale of the leakage is small. In some failed cases, the residual increased when leakage occurred, but was not sufficient to exceed the set threshold. If the thresholds were set lower, these cases would be detectable; however, this may increase the number of false positives.

As shown in Tables 4-6, as the number of consumers at the supply destination increases, it becomes more difficult to deal with a small leak diameter. This is because the larger the number of consumers at the supply destination, the greater the volume of hydrogen supplied, and the greater the pressure and flow rate fluctuations under non-leakage conditions. This makes it difficult to distinguish the differences in the pressure and flow rate fluctuations under leakage conditions.

Table 4. Result of identification (leak points: HPL (1))

Leak diameter [mm]	Hydrogen supply pressure [MPa]			
	0.7	0.8	0.9	1.0
0.2	0 %	0 %	0 %	0 %
0.4, 0.6, 0.8, 1.0	100 %	100 %	100 %	100 %

Table 5. Result of identification (leak points: HPL (2))

Leak diameter [mm]	Hydrogen supply pressure [MPa]			
	0.7	0.8	0.9	1.0
0.2, 0.4	0 %	0 %	0 %	0 %
0.6	0 %	100 %	100 %	100 %
0.8, 1.0	100 %	100 %	100 %	100 %

Table 6. Result of identification (leak points: HPL (3))

Leak diameter [mm]	Hydrogen supply pressure [MPa]			
	0.7	0.8	0.9	1.0
0.2, 0.4, 0.6	0 %	0 %	0 %	0 %
0.8	0 %	80 %	80 %	100 %
1.0	100 %	100 %	100 %	100 %

3.3. Identification of leakage points

Locating leakage plays one of the most important roles in leakage detection. When the leakage point was HPL (1), the relationship between the pressure of HPL (1) and the pressure of the fuel cell (i) contributed to the leak determination in all test data. The relationship between the pressure of HPL (2) and the pressure of the fuel cell (i), the pressure of HPL (1), and the pressure of HPL (2) also contributed to the leak determination. Because many sensors are located near the leak point, the leak points can be identified based on the collapsed relationships. Similarly, when the leakage point was HPL (2), the relationship between the pressure of HPL (4) and the pressure of HPL (2), the pressure of HPL (1) and the pressure of HPL (2), and the pressure of HPL (2) and the pressure of fuel cell (i) contributed to leak determination in all test data. When the leakage point was HPL (3), the relationships between the pressure of the fuel cell (ⅲ) and the pressure of HPL (3), and the pressure of the tank and the pressure of the fuel cell (i) contributed to leak determination. Similarly, it was also possible to identify the leak point based on the collapsed relationships when the leakage point was HPL (2) and (3).

3.4. Advantages and challenges of the ML model

As mentioned previously, the ML model can distinguish between non-leakage and leakage under certain conditions. In addition, it can identify leakage points based on collapsed relationships. The ability to identify leakage points is an advantage of ML models. It is also possible to determine an optimal sensor solution for leak detection during the process design stage. Leakage can be detected not only using the values of the sensors near the leakage points but also using the values of the sensors far from the leakage points. However, in some cases, the ML model does not function. Thus, it is necessary to consider methods to improve the performance of the ML model. For example, consideration could be given to improving the performance when the number and variation of training data are increased. Additionally, the amount of validation data must be increased. Moreover, it is necessary to confirm in demonstration projects that the proposed method works because this study worked mainly by simulation.

4. Conclusion

In this study, we constructed an ML model to detect hydrogen leakage from an HPL. First, an HPL model (written in the modeling language Modelica) designed to supply hydrogen to multiple consumers was constructed. Physical modeling allows the generation of a large amount of various data and the flexibility to respond to future design changes. The ML model was then constructed based on the data simulated by the HPL model using unsupervised learning algorithms. The leak detection capability of the ML model was verified by comparing the anomaly scores of the non-leakage and leakage conditions. From the verification results, the ML model can distinguish between leakage and non-leakage conditions under certain conditions. In addition, the ML model can identify leakage points based on collapsed relationships. However, the ML model cannot work well in some cases and needs to be improved to deal with more cases.

Acknowledgment

This study was supported by "JPEC Commissioned Basic Research Project for Young Researchers" of the Japan Petroleum Energy Center (JPEC).

References

ESI ITI GmbH, https://www.simulationx.com/.

Faye, O., Szpunar, J., & Eduok, U. (2022). A critical review on the current technologies for the generation, storage, and transportation of hydrogen. *International Journal of Hydrogen Energy*, Vol. 47, pp. 13771-13802. doi: 10.1016/j.ijhydene.2022.02.112

Hanae, D. (2011). Models, methods and approaches for the planning and design of the future hydrogen supply chain. *International Journal of Hydrogen Energy*, Vol. 37, pp. 5318-5327. doi: 10.1016/j.ijhydene.2011.08.041

Idachaba, F., & Tomomewo, O. (2023). Framework for generating pipeline leak datasets using PIPESIM. Journal of Pipeline Science and Engineering, pp. 1-30. doi: 10.1016/j.jpse.2023.100113

Imamura, T., Hamada, S., Mogi, T., Wada, Y., Horiguchi, S., Miyake, A., & Ogawa, T. (2008). Experimental investigation on the thermal properties of hydrogen jet flame and hot currents in the downstream region. *International Journal of Hydrogen Energy*, Vol. 33, pp. 3426-3435. doi: 10.1016/j.ijhydene.2008.03.063

Kawatsu, K. (2018). System-level modeling and simulation-based approach to risk assessment for space systems. *Annual Reliability and Maintainability Symposium* (pp. 1-8), January 22-25. Reno, NV, USA. doi:10.1109/RAM. 2018.8463034

Nakayama, J., Suzuki, T., Owada, S., Izato, Y., Noguchi, K., & Miyake, A. (2022). Qualitative risk analysis of the overhead hydrogen piping at the conceptual process design stage. *International Journal of Hydrogen Energy*, Vol. 47, pp. 11725-11738. doi: 10.1016/j.ijhydene.2022.01.199

NEC, https://www.nec.com/en/global/solutions/ai/analyze/invariant.html

Mingbin, Z., Teng, H., Chenhui, L., Mingjia, C., Shui, J., David, C., & Xuefang, L. (2021). Leak localization using distributed sensors and machine learning for hydrogen releases from a fuel cell vehicle in a parking garage. *International Journal of Hydrogen Energy*, Vol. 46, pp. 1420-1433. doi: 10.1016/j.ijhydene.2020.09.218

Omata, N., Satoh, D., Tsutsusmi, S., Kawatsu, K., & Abe, M. (2022). Model-based supervised sensor placement optimization to detect propellant leak in a liquid rocket engine. *Acta Astronautica*, Vol. 195, pp. 234-242. Doi: 10.1016/j.actaastro.2022.02.009

Rong-Heng, L., Ying-Ying, Z., & Bu-Dan, W. (2020). Toward a hydrogen society: Hydrogen and smart grid integration. *International Journal of Hydrogen Energy*, Vol. 45, pp. 20164-20175. doi: 10.1016/j.ijhydene.2020.01.047

Soma, T., Ishii, K., Fukuta, Y., & Shiga, M. (2018). Applying of system analysis technology (SIAT) to J-PARC accelerator system. *Proceedings of the 15th Annual Meeting of Particle Society of Japan* (pp. 114-118), August 7-10. Nagaoka, Japan

Suzuki, T., Kawatsu, K., Shiota, K., Izato, Y., Komori, M., Sato, K., Takai, Y., Ninomiya, T., & Miyake, A. (2021). Quantitative risk assessment of a hydrogen refueling station using a dynamic physical model based on multi-physics system-level modeling. *International Journal of Hydrogen Energy*, Vol. 46, pp. 38923-38933. doi: 10.1016/j.ijhydene.2021.09.125

The Modelica Association. Modelica Language, https://modelica.org/

Zhonglin, Z., Li, M., Shan, L., Jing, L., Hao, Z., & Tong, L. (2022). A semi-supervised leakage detection method driven by multivariate time series for natural gas gathering pipeline. Process Safety and Environmental Protection, Vol. 164, pp. 468-478. doi: 10.1016/j.psep.2022.06.036