Poultry Science

Multivariate analysis for data mining to characterize poultry house environment in winter

Mingyang Li,* Zilin Zhou,* Qiang Zhang ,† Jie Zhang,* Yunpeng Suo,* Junze Liu,* Dan Shen,* Lu Luo,* Yansen Li,* and Chunmei Li*,[1]

*Research Center for Livestock Environmental Control and Smart Production, College of Animal Science and Technology, Nanjing Agricultural University, Nanjing, Jiangsu Province 210095, China; and †Univ Manitoba, Department of Biosystems Engineering, Winnipeg, MB R3T 5V6, Canada

ABSTRACT The processing and analysis of massive high-dimensional datasets are important issues in precision livestock farming (**PLF**). This study explored the use of multivariate analysis tools to analyze environmental data from multiple sensors located throughout a broiler house. An experiment was conducted to collect a comprehensive set of environmental data including particulate matter (TSP, PM_{10}, and $PM_{2.5}$), ammonia, carbon dioxide, air temperature, relative humidity, and in-cage and aisle wind speeds from 60 locations in a typical commercial broiler house. The dataset was divided into 3 growth phases (wk 1−3, 4−6, and 7−9). Spearman's correlation analysis and principal component analysis (**PCA**) were used to investigate the latent associations between environmental variables resulting in the identification of variables that played important roles in indoor air quality. Three cluster analysis methods; k-means, k-medoids, and fuzzy c-means cluster analysis (**FCM**), were used to group the measured parameters based on their environmental impact in the broiler house. In general, the Spearman and PCA results showed that the in-cage wind speed, aisle wind speed, and relative humidity played critical roles in indoor air quality distribution during broiler rearing. All 3 clustering methods were found to be suitable for grouping data, with FCM outperforming the other 2. Using data clustering, the broiler house spaces were divided into 3, 2, and 2 subspaces (clusters) for wk 1 to 3, 4 to 6, and 7 to 9, respectively. The subspace in the center of the house had a poorer air quality than other subspaces.

Key words: broiler house, microclimate, air quality, multivariate analysis, data mining

2024 Poultry Science 103:103633
https://doi.org/10.1016/j.psj.2024.103633

INTRODUCTION

Monitoring and controlling the indoor environment is critical for intensive poultry production (Ni et al., 2012). Unfavorable indoor environmental conditions can reduce chicken growth and welfare while increase the risk of disease and mortality (Daghir, 2008). The complex structures of multilayer broiler houses may cause uneven environmental conditions, particularly in winter, when ventilation is restricted to reduce the consumption of heating energy.

Multiple parameters influence environmental conditions in broiler houses including temperature, humidity, air movement (speed), harmful gases, and particulate matter (**PM**) (Ni et al., 2021). Air temperature and relative humidity are environmental factors that are significantly related to broiler mortality, because birds do not have sweat glands and are highly sensitive to heat stress (Dawkins et al., 2004; Nawab et al., 2018). Furthermore, inadequate relative humidity can increase the prevalence of infectious diseases by impairing the animal respiratory tract (Xiong et al., 2017). Ventilation is a critical management strategy for ensuring good air quality, production efficiency, and animal well-being (Luck et al., 2017). In chicken houses, harmful gases such as ammonia (**NH_3**) can damage the respiratory tracts of chickens and humans (Portejoie et al., 2002); while high concentrations of CO_2 can affect animal health and welfare in confined buildings. The impact of particulate matter on animals and humans dependent on the particle size, which is commonly measured as total suspended particles (**TSP**) with aerodynamic diameters (**AD**) ≤ 100 μm, inhalable particulate matter (**PM_{10}**) with AD ≤ 10 μm, and fine particulate matter ($PM_{2.5}$) with AD ≤ 2.5 μm (Banhazi et al., 2008; Bonifacio et al., 2015). All 3 PM categories are indicators of air quality in animal buildings.

Received October 26, 2023.
Accepted March 5, 2024.

[1]Corresponding author: chunmeili@njau.edu.cn

Currently, only a few parameters such as temperature, relative humidity, carbon dioxide, or ammonia are monitored in typical poultry barns. Environmental control systems based on a few environmental factors are insufficient for achieving optimal environmental control during broiler rearing (Martinez et al., 2021). Broiler chicken's environmental requirements and management strategies differ depending on their growth periods, and the relationships between environmental variables can also vary (Zhao et al., 2015). With a single input parameter, such as temperature, as in most barns in current practices, environmental control is relatively simple; however, with multiple parameters at multiple locations, sophisticated data analytics are required to develop optimal environmental control strategies (Miles et al., 2008).

One of the challenges in precision livestock farming (**PLF**) is the processing of massive amounts of data from various sensors for environmental control. Therefore, there is a need to develop effective methods for integrating, analyzing, and interpreting data from multiple sensors. Multivariate analysis and clustering are powerful digital technologies for processing large amounts of data from various sensors for the environmental control of PLF. Principal component analysis (**PCA**) and clustering methods have been used to analyze multidimensional data from ecological environments (Kim et al., 2011; Salvati et al., 2016), genetic analysis (Zhang et al., 2016), and microbial identification (Hou et al., 2012; Hug et al., 2015). Principal component analysis and clustering are widely used for data mining to identify groups and classify individuals, as well as machine learning (Thevenot et al., 2013). However, each of these methods has its strengths and limitations, and their suitability for a particular problem is often determined experimentally (Mukhopadhyay et al., 2014).

The goal of this study was to explore various data analytics methods for their suitability and effectiveness in processing and analyzing multivariate heterogeneous environmental data from broiler houses during winter. Specifically, feature selection and principal component analysis were performed to investigate the latent relationships among the environmental parameters, and the key environmental parameters that reflected the indoor air quality distribution were identified. We also used clustering methods to cluster the measurement points throughout the house and evaluated the environmental quality of each cluster. Cluster division evaluation of broiler house indoor air quality may help farmers improving environmental monitoring through strategic selection and sensor placement.

MATERIALS AND METHODS

An experiment was conducted to collect an array of environmental data in a commercial multi-layer broiler barn, including aerosols (TSP, PM_{10}, and $PM_{2.5}$), NH_3, CO_2, wind speed (in-cage and aisle), temperature, and relative humidity. Spearman correlation analysis and PCA were used to investigate the relationships between environmental variables and determine the target variables that best described the environment in the broiler house. Clustering analyses were used to evaluate temporal and spatial variations in the environment within multiple-layer broiler houses.

Broiler House Description

This study was conducted in a typical commercial broiler house located in Jiangsu Province (34°25′ N, 118° 3′ E) from November 19, 2020 to January 20, 2021 with 34,000 one-day-old yellow-feathered broilers. The barn was 90 m long and 16 m wide, and oriented north-south (Figure 1). There were 7 rows of cage stacks, each consisting of 3 tiers with a manure belt. The feed lines were on the sides of each row of cage stack, while the drinker lines were in the center of each row. The chickens were fed a diet formulated to meet NRC (2012) standards for all ingredients and received feed and water ad libitum. Fresh air entered the house through evenly distributed air inlets of 27 cm × 56 cm spaced at 0.64 m along both sidewalls. There were 20 belt-driven exhaust fans of 1.4-m diameter (50″ BOX, Big Herdsman, Qingdao, China) distributed along the south-end wall, while up to 3 fans located in the middle of the wall opened during the study period. The design ventilation rate was approximately 0.15 to 1.85 m^3/h per bird during the study period. A simple environmental control system was installed in the broiler house based on temperature, and a coal-burning boiler system supplied heating energy for 1- to 5-wk-old broilers.

Environmental Monitoring

Due to symmetric housing structures and ventilation operations, only half of the broiler house was included in the environmental measurement process. The interior environment of the broiler houses was measured using multi-attribute sensors at 60 locations (Figure 1). The environmental parameters measured included TSP, PM_{10}, $PM_{2.5}$, NH_3, CO_2, air temperature, relative humidity, and in-cage and aisle air movement (speed). Concentrations of different sizes of PM were measured using a DustTrak II 8533 aerosol monitor (TSI Instruments Co., Ltd., Shoreview, MN). The concentrations of NH_3 and CO_2 were measured using a Korno GT-1000-X5 instrument (Shenzhen Korno Electronic Technology Co., Ltd., Shenzhen, Guangdong, China). The air temperature and relative humidity sensors used were Elitech RC-4HC (Elitech Technology, Co., Ltd., San Jose, CA). The wind speed sensor used for the in-cage and aisle air movement measurements was a Testo 425 (Testo SE & Co. KGaA, Lenzkirch, Germany). All the environmental measurement sensors were checked and calibrated prior to the study. The TSP, PM_{10}, $PM_{2.5}$, NH_3, CO_2, and air movement (speed) in the cage and aisle were measured at 60 points 3 times a week and 3 times a day (8:00 am, 14:00 am, and 20:00 pm). Two people approached the measurement point from both the front and back, collecting data at the same time. To minimize errors, each

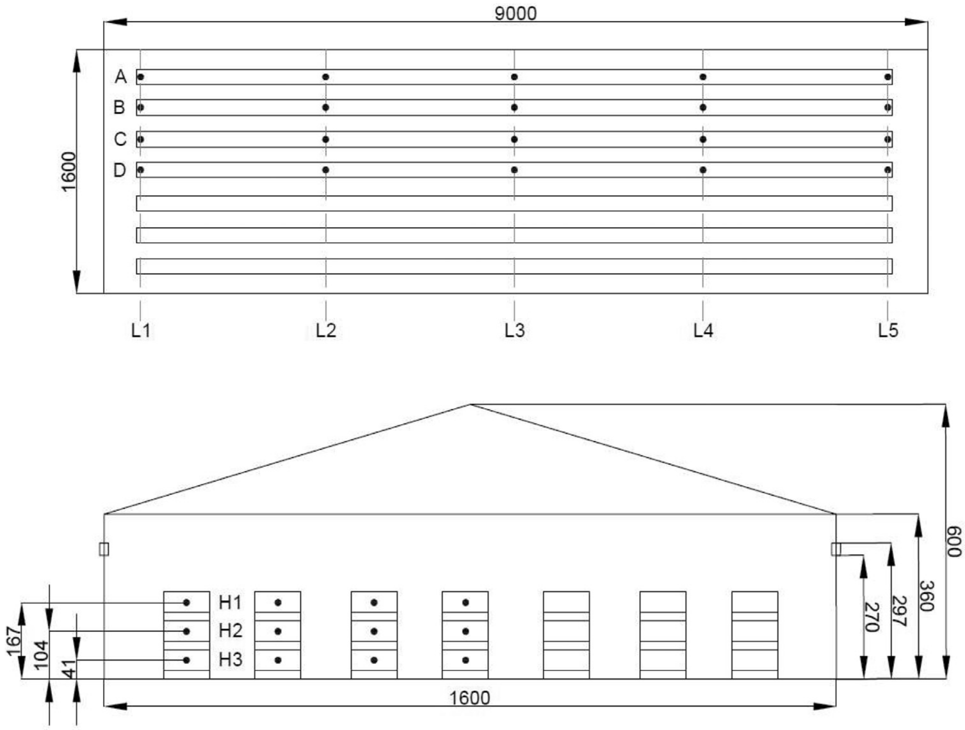

Figure 1. Plan and sectional views of the broiler house with approximate measurement (unit: cm) and 60 measuring locations indicated.

environmental parameter value was collected after one minute of stabilization and repeated 3 times. Temperature and humidity sensors were installed at 60 measuring points to record temperature and relative humidity at 10-min intervals throughout the study period. To accurately reflect the bird's true microclimate, all nine environmental parameters were measured at or near their height. All variables were recorded without interference from commercial broiler production.

Preprocessing of Data and Correlation Analysis

We divided the data into 3 datasets based on the production phases of yellow-feathered broilers: the brooding period (wk 1−3), growing period (wk 4−6), and fattening period (wk 7−9). The means of all measured parameters for each location were calculated for each dataset, which was then subjected to multivariate analysis. A similar approach has been used in previous studies (Hannah et al., 2000; Ouyang et al., 2010).

When 2 features were highly correlated, removing one of them did not significantly affect the variations in all measured parameters in the PCA. Thus, the correlation analysis and PCA may reveal the potential for reducing the number of environmental parameters (sensors) to be monitored in barns. The collected data was analyzed by Spearman's correlation analysis in R with the Hmisc package (Harrell, 2022). Feature selection was performed using the Caret package to determine highly correlated variables (Peng et al., 2005), with a pairwise absolute correlation cutoff of 0.9.

Principal Component Analysis

The relationships between environmental variables were examined using PCA (Wang et al., 2021). In addition, PCA is typically used as a preprocessing tool before cluster analysis (Delaval et al., 2021). PCA was performed by using the R package "FactoMineR" (Le et al., 2008). Principal component loadings were used to determine the relationships between the original variables and principal components, while score calculations were used to display the principal component classification results for each location.

K-Means Cluster Analysis (k-Means)

A useful approach to k-means clustering for determining the optimal number of clusters in data without prior knowledge is to run different simulations with different k values and then use the silhouette method to assess clustering efficiency. Based on previous studies (Javed et al., 2021; Bodereau et al., 2022) and our preliminary observations, we set the test interval for the k values as [2, 8]. This test interval was also used for k-medoids and the fuzzy c-means clustering, as discussed in the following sections. The R statistics package was used for k-means clustering (Grunsky, 2002). The silhouette method is described in the *Cluster validation*.

K-Medoids Cluster Analysis

K-medoids clustering is a robust alternative to k-means clustering (Xu and Wunsch, 2005). Compared with k-means, k-medoids are less sensitive to noise and outliers because they use medoids as cluster centers instead of means. This study adopted the PAM algorithm using the R fpc package, which is the most common k-medoid clustering method (Husson et al., 2017).

Fuzzy c-Mean Cluster Analysis

Fuzzy c-mean clustering is widely regarded as a reliable algorithm in environmental sciences (Hamedian et al., 2016). Fuzzy clustering is considered soft clustering, which differs from k-means and k-medoids, where each object is assigned exactly to one cluster. Fuzzy c-means cluster analysis (**FCM**) involves an additional parameter called the fuzzifier m. The value of m defines the maximum fuzziness or noise in a dataset. To determine the optimal value of m, the fitting equation (Eq. 1) was used to carefully adjust the fuzzifier based on the number of objects and the dimensions of the dataset (Schwammle and Jensen, 2010):

$$m = 1 + \left(\frac{1418}{N} + 22.05\right)^{D^{-2}} + \left(\frac{12.33}{N} + 0.243\right)^{D^{-0.0406\ln(N) - 0.1134}} \quad (1)$$

where N is the number of objects and D is the number of dimensions of an object. The fclust package was used to carry out FCM (Ferraro et al., 2019).

Cluster Validation

Cluster validation is indispensable to avoid finding clusters in random data and to compare different clustering methods. Cluster validation includes clustering tendency assessment, determination of the optimal number of clusters, cluster validation statistics, and selection of the best clustering method. The hopkins package in R was used to calculate the Hopkins index for assessing clustering tendency (Cross and Jain, 1982) and Hopkins values of 0.7 to 1.0 was considered to be acceptable clustering. To determine the optimal number of clusters, the silhouette coefficient method in the NbClust package (Charrad et al., 2014) and the factoextra package were used for both k-means and k-medoids (Kassambara and Mundt, 2020). For the FCM, the fuzzy silhouette index from the fclust package was used (Ferraro et al., 2019). The vegan package was used for the Analysis of Similarities (**ANOSIM**) method to compare within- and between-group similarities (Dixon, 2003). To identify significant differences between the environmental variables for the groups formed by cluster analysis, the Kruskal−Wallis test was performed using the PMCMRplus package (Kruskal and Wallis, 1987). This test identifies the parameters responsible for differentiating the groups formed by clustering. Finally, the best clustering method was selected based on a combination of the above results.

Statistical Analysis

All data were analyzed with the R software 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria). The details and R-packages of the Spearman correlation analysis, PCA, and clustering analyses (k-means, k-medoids, and FCM) are mentioned in the previous sections. Values are expressed as the mean ± standard error of the mean (**SEM**). Significant differences were observed at $p < 0.05$ or $p < 0.01$.

RESULTS

General Observations

The statistical results showed that most environmental variables were significantly different between the 3 growth periods, including TSP, PM_{10}, $PM_{2.5}$, NH_3, CO_2, temperature (**TEM**), and relative humidity (**HUM**) (Table S1). The highest values of NH_3, CO_2, TEM, and HUM were observed in wk 1 to 3 than other weeks ($p < 0.05$), while the concentrations of TSP, PM_{10}, and $PM_{2.5}$ were the lowest ($p < 0.05$); TSP, PM_{10}, and $PM_{2.5}$ were the highest in wk 7 to 9 than other weeks ($p < 0.05$), while the other environmental variables were the lowest ($p < 0.05$).

As shown in Figure 2, the Spearman correlation analysis indicated that there was a positive correlation between TSP, PM_{10}, and $PM_{2.5}$ ($R_{TSP\text{-}PM10} = 0.67$, $R_{TSP\text{-}PM2.5} = 0.59$, $R_{PM10\text{-}PM2.5} = 0.85$, $P < 0.01$) in wk 1 to 3. There was also a positive correlation between NH_3 and CO_2 and between INCAGE and AISLE ($R_{NH3\text{-}CO2} = 0.54$, $R_{INCEGE\text{-}AISLE} = 0.69$, $P < 0.01$). Meanwhile, NH_3 and CO_2 had a negative correlation with TSP, PM_{10}, and $PM_{2.5}$ ($R_{NH3\text{-}TSP} = -0.50$, $R_{NH3\text{-}PM10} = -0.41$, $R_{NH3\text{-}PM2.5} = -0.45$, $R_{CO2\text{-}TSP} = -0.36$, $R_{CO2\text{-}PM10} = -0.58$, $R_{CO2\text{-}PM2.5} = -0.79$, $P < 0.01$), and TEM had a

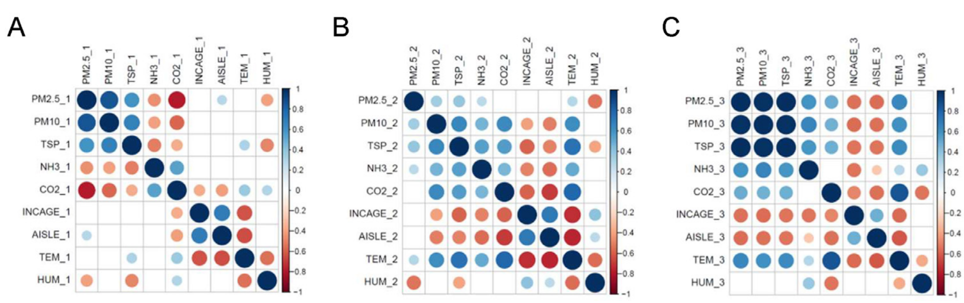

Figure 2. Spearman correlation coefficient results for different production periods. (A) wk 1 to 3; (B) wk 4 to 6; (C) wk 7 to 9.

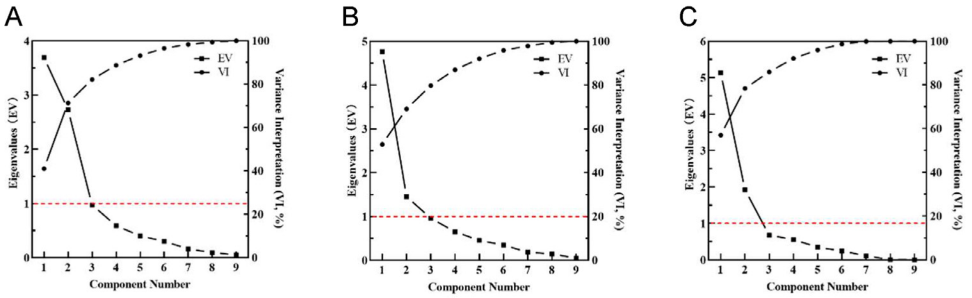

Figure 3. Principal component eigenvalues and variance interpretation percentages for different production periods. (A) wk 1 to 3; (B) wk 4 to 6; (C) wk 7 to 9.

negative correlation with INCAGE, AISLE, and HUM ($R_{TEM\text{-}INCAGE}$ = -0.63, $R_{TEM\text{-}AISLE}$ = -0.64, $R_{TEM\text{-}HUM}$ = -0.54, $P < 0.01$). But in wk 4 to 6, INCAGE and AISLE were negatively correlated to TSP, PM_{10}, NH_3, CO_2, and TEM ($R_{INCAGE\text{-}TSP}$ = -0.59, $R_{INCAGE\text{-}PM10}$ = -0.41, $R_{INCAGE\text{-}NH3}$ = -0.49, $R_{INCAGE\text{-}CO2}$ = -0.59, $R_{INCAGE\text{-}TEM}$ = -0.74, $R_{AISLE\text{-}TSP}$ = -0.50, $R_{AISLE\text{-}PM10}$ = -0.50, $R_{AISLE\text{-}NH3}$ = -0.55, $R_{AISLE\text{-}CO2}$ = -0.70, $R_{AISLE\text{-}TEM}$ = -0.75, $P < 0.01$) and positively correlated to HUM ($R_{INCAGE\text{-}HUM}$ = -0.40, $R_{AISLE\text{-}HUM}$ = -0.26, $P < 0.05$). In wk 7 to 9, most environmental variables (TSP, PM_{10}, $PM_{2.5}$, NH_3, CO_2, and TEM) were positively correlated with each other ($P < 0.01$), with 2 exceptions: most environmental variables have a negative correlation to INCAGE and AISLE ($R_{INCAGE\text{-}TSP}$ = -0.54, $R_{INCAGE\text{-}PM10}$ = -0.56, $R_{INCAGE\text{-}PM2.5}$ = -0.54, $R_{INCAGE\text{-}NH3}$ = -0.53, $R_{INCAGE\text{-}CO2}$ = -0.48, $R_{INCAGE\text{-}TEM}$ = -0.56, $R_{AISLE\text{-}TSP}$ = -0.53, $R_{AISLE\text{-}PM10}$ = -0.54, $R_{AISLE\text{-}TSP}$ = -0.55, $R_{AISLE\text{-}NH3}$ = -0.27, $R_{AISLE\text{-}CO2}$ = -0.55, $R_{AISLE\text{-}TEM}$ = -0.61, $P < 0.01$); and a negative correlation between HUM and CO_2 and TEM ($R_{HUM\text{-}CO2}$ = -0.53, $R_{HUM\text{-}TEM}$ = -0.38, $P < 0.01$). The feature selection results showed absolute correlation values greater than 0.9 between PM_{10} and $PM_{2.5}$ in wk 1 to 3 and between TSP, PM_{10}, and $PM_{2.5}$ in wk 7 to 9.

PCA Results

PCA showed that the first 2 principal components (PC1 and PC2) were above the scree (eigenvalues > 1) and accounted for 71.82, 69.05, and 76.14% of the variation in all measured parameters for wk 1 to 3, 4 to 6, and 7 to 9, respectively (Figure 3). Specifically, in wk 1 to 3 PC1 represented 38.85% of the variance in all measured parameters, with $PM_{2.5}$, INCAGE, and AISLE in the positive direction and CO_2 and NH_3 in the negative direction (Figure 4A); PC2 explained 32.97% of the variance, with HUM in the positive direction and TEM and TSP in the negative direction. In wk 4 to 6, PC1 explained 52.93% of the variance, with TSP, PM_{10}, NH_3, CO_2, and TEM in the positive direction, and INCAGE and AISLE in the negative direction (Figure 4B). PC2 accounted for 16.12% of the variance, with $PM_{2.5}$ in the positive direction of PC2, and HUM in the negative direction. In wk 7 to 9, PC1 explained 51.89% of the variance, with $PM_{2.5}$, TEM, and CO_2 in the positive direction, and INCAGE and AISLE in the negative direction; PC2 explained 24.25% of the variance, with HUM and NH_3 in the positive direction (Figure 4C).

Data Clustering

The Hopkins values for the 3 growth phases were 0.978, 0.997, and 0.968, indicating a clear clustering trend. Specifically, the 60 locations were divided into 3, 2, and 2 clusters for wk 1 to 3, 4 to 6, and 7 to 9, respectively (Figure S1). Furthermore, FCM yields the same clustering results based on the fuzzy silhouette index.

The data clusters obtained using different methods are summarized in Figure 5. In wk 1 to 3 (Figures. 5A, 5D, and 5G), cluster 1 was in the positive direction of PC1, while clusters 2 and 3 were in the negative direction of PC1 and were separated by PC2. In wk 4 to 6, Clusters 1 and 2 were distinguished using PC1. Cluster 1 included INCAGE, AISLE, and HUM (Figures 5B, 5E, and 5H), whereas Cluster 2 contained the other 6 variables (TSP, PM_{10}, $PM_{2.5}$, NH_3, CO_2, and TEM). Vectors in PC2 could explain the degree of overlap between clusters, including $PM_{2.5}$ and HUM. Furthermore, the

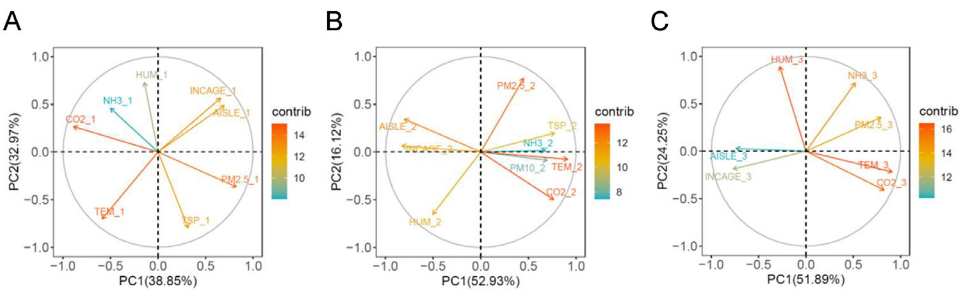

Figure 4. Principal component contribution values (PC 1: x-axis vs. PC 2: y-axis). (A) wk 1 to 3; (B) wk 4 to 6; (C) wk 7 to 9.

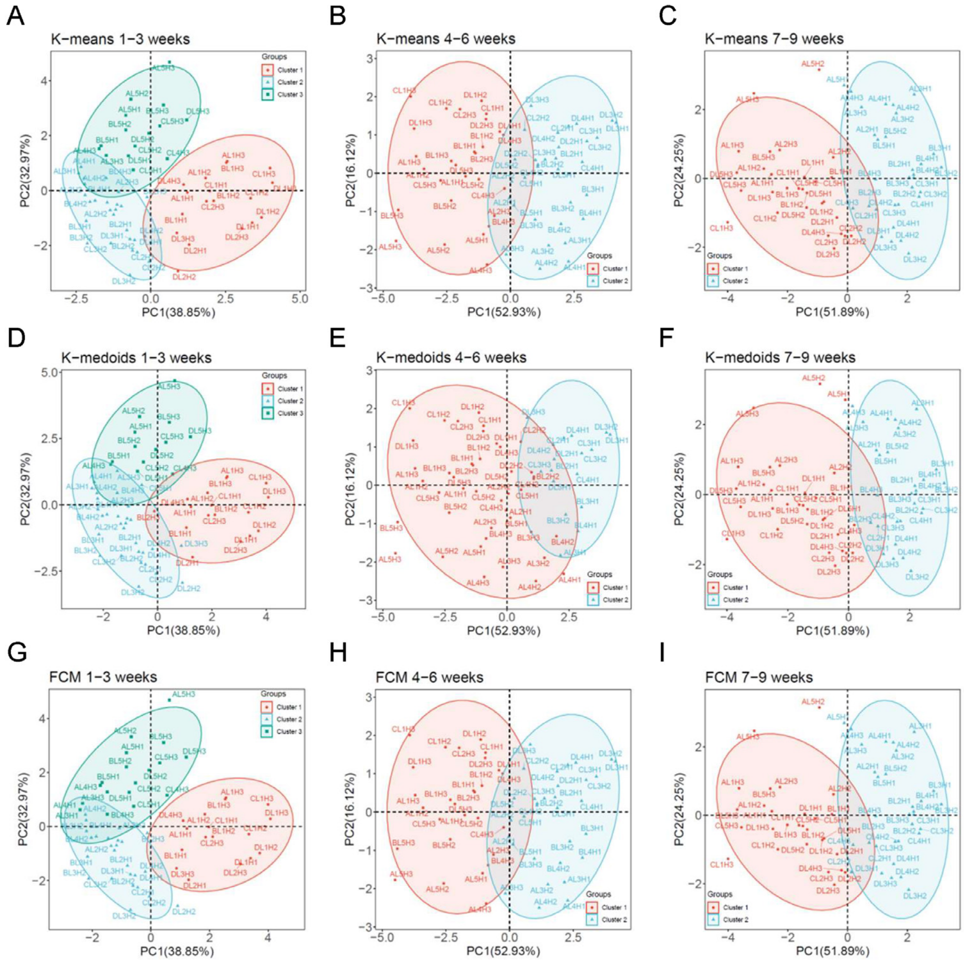

Figure 5. Cluster results of different cluster methods (k-means, k-medoids, and FCM) for 3 production periods.

k-means clustering results were consistent with the FCM results from wk 4 to 6. In wk 7 to 9, Cluster 1 integrated the INCAGE and AISLE vectors of the first principal component, while cluster 2 housed the remaining vectors. The HUM vector was located between Clusters 1 and 2, indicating some degree of overlap between the clusters. Variables on the PC2 axis also showed evidence of overlap.

Cluster Validation

The ANOSIM results of the 3 different cluster methods (k-means, k-medoids, and FCM) showed greater between-group dissimilarity than within-group dissimilarity for all 3 phases (Figure S2). Tables S2, S3, and S4 further confirm that the 3 cluster methods yielded similar results (i.e., the same locations were clustered together by all 3 methods), although slight differences in the cluster means and standard deviations were observed.

The spatial distribution of clusters after applying partition-based clustering methods (k-means, k-medoids, and FCM) revealed that the 60 measurement locations were divided into different clusters, primarily along the longitudinal direction of the broiler house (Fig. 6). In wk 1 to 3, Cluster 1 had the highest aerosol concentration and wind speed among all clusters ($P < 0.05$), Cluster 2 had the highest temperature ($P < 0.05$), and Cluster 3 had the highest NH_3 and CO_2 concentrations and humidity ($P < 0.05$). In wk 4 to 6, Cluster 1 had a higher concentration of aerosols, NH_3, CO_2, and temperature than Cluster 2 ($P < 0.05$) but lower wind speed and humidity than Cluster 2 ($P < 0.05$). The statistical results for wk 7 to 9 were similar to those for wk 4 to 6, except for humidity, which showed no significant differences between the clusters in wk 7 to 9 ($P > 0.05$).

DISCUSSION

Compared with the literature data (Winkel et al., 2015), the PM_{10}, $PM_{2.5}$, and CO_2 concentrations were higher. Several factors could explain this finding. First, as broilers age, the concentration of PM in houses tends to increase logarithmically. The PM concentration of particulate matter in the house was also influenced by broiler feeding activities and farm personnel management. Similar NH_3 concentration trends were observed in 2 large mechanically ventilated layer hen houses equipped with manure belts (Chai et al., 2010). The CO_2 concentrations in this study were higher than those reported in previous studies (Alberdi et al., 2016; Zheng et al., 2020). This inconsistency may be due to low ventilation and the accumulation of CO_2 in broiler houses during winter. In addition, no significant differences

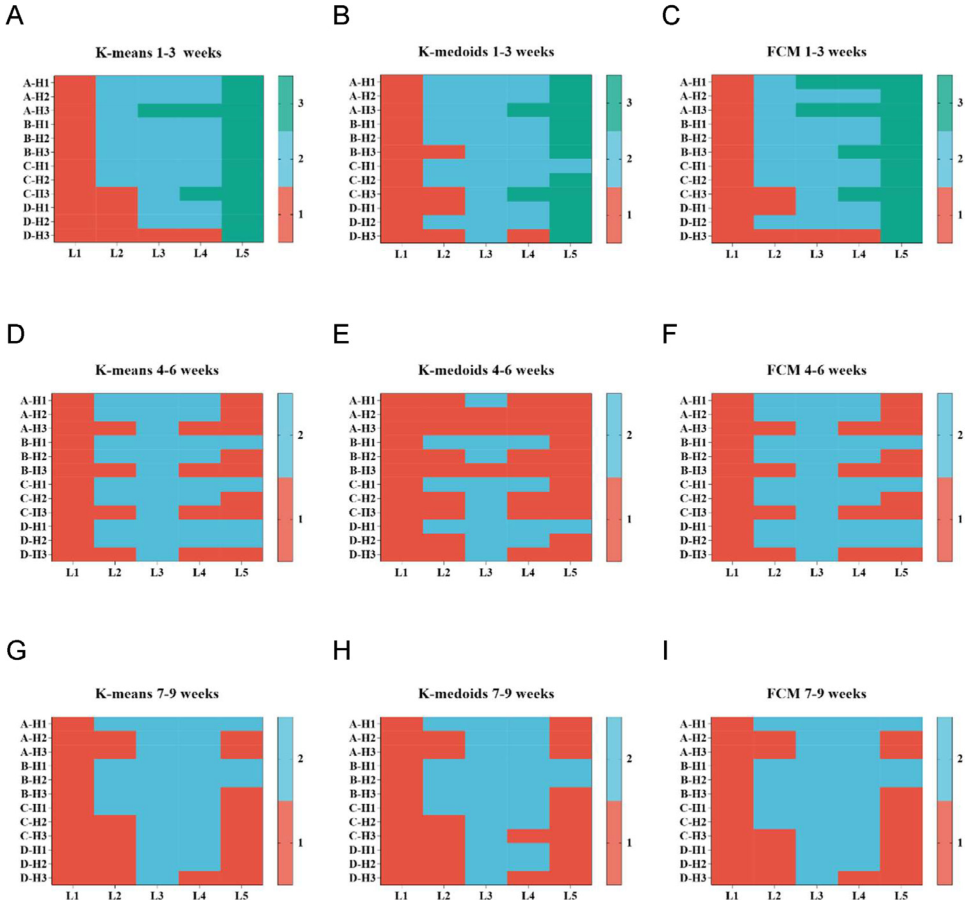

Figure 6. Distribution of clusters obtained from different cluster methods (k -means, k -medoids, and FCM) for 3 production periods.

were observed between the in-cage and aisle airspeeds, with low values in all 3 production phases.

As expected, with an increase in bird body in wk 4 to 6 and 7 to 9, ventilation and wind speed also increased; therefore, higher wind speeds had a greater influence on other environmental variables (INCAGE and AISLE). These results are consistent with previous findings, which demonstrated that the accumulation of hot and harmful air with low ventilation in winter, as well as humidity, is a major factor in environmental control systems in multiple-layer poultry houses (Zajicek and Kic, 2013; Ni et al., 2017). The feature selection results indicated that PM_{10} could be excluded in wk 1 to 3 and TSP and PM_{10} in wk 7 to 9 for subsequent PCA and clustering analyses because their effects could be represented by $PM_{2.5}$ in the original data. The absolute correlation value between the in-cage and aisle wind speed variables was below 0.9, which could be attributed to the different effects of dilution and dissipation on other environmental parameters.

The results of the PCA showed that PC1 and PC2 could adequately represent all measured environmental variables, implying that prediction models, such as machine learning models, can be used in barns with limited sensors/measurements of environmental parameters; that is, variable selection feasible. The directional relationship between the environmental vectors was also consistent with the Spearman's correlation results. The main factors affecting the environmental conditions in broiler houses were identified by examining the combinations of PC1 and PC2. Table 1 shows that the contributions of environmental variables to PC1 and PC2 varied across production periods. The rankings of the contributions were also different. The most important result was that both in-cage and aisle wind speeds were always on the PC1 axis, and their contributions in PC1 were higher than in PC2, whereas HUM was always on the PC2 axis, with a higher contribution in PC2 than in PC1. The stable contributions suggest that INCAGE, AISLE, and HUM are key variables representing the other variables in the PCA results.

The clustering results showed that the variables on the PC1 axis were mainly responsible for cluster division while the variables on the PC2 axis were primarily responsible for cluster overlap. Combined with the PCA results, the in-cage wind speed, aisle wind speed, and relative humidity could be considered the key variables representing the indoor air quality distribution in the broiler house. All 3 variables are closely related to ventilation, which is critical for environmental control and reduction of harmful gases and particulates (Carvalho et al., 2011; Gillespie et al., 2017). An excess or lack of ventilation may interfere with broiler production (Kucuktopcu et al., 2022). The HUM vector was on the PC2 axis throughout all 3 growth periods, indicating that the humidity distribution pattern was different from the other environmental variables. This was consistent with our previous findings, which showed that the cold draft

Table 1. Contributions of environmental variables in PC1 and PC2 for 3 production periods.

Environmental variables	1–3 wk		4–6 wk		7–9 wk		Average contribution	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
TSP	3.23	24.05	12.67	2.70	NA	NA	7.95	13.37
PM_{10}	NA	NA	10.26	0.52	NA	NA	10.26	0.52
$PM_{2.5}$	21.80	5.06	4.26	41.20	17.38	7.57	14.48	17.94
NH_3	8.08	8.02	9.70	0.02	7.50	30.66	8.43	12.90
CO_2	25.52	2.70	12.29	17.22	18.81	9.89	18.88	9.94
INCAGE	13.94	12.11	14.43	0.27	16.29	2.06	14.89	4.81
AISLE	15.63	9.05	13.52	8.17	15.05	0.05	14.74	5.76
TEM	11.14	18.73	17.53	0.40	22.86	2.79	17.17	7.31
HUM	0.66	20.28	5.33	29.49	2.11	46.98	2.70	32.25

NA, not available because of the feature selection; INCAGE, in-cage wind speed; AISLE, aisle wind speed; TEM, temperature; HUM, relative humidity.

generated by the outdoor air created a zone with high relative humidity in the cage rows near the sidewalls. This finding was also reported by Dain et al., 2022.

The ANOSIM results showed that all 3 methods produced consistent results and were suitable for analyzing the environmental data collected for this study. Clusters containing data from the middle section of the house consistently had higher concentrations of harmful gases and higher temperatures, indicating poor environmental conditions. Interestingly, for wk 1 to 3 there were significant differences in PM_{10} levels between various FCM clusters. Given that the clustering results of FCM were identical to those of k-means at wk 4 to 6, FCM seemed more discriminative than k-means or k-medoids. This finding broadly agrees with those of other studies on air pollution area (Yu et al., 2012; Suris et al., 2022). In this study, fuzzifier m was optimized based on a specific dataset. However, when the number of objects and the dimensions were changed in the new dataset, the appropriate value of m was determined again. Furthermore, applying the findings of this study using a different sensor or different broiler houses (e.g., different numbers of air inlets and outlets, and different building dimensions) would help ensure its generalization.

CONCLUSIONS

This study found that multivariate data mining methods can analyze complex datasets describing the environmental conditions in a multi-layer broiler house and extract additional information to facilitate PLF development. Spearman's correlation analysis and PCA indicated that different environmental variables were correlated in a complex manner during different growth periods. Spatial variations in environmental conditions in the broiler house were causally related to in-cage and aisle air movements (speed) and relative humidity. Although this observation was limited to the dataset used in this study and may not be universally applicable to other conditions, it demonstrated that PCA could be used to reduce the dimensions of data containing multiple environmental variables, and the results could be used to optimize the design of environmental monitoring systems with the least number of sensors. Three different clustering algorithms (k-means, k-medoids, and fuzzy c-means clustering) performed well, with fuzzy c-mean being slightly more discriminative. The clustering methods divided the broiler house space into 3, 2, and 2 clusters for wk 1 to 3, 4 to 6, and 7 to 9, respectively. The cluster for data collected in the middle of the house indicated poorer environmental conditions than the other clusters. This study focused on the effects of individual environmental variables on overall environmental quality. Although it is generally agreed that better environmental quality results in better production, the impact of different environmental parameters on production when reducing the dimensions of diverse environmental data should be further studied.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (No. 32372935 and 32072781), Jiangsu Agricultural Industry Technology System (JATS(2023)437) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_0637).

DISCLOSURES

The authors declare no conflicts of interest.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.psj.2024.103633.

REFERENCES

Alberdi, O., H. Arriaga, S. Calvet, F. Estelles, and P. Merino. 2016. Ammonia and greenhouse gas emissions from an enriched cage laying hen facility. Biosyst. Eng 144:1–12.

Banhazi, T. M., J. Seedorf, M. Laffrique, and D. L. Rutley. 2008. Identification of the risk factors for high airborne particle concentrations in broiler buildings using statistical modelling. Biosyst. Eng 101:100–110.

Bodereau, N., A. Delaval, H. Lepage, F. Eyrolle, P. Raimbault, and Y. Copard. 2022. Hydrological classification by clustering approach of time-integrated samples at the outlet of the Rhone river: application to Δ^{14}C-POC. Water Res 220.

Bonifacio, H. F., R. G. Maghirang, S. L. Trabue, L. L. McConnell, J. H. Prueger, and E. R. Bonifacio. 2015. TSP, PM_{10}, and $PM_{2.5}$ emissions from a beef cattle feedlot using the flux-gradient technique. Atmos. Environ 101:49–57.
Carvalho, T. M. R., D. J. Moura, Z. M. Souza, G. S. Souza, and L. G. D. Bueno. 2011. Litter and air quality in different broiler housing conditions. Pesqui Agropecu Bras 46:351–361.
Chai, L. L., J. Q. Ni, Y. Chen, C. A. Diehl, A. J. Heber, and T. T. Lim. 2010. Assessment of long-term gas sampling design at two commercial manure-belt layer barns. J. Air Waste Manage 60:702–710.
Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs. 2014. Nbclust: an R package for determining the relevant number of clusters in a data set. J. Stat. Softw 61:1–36.
Daghir, N. J. 2008. Broiler feeding and management in hot climates. Pages 227−260 in Poultry Production in Hot Climates. N. J. Daghir, ed. CABI Books, Wallingford, UK.
Dain, K., I. B. Lee, S. Y. Lee, S. J. Park, J. G. Kim, J. Cho, H. H. Jeong, and D. Y. Jeong. 2022. 강솔뫼. 2022. Internal thermal environment uniformity analysis of mechanically ventilated broiler house. J. Korean Soc. Agric. Eng 64:65–75.
Dawkins, M. S., C. A. Donnelly, and T. A. Jones. 2004. Chicken welfare is influenced more by housing conditions than by stocking density. Nature 427:342–344.
Delaval, A., C. Duffa, I. Pairaud, and O. Radakovitch. 2021. A fuzzy classification of the hydrodynamic forcings of the Rhone River plume: An application in case of accidental release of radionuclides. Environ. Modell Softw 140:105005.
Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci 14:927–930.
Ferraro, M. B., P. Giordani, and A. Serafini. 2019. fclust: An R package for fuzzy clustering. R. J 11:211–229.
Gillespie, J., R. Nehring, and C. Hallahan. 2017. New versus old broiler housing technology: which leads to greater profit? J. Appl. Poultry. Res 26:72–83.
Grunsky, E. C. 2002. R: A data analysis and statistical programming environment - an emerging tool for the geosciences. Comput. Geosci 28:1219–1222.
Hamedian, A. A., A. Javid, S. M. Zarandi, Y. Rashidi, and M. Majlesi. 2016. Air quality analysis by using fuzzy inference system and fuzzy c-mean clustering in Tehran, Iran from 2009-2013. Iran J. Public Health 45:917–925.
Hannah, D. M., B. P. G. Smith, A. M. Gurnell, and G. R. McGregor. 2000. An approach to hydrograph classification. Hydrol Process 14:317–338.
Cross, G. R., and A. K. Jain. 1982. Measurement of clustering Tendency. Theory and Application of Digital Control. Proceedings of the IFAC Symposium, New Delhi, India.
Harrell, J. F. 2022. Hmisc: Harrell miscellaneous. R package version 4.7-1. Accessed Jan. 2023. https://CRAN.R-project.org/package=Hmisc.
Hou, Y., D. R. Braun, C. R. Michel, J. L. Klassen, N. Adnani, T. P. Wyche, and T. S. Bugni. 2012. Microbial strain prioritization using metabolomics tools for the discovery of natural products. Anal. Biochem 84:4277–4283.
Hug, C., X. Zhang, M. Guan, M. Krauss, R. Bloch, T. Schulze, T. Reinecke, H. Hollert, and W. Brack. 2015. Microbial reporter gene assay as a diagnostic and early warning tool for the detection and characterization of toxic pollution in surface waters. Environ. Toxicol. Chem 34:2523–2532.
Kassambara, A., Mundt, F., 2020. Factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7, Accessed Jan. 2023. https://CRAN.R-project.org/package=factoextra.
Husson, F., S. Le, and J. Pagès. 2017. Exploratory Multivariate Analysis by Example Using R. . 2nd ed CRC Press, New York.
Javed, A., S. D. Hamshaw, B. S. Lee, and D. M. Rizzo. 2021. Multivariate event time series analysis using hydrological and suspended sediment data. J Hydrol 593:125802.
Kim, H. S., J. H. Kim, C. H. Ho, and P. S. Chu. 2011. Pattern classification of typhoon tracks using the fuzzy c-means clustering method. J. Climate 24:488–508.
Kruskal, W. H., and W. A. Wallis. 1987. Citation-classic - use of ranks in one-criterion variance analysis. Curr. Contents/Soc. Behav. Sci 40:20−20.
Kucuktopcu, E., B. Cemek, H. Simsek, and J. Q. Ni. 2022. Computational fluid dynamics modeling of a broiler house microclimate in summer and winter. Animals 12:867.
Le, S., J. Josse, and F. Husson. 2008. FactoMineR: An R package for multivariate analysis. J. Stat. Softw 25:1–18.
Luck, B. D., J. D. Davis, J. L. Purswell, A. S. Kiess, and S. J. Hoff. 2017. Assessing air velocity distribution in three sizes of commercial broiler houses during tunnel ventilation. T. Asabe 60:1313–1323.
Martinez, A. A. G., I. D. Naas, T. M. R. de Carvalho-Curi, J. M. Abe, and N. D. D. Lima. 2021. A heuristic and data mining model for predicting broiler house environment suitability. Animals 11:2780.
Miles, D. M., D. E. Rowe, and P. R. Owens. 2008. Winter broiler litter gases and nitrogen compounds: Temporal and spatial trends. Atmos. Environ 42:3351–3363.
Mukhopadhyay, A., U. Maulik, S. Bandyopadhyay, and C. A. C. Coello. 2014. A survey of multiobjective evolutionary algorithms for data mining: Part I. Ieee. T. Evolut. Comput 18:4–19.
Nawab, A., F. Ibtisham, G. H. Li, B. Kieser, J. Wu, W. C. Liu, Y. Zhao, Y. Nawab, K. Q. Li, M. Xiao, and L. L. An. 2018. Heat stress in poultry production: Mitigation strategies to overcome the future challenges facing the global poultry industry. J. Therm. Biol 78:131–139.
Ni, J. Q., L. Chai, L. Chen, B. W. Bogan, K. Wang, E. L. Cortus, A. J. Heber, T.-T. Lim, and C. A. Diehl. 2012. Characteristics of ammonia, hydrogen sulfide, carbon dioxide, and particulate matter concentrations in high-rise and manure-belt layer hen houses. Atmos. Environ 57:165–174.
Ni, J. Q., C. A. Diehl, L. Chai, Y. Chen, A. J. Heber, T. T. Lim, and B. W. Bogan. 2017. Factors and characteristics of ammonia, hydrogen sulfide, carbon dioxide, and particulate matter emissions from two manure-belt layer hen houses. Atmos. Environ 156:113–124.
Ni, J. Q., M. A. Erasmus, C. C. Croney, C. Li, and Y. Li. 2021. A critical review of advancement in scientific research on food animal welfare-related air pollution. J. Hazard Mater 408:124468.
NRC. 2012. Nutrient Requirements of Swine. 11th ed. The National Academies Press, Washington.
Ouyang, R. L., L. L. Ren, W. M. Cheng, and C. H. Zhou. 2010. Similarity search and pattern discovery in hydrological time series data mining. Hydrol Process 24:1198–1210.
Peng, H. C., F. H. Long, and C. Ding. 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. Ieee T Pattern Anal 27:1226–1238.
Portejoie, S., J. Martinez, and G. Landmann. 2002. Ammonia of farm origin: impact on human and animal health and on the natural habitat. Inra. Prod. Anim 15:151–160.
Salvati, L., C. Kosmas, O. Kairis, C. Karavitis, S. Acikalin, A. Belgacem, A. Sole-Benet, M. Chaker, V. Fassouli, C. Gokceoglu, H. Gungor, R. Hessel, H. Khatteli, A. Kounalaki, A. Laouina, F. Ocakoglu, M. Ouessar, C. Ritsema, M. Sghaier, H. Sonmez, H. Taamallah, L. Tezcan, J. de Vente, C. Kelly, A. Colantoni, and M. Carlucci. 2016. Assessing the effectiveness of sustainable land management policies for combating desertification: a data mining approach. J. Environ. Manage 183:754–762.
Schwammle, V., and O. N. Jensen. 2010. A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. Bioinformatics 26:2841–2848.
Suris, F. N. A., M. A. Abu Bakar, N. M. Ariff, M. S. M. Nadzir, and K. Ibrahim. 2022. Malaysia PM_{10} air quality time series clustering based on dynamic time warping. Atmosphere 13:503.
Thevenot, A., J. Aubin, E. Tillard, and J. Vayssieres. 2013. Accounting for farm diversity in life cycle assessment studies - the case of poultry production in a tropical island. J Clean Prod 57:280–292.
Wang, X., Z. Yang, X. Liu, G. Huang, W. Xiao, and L. Han. 2021. Characteristics and non-parametric multivariate data mining analysis and comparison of extensively diversified animal manure. Waste Biomass Valori 12:2343–2355.
Winkel, A., J. Mosquera, P. W. G. G. Koerkamp, N. W. M. Ogink, and A. J. A. Aarnink. 2015. Emissions of particulate matter from animal houses in the Netherlands. Atmos. Environ 111:202–212.
Xiong, Y., Q. S. Meng, J. Gao, X. F. Tang, and H. F. Zhang. 2017. Effects of relative humidity on animal health and welfare. J Integr Agr 16:1653–1658.

Xu, R., and D. Wunsch. 2005. Survey of clustering algorithms. Ieee T. Neural Networ 16:645–678.

Yu, S. W., Y. M. Wei, J. L. Fan, X. Zhang, and K. Wang. 2012. Exploring the regional characteristics of inter-provincial CO_2 emissions in China: an improved fuzzy clustering analysis based on particle swarm optimization. Appl. Energ 92:552–562.

Zajicek, M., and P. Kic. 2013. CFD analysis of broiler house ventilation patterns with respect to the poultry welfare. Pages 151−156 in 6th International Scientific Conference on Rural Development - Innovations and Sustainability.

Zhang, H., T. Liu, Z. Zhang, S. H. Payne, B. Zhang, J. E. McDermott, J. Y. Zhou, V. A. Petyuk, L. Chen, D. Ray, S. S. Sun, F. Yang, L. J. Chen, J. Wang, P. Shah, S. W. Cha, P. Aiyetan, S. Woo, Y. Tian, M. A. Gritsenko, T. R. Clauss, C. Choi, M. E. Monroe, S. Thomas, S. Nie, C. C. Wu, R. J. Moore, K. H. Yu, D. L. Tabb, D. Fenyo, V. Bafna, Y. Wang, H. Rodriguez, E. S. Boja, T. Hiltke, R. C. Rivers, L. Sokoll, H. Zhu, I. M. Shih, L. Cope, A. Pandey, B. Zhang, M. P. Snyder, D. A. Levine, R. D. Smith, D. W. Chan, K. D. Rodland, and C. Investigators. 2016. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell 166:755–765.

Zhao, Y., T. A. Shepherd, H. Li, and H. Xin. 2015. Environmental assessment of three egg production systems-Part I: monitoring system and indoor air quality. Poult. Sci 94:518–533.

Zheng, W., Y. Xiong, R. S. Gates, Y. Wang, and K. W. Koelkebeck. 2020. Air temperature, carbon dioxide, and ammonia assessment inside a commercial cage layer barn with manure-drying tunnels. Poult. Sci 99:3885–3896.