Habitus Aquatica J, August 2026, 7(2):120–132
DOI: https://doi.org/10.29244/HAJ.7.2.120

E-ISSN: 2721-1525

Habitus Aquatica
Journal of Aquatic Resources and Fisheries Management
Journal Homepage:
http://journal.ipb.ac.id/index.php/habitusaquatica

Correlation between household connections and influent domestic wastewater characteristics: Evidence from a small island WWTP on Tidung Besar Island, Indonesia

Hubungan antara jumlah sambungan rumah tangga dan karakteristik air limbah domestik yang masuk: Bukti dari IPAL di Pulau Tidung Besar, Indonesia

Gatot Prayoga[1,2,3*], Rizki Moch Rijaldi[3], Rizqi Fahma Sidik[3], Hefni Effendi[2], Gatot Yulianto[2]

[1] *Coastal and Marine Resources Management Study Program, Faculty of Fisheries and Marine Sciences, IPB University, Bogor, Indonesia*
[2] *Department of Aquatic Resources Management, Faculty of Fisheries and Marine Sciences, IPB University, Bogor, Indonesia*
[3] *Center for Environmental Research, IPB University, Bogor, Indonesia*

Received 31 March 2026 Accepted 15 July 2026 Published 31 August 2026

ABSTRACT

This study examined whether the number of household connections (SR) is associated with influent domestic wastewater characteristics at decentralized WWTPs on Tidung Besar Island, Indonesia. The dataset covered four service zones (North, West, East, and South) observed annually from 2022 to 2025. SR was treated as the independent variable, while pH, BOD5, COD, TSS, oil and grease, ammonia, surfactant anionic (MBAS), free chlorine, total coliform, fecal coliform, and discharge were treated as dependent variables according to data availability. Although monitoring was conducted at both WWTP inlets and outlets, this study used inlet data only to represent untreated wastewater entering the system and to avoid confounding from treatment processes. Exploratory simple linear regression and Pearson correlation were applied to pooled zone-year observations, supplemented by leave-one-out cross-validation (LOOCV) as a predictive robustness check. Total household connections increased from 729 in 2022 to 831 in 2025. Across the pooled 2022-2025 dataset, the highest positive correlations among the main parameters were observed for BOD5 ($r = 0.362$) and COD ($r = 0.346$), while pH ($r = 0.061$), TSS ($r = -0.135$), oil and grease ($r = 0.216$), and ammonia ($r = -0.083$) showed negligible relationships. Discharge showed a low positive correlation with higher value ($r = 0.446$) but should be interpreted cautiously because of possible flow-meter problems. Parameters introduced or discontinued after regulatory changes were treated as exploratory only. None of the pooled relationships were statistically significant ($p > 0.05$), and LOOCV yielded negative predicted R^2 values, indicating poor predictive stability. Overall, SR alone was not a reliable predictor of influent wastewater characteristics.

Keywords: correlation analysis, decentralized WWTP, household connections, influent domestic wastewater, small island

1. Introduction

Domestic wastewater management remains a major environmental challenge in Indonesia, where household and other domestic-source discharges continue to contribute organic matter, suspended solids, nutrients, oil and grease, and microbial contaminants to receiving waters (Widyarani et al., 2022). At the national level, safe wastewater management coverage remains limited, while decentralized systems continue to play an important role in extending sanitation services in areas where centralized sewerage is not yet feasible (Eales et al., 2013; World Bank, 2013). These challenges are often more pronounced in small-island settings, where land availability is constrained, freshwater resources are limited, settlements are relatively dense, and coastal and marine environments are highly sensitive to pollution (Rouse, 2015).

Within such systems, understanding influent domestic wastewater characteristics is important for service planning and operational

*Corresponding author
E-mail address: gatotprayoga@apps.ipb.ac.id

management. Influent quality represents the untreated wastewater entering the WWTP and therefore provides a more direct indication of source-side wastewater characteristics than outlet data, which are already affected by treatment processes. Previous studies have shown that influent characteristics may vary according to source composition and local wastewater generation patterns (Widyarani et al., 2022). Influent quality may also fluctuate with occupancy patterns and tourism-related dynamics, particularly in areas with seasonal variation in visitor numbers (Rodríguez-Alcántara et al., 2024). Accordingly, in practical planning, it is often assumed that wastewater strength and composition will change in a broadly predictable manner as the number of connected households increases. Nevertheless, this assumption may not always hold, because inflow and infiltration within the collection network can dilute or otherwise alter wastewater quantity and quality before it reaches the treatment plant (Bugajski et al., 2025).

Tidung Besar Island provides a relevant case for examining this issue because domestic wastewater on the island is managed through a decentralized system consisting of four WWTP service zones (North, West, East, and South) and the number of household connections (SR) has increased over time. A recent multi-aspect assessment reported that Tidung Besar already has domestic wastewater infrastructure, but the system still faces environmental and operational challenges, including parameter exceedances, uneven service pressure, and equipment-related constraints (Prayoga et al., 2025). This makes Tidung Besar an appropriate setting for evaluating whether service coverage expansion is accompanied by measurable changes in the characteristics of wastewater entering the treatment system. In addition, although its practical relevance, published work has focused more on domestic wastewater generation, characteristics, treatment systems, and system management than on statistically testing whether household connections (SR) are associated with monitored influent parameters (Eales et al., 2013; Widyarani et al., 2022). Accordingly, this study aims to evaluate the correlation between household connections (SR) and influent domestic wastewater characteristics at the WWTP inlets on Tidung Besar Island using time series data from 2022 to 2025. Furthermore, the study assesses whether SR can be interpreted not only as a service coverage indicator but also as a potentially credible standalone predictor of influent domestic wastewater characteristics. In doing so, this paper contributes a more cautious and evidence-based assessment of the opportunities and limitations of using SR-based relationships in decentralized small-island WWTP systems.

2. Methods

2.1. Study context and data structure

This study used a processed dataset for Tidung Besar Island, Indonesia, covering four decentralized WWTP service zones: North, West, East, and South of the Tidung Besar Island. The analytical unit was a zone-year observation, meaning that each observation represented one service zone in one year. Household connections (SR) were recorded for each zone and year, while wastewater quality variables were taken from inlet monitoring data for the corresponding WWTP. Because the objective of this study was to assess the relationship between SR and wastewater characteristics before treatment, only inlet data were used in the statistical analysis, even though monitoring was conducted at both the inlet and outlet.

For parameters monitored continuously from 2022 to 2025, the pooled dataset contained 16 observations (4 zones × 4 years). However, not all parameters were available for the full study period. Some variables were monitored only up to 2024 or only in 2025 because of changes in wastewater quality standard following regulatory updates (Regulation of the Minister of Environment and Forestry No P.68 Year 2016 on Domestic Wastewater Quality Standard, 2016; Regulation of the Minister of Environment/Environmental Management Agency No. 11 of 2025 on Wastewater Quality

Standards and Wastewater Treatment Technology Standards for Domestic Wastewater, 2025).

2.2. Variables and parameter grouping

Household connections (SR) were treated as the independent variable (x). The dependent variables (y) were pH, BOD_5, COD, TSS, oil and grease, ammonia (NH_3), surfactant anionic (MBAS), free chlorine (Cl_2), total coliform, fecal coliform, and discharge, subject to data availability. The values analyzed in this study were annual zone-level values compiled by the author from monitoring records provided by the Seribu Islands Water Resources Sub-Agency. Accordingly, the statistical analysis was performed on the processed annual dataset.

To support interpretation, the parameters were grouped into three analytical categories. Parameters with complete observations for 2022-2025 were treated as the main analysis set. Parameters available only for 2022-2024 were treated as secondary analysis, while parameters available only in 2025 were treated as exploratory only, because their time series were too limited for robust interpretation.

2.3. Statistical Analysis

The main statistical approach combined simple linear regression and Pearson correlation analysis to examine the relationship between household connection and each influent wastewater parameter (James et al., 2021; Schober et al., 2018). For each parameter, the following linear model was fitted to **Equation 1**.

$$y = a + bx \qquad (1)$$

where y is wastewater parameter value, x is number of households connection (SR), a is intercept, and b is slope.

Pearson's correlation coefficient (r) was used to describe the direction and strength of the linear relationship between SR and each parameter (Schober et al., 2018). The coefficient of determination (R^2) was used to indicate the proportion of variance in the dependent variable explained by SR under the fitted linear model (James et al., 2021). Because this was a simple bivariate exploratory study, statistical significance was evaluated using a nominal threshold of $p < 0.05$.

For interpretive consistency, the strength of correlation was classified following (Hinkle et al., 2003) in (Mukaka, 2012) as presented in Table 1. These analyses were applied to the pooled 2022-2025 zone-year dataset for each parameter according to its available observation period.

2.4. Validation and robustness checks

Because the study also generated regression equations, a predictive robustness check was added using leave-one-out cross-validation (LOOCV), a resampling-based approach widely used to assess out-of-sample predictive performance (James et al., 2021; Picard & Cook, 1984; Stone, 1974). In LOOCV, one observation is omitted at a time, the model is re-fitted to the remaining observations, and the omitted observation is then predicted. This procedure was repeated until each observation had served once as the validation case. LOOCV was used here as a secondary robustness check, rather than as the primary basis for inference, to assess whether the fitted equations showed any practical predictive stability beyond the calibration sample.

Table 1. Classification and interpretation of correlation coefficients.

Size of Correlation	Interpretation
0.90 to 1.00 (-0.90 to -1.00)	Very high positive (negative) correlation
0.70 to 0.90 (-0.70 to -0.90)	High positive (negative) correlation
0.50 to 0.70 (-0.50 to -0.70)	Moderate positive (negative) correlation
0.30 to 0.50 (-0.30 to -0.50)	Low positive (negative) correlation
0.00 to 0.30 (0.00 to -0.30)	Negligible correlation

For each omitted observation i, the cross-validated prediction error was calculated with Equation 2.

$$e_i = y_i - \hat{y}_{(-i)} \quad (2)$$

Where e_i is the cross-validated prediction error for observation i, y_i is the observed value for observation i, and $\hat{y}_{(-i)}$ is the value predicted from the model fitted without observation i. Based on these LOOCV prediction errors, the cross-validated root mean square error (LOOCV RMSE) and cross-validated mean absolute error (LOOCV MAE) were calculated using Equation 3 and Equation 4.

$$LOOCV\ RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_{(-i)}\right)^2} \quad (3)$$

$$LOOCV\ MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_{(-i)}\right| \quad (4)$$

In addition, predicted R^2, here denoted as Q^2, was calculated to evaluate cross-validated goodness-of-prediction (Equation 5).

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{(-i)}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

where $\bar{y}$ is the mean of the observed values in the full sample. Positive predicted R^2 (Q^2) values indicate some predictive utility beyond the calibration sample, whereas negative predicted R^2 (Q^2) values indicate poor predictive stability.

For comparison with the cross-validated metrics, the in-sample RMSE of the fitted model was also calculated using Equation 6.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (6)$$

where $\hat{y}_i$ is the fitted value from the regression model estimated using the full dataset.

In addition, year-specific Pearson correlations were calculated separately for 2022, 2023, 2024, and 2025 to examine whether pooled relationships were temporally consistent. Because each year contained only four zone-level observations, these annual correlations were treated as descriptive checks only and were not emphasized as stand-alone inferential results

2.5. Data-quality constraints and interpretation rules

This study was intended as an exploratory analysis of association rather than a test of causality. Three limitations should therefore be considered. First, the same four service zones were observed repeatedly from year to year, so the pooled dataset does not consist of fully independent observations. Second, some parameters were not available for the full 2022-2025 period because monitoring requirements changed during the study period following regulatory updates. In addition,

Table 2. Domestic wastewater parameter availability and analytical status.

Parameter	Available years	n	Analytical status
pH	2022-2025	16	Main analysis
BOD5	2022-2025	16	
COD	2022-2025	16	
TSS	2022-2025	16	
Oil and Grease	2022-2025	16	
Ammonia (NH_3)	2022-2025	16	
Total Coliform	2022-2024	12	Secondary analysis; 2025 unavailable and field note indicates possible flow-meter issues
Discharge	2022-2024	12	
Surfactant Anionic (MBAS)	2025 only	4	Exploratory only
Free Chlorine (Cl_2)	2025 only	4	
Fecal Coliform	2025 only	4	

missing values were not imputed, and each parameter was analyzed only on the basis of the available observations. Third, the discharge variable was treated with special caution because field notes indicated possible flow-meter problems and uncertainty in some discharge records. For this reason, discharge was retained only as a secondary variable and interpreted conservatively.

Accordingly, parameters with full 2022-2025 coverage were treated as the main analytical set, parameters available only through 2024 were treated as secondary, and parameters available only in 2025 were treated as exploratory only (Table 2).

3. Result and Discussion

3.1. Trend of household connections

Household connections (SR) increased in all four WWTP service zones during the study period. The total number of connections rose from 729 in 2022 to 763 in 2023, 804 in 2024, and 831 in 2025 (Figure 1), representing an overall increase of approximately 14.0% over four years. In absolute terms, the West Zone consistently carried the largest service load, increasing from 407 to 449 connections, while the South Zone remained the smallest, increasing from 20 to 23 connections. In relative terms, the North Zone showed the fastest growth, increasing from 122 to 155 connections (about 27.0%), followed by the South, East, and West zones. Overall, these figures indicate a pattern of gradual service expansion rather than a major step-change in connected households.

3.2. Pattern of influent wastewater parameters

Despite the steady increase in SR, the influent wastewater parameters did not change monotonically over time. Several parameters fluctuated substantially across years even when SR increased only modestly. As shown in Table 3, the annual mean BOD_5 decreased from 142.9 mg/L in 2022 to 126.1 mg/L in 2023, then increased to 203.9 mg/L in 2024 and remained high in 2025 (196.5 mg/L). A similar pattern was observed for COD, which declined slightly from 188.0 mg/L in 2022 to 167.8 mg/L in 2023, increased to 270.8 mg/L in 2024, and rose sharply to 393.8 mg/L in 2025. By contrast, ammonia dropped markedly in 2025 relative to the earlier years, while pH shifted considerably between 2023 and 2024. These descriptive patterns suggest that SR alone may not fully explain year-to-year variation in influent concentration.

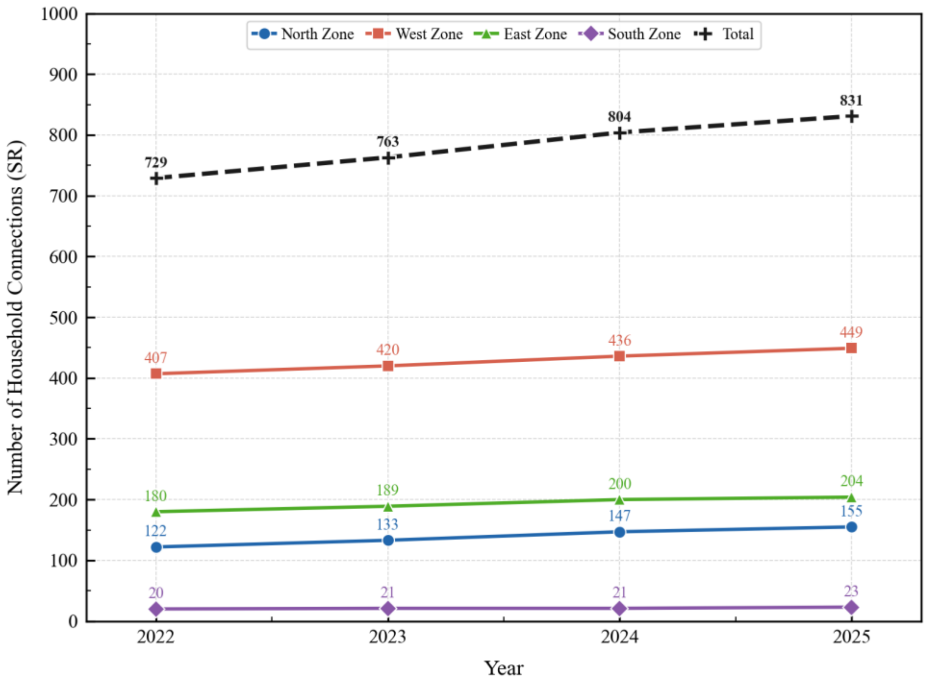

Figure 1. Number of household connections (SR) by WWTP service zone and year.

Table 3. Annual mean influent wastewater parameters across the four service zones in 2022-2025.

Year	Parameter and Unit										
	pH	BOD_5	COD	TSS	Oil and Grease	Ammonia (NH_3)	Surfactant Anionic (MBAS)	Free Chlorine (Cl_2)	Total Coliform	Fecal Coliform	Discharge
	-	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	MPN/100 mL	MPN/100 mL	m^3/s
2022	7.4	142.9	188.0	18.3	6.0	11.4	-	-	220,358	-	0.0003
2023	6.3	126.1	167.8	22.5	8.3	26.3	-	-	149,500	-	0.0011
2024	7.6	203.9	270.8	19.0	8.3	17.5	-	-	149,500	-	0.0005
2025	7.1	196.5	393.8	37.8	4.3	0.6	2.3	0.03	-	73,125	-
QS*	6-9	30	100	30	5	10	5**	1**	3,000	1,000**	-

*Primary and currently applicable reference/quality standard (QS): Regulation of the Minister of Environment/Environmental Management Agency No. 11 of 2025 (App. I.A.2 Wastewater Quality Standards for Non-Toilet Wastewater, or a Mixture of Toilet Wastewater and Non-Toilet Wastewater Discharged into Water Bodies, with a discharge volume of x > 50 m³/day) (Regulation of the Minister of Environment/Environmental Management Agency No. 11 of 2025 on Wastewater Quality Standards and Wastewater Treatment Technology Standards for Domestic Wastewater, 2025)

**Older references/quality standard (QS), used for parameters missing in the primary reference/QS: Regulation of the Minister of Environment and Forestry No P.68 Year 2016 on Domestic Wastewater Quality Standard, 2016 (Regulation of the Minister of Environment and Forestry No P.68 Year 2016 on Domestic Wastewater Quality Standard, 2016)

These interannual changes are interpreted as descriptive evidence of influent variability rather than as direct effects of SR expansion. No specific management intervention was documented during the period that could directly explain the marked COD increase or the ammonia decrease in 2025, and the laboratory reports did not indicate a change in analytical methods for these parameters. Therefore, these changes are better interpreted as year-to-year variability in influent wastewater characteristics that cannot be explained by SR alone. Possible unmeasured contributors include variation in wastewater source composition, water-use behavior, sampling timing, tourism-related occupancy, and hydraulic conditions in the collection system. Such factors have been reported to influence domestic wastewater characteristics and influent quality in decentralized or tourism-affected wastewater systems (Bugajski et al., 2025; Rodríguez-Alcántara et al., 2024; Sururi et al., 2023; Watanabe et al., 2021; Widyarani et al., 2022).

3.3. Pooled regression and correlation results

The pooled regression and correlation results are presented in detail on Table 4, while selected scatter plots are shown in Figure 2. Among the parameters with complete 2022-2025 coverage, the highest positive correlations with SR were observed for BOD5 ($r = 0.362$; $R^2 = 0.131$) and COD ($r = 0.346$; $R^2 = 0.120$). Under the correlation classification following (Hinkle et al., 2003) in (Mukaka, 2012) adopted in this study, both relationships fall within the low positive range. However, neither relationship was statistically significant ($p > 0.05$). The remaining main parameters showed even weaker relationships; oil and grease ($r = 0.216$), pH ($r = 0.061$), TSS ($r = -0.135$), and ammonia ($r = -0.083$) all fall within the negligible range. Thus, for the main analytical set, SR explained only a small proportion of the variation in influent wastewater characteristics.

Among the secondary parameters, discharge showed the highest pooled correlation ($r = 0.446$; $R^2 = 0.199$), which corresponds to a low positive relationship.

Table 4. Pooled regression equations and correlation statistics.

Parameter	n	Equation (y = a + bX)	r	R^2	p-value[a]	Strength	Status
BOD_5	16	y = 135.363 + 0.163635x	0.362	0.131	0.168	Low	
COD	16	y = 197.502 + 0.294520x	0.346	0.120	0.189	Low	
Oil and Grease	16	y = 6.133 + 0.002838x	0.216	0.047	0.422	Negligible	Main
TSS	16	y = 25.940 − 0.008008x	-0.135	0.018	0.619	Negligible	
pH	16	y = 7.036 + 0.000215x	0.061	0.004	0.823	Negligible	
Ammonia (NH_3)	16	y = 15.069 − 0.005887x	-0.083	0.007	0.759	Negligible	
Discharge	12	y = 0.000 + 0.000001x	0.446	0.199	0.146	Low	Secondary
Total Coliform	12	y = 145837.484 + 142.587191x	0.225	0.051	0.481	Negligible	
Fecal Coliform	4	y = 47717.974 + 122.296156x	0.523	0.273	0.477	Moderate	
Free Chlorine (Cl_2)	4	y = 0.048 − 0.000062x	-0.627	0.393	0.373	Moderate	Exploratory
Surfactant Anionic (MBAS)	4	y = 2.653 − 0.001942x	-0.692	0.478	0.308	Moderate	

[a]*Statistical significance was evaluated at $p < 0.05$.*

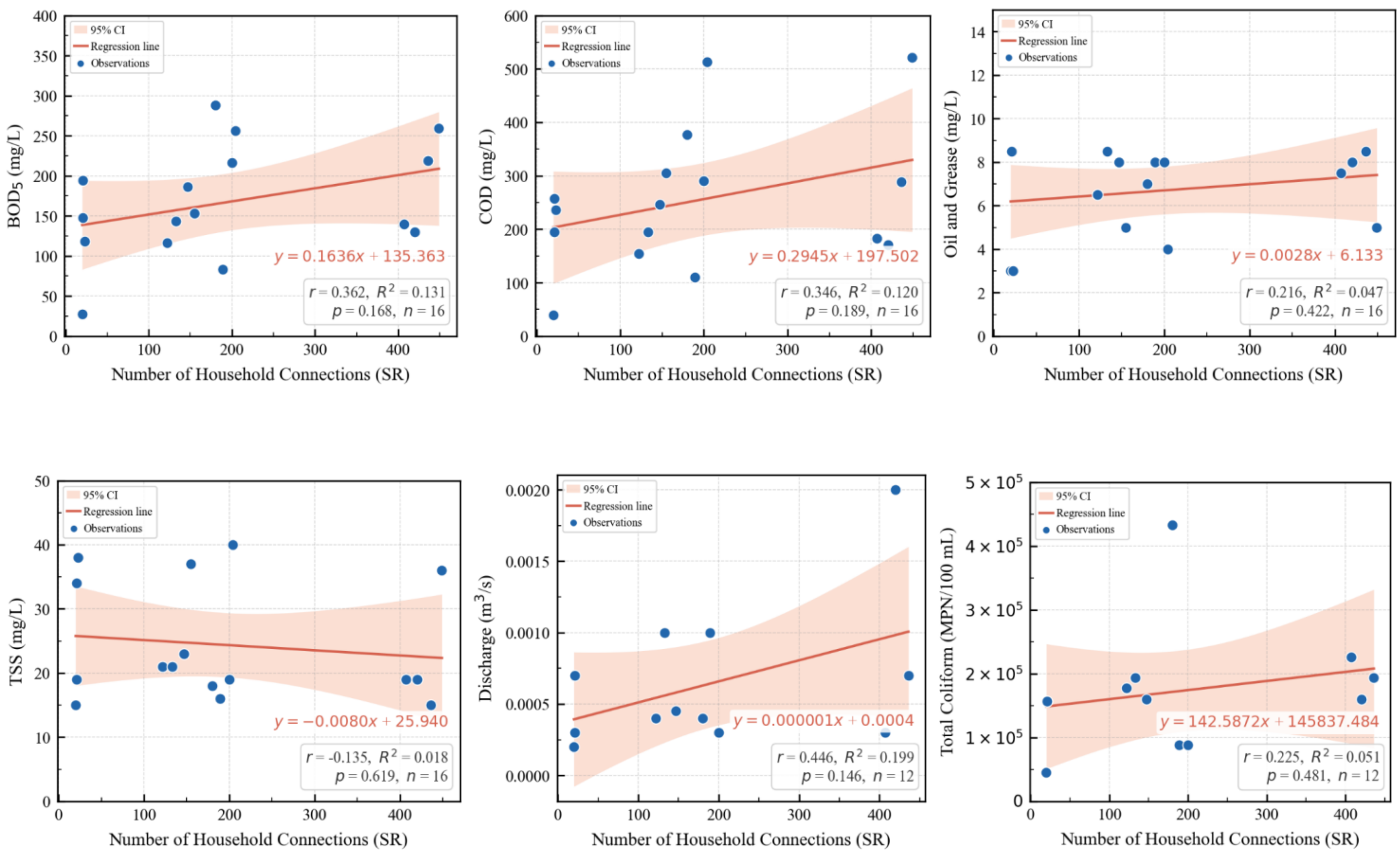

Figure 2. Scatter plot of the number of household connections (SR) against several influent wastewater parameters.

However, this result should be interpreted cautiously because the discharge series is incomplete, especially in 2025, and accompanied by field notes indicating possible flow-meter problems. Total coliform showed only a negligible positive relationship ($r = 0.225$; $R^2 = 0.051$).

The exploratory parameters measured only in 2025 yielded apparently stronger coefficients, including MBAS ($r = -0.692$), free chlorine ($r = -0.627$), and fecal coliform ($r = 0.523$) (Table 4). Under the adopted classification, these represent moderate relationships. However, because these estimates are based on only four observations, they should be regarded as preliminary signals rather than stable or transferable equations.

Overall, the pooled regression results indicate that household connections (SR) were not a strong standalone predictor of influent domestic wastewater characteristics. This is reflected in the generally low to negligible correlation coefficients, small R^2 values, and the absence of statistically significant relationships across the main analytical parameters.

3.4. Model validation and temporal consistency

The LOOCV results confirm that the regression equations presented in Table 4 have limited predictive value when SR is used as the sole predictor. Across parameters, the cross-validated prediction errors were generally larger than the corresponding in-sample errors, as shown by the higher LOOCV RMSE compared with the in-sample RMSE (Table 5). This pattern indicates that the fitted equations performed less well when each observation was predicted outside the calibration subset.

All models produced negative predicted R^2 (Q^2) values (Table 5), meaning that cross-validated predictions performed worse than a simple mean-based benchmark. Therefore, the fitted equations should be interpreted primarily as descriptive relationships rather than predictively robust models. The poorest predictive performance was observed among the exploratory parameters, particularly surfactant anionic/MBAS ($Q^2 = -4.143$) and fecal coliform ($Q^2 = -3.184$), which is consistent with their very small sample size. However, these results should be interpreted cautiously because these parameters were based on only four observations, making the validation metrics highly sensitive to individual data points.

The annual correlations also varied substantially in both direction and magnitude. For example, BOD_5 and COD were positively correlated with SR in 2022, became negative in 2023, and turned positive again in 2024-2025. Similar reversals were observed for oil and grease, while pH shifted from negative in the earlier years to positive in the later years. These reversals suggest that the pooled equations are not temporally stable and should not be interpreted as evidence of a single persistent SR-parameter relationship. The dynamics of year-specific correlation coefficients for each parameter are presented in Figure 3.

Table 5. Validation metrics for pooled regression models.

Parameter	n	LOOCV RMSE	LOOCV MAE	Predicted R^2 (Q^2)	RMSE (in-sample RMSE)	Status
BOD_5	16	70.986	58.833	-0.122	62.469	Main
COD	16	137.706	109.976	-0.191	118.369	
Oil and Grease	16	2.212	1.886	-0.284	1.906	
TSS	16	10.123	8.959	-0.318	8.736	
pH	16	0.615	0.534	-0.374	0.524	
Ammonia (NH_3)	16	11.743	9.082	-0.251	10.465	
Discharge	12	0.001	0.000	-0.396	0.000	Secondary
Total Coliform	12	100956.849	65936.751	-0.191	90135.951	
Fecal Coliform	4	73810.696	65152.762	-3.184	30765.072	Exploratory
Free Chlorine (Cl_2)	4	0.017	0.015	-0.240	0.012	
Surfactant Anionic (MBAS)	4	0.982	0.853	-4.143	0.313	

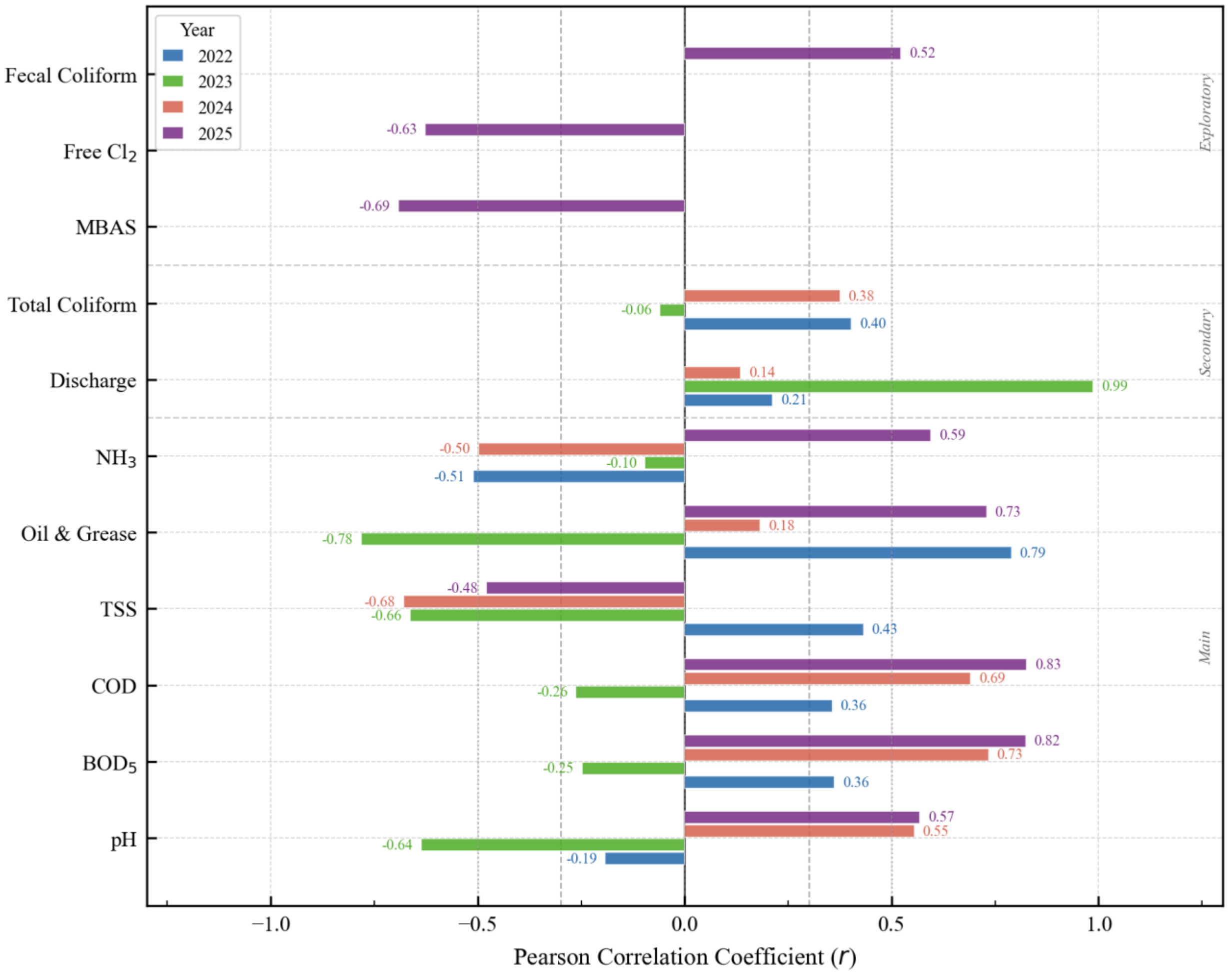

Figure 3. Annual Pearson correlation coefficients (r).

3.5. Discussion

The principal finding of this study is that household connections (SR) were not a strong standalone predictor of influent domestic wastewater characteristics at the WWTP inlets on Tidung Besar Island. This conclusion is supported by three consistent signals from the dataset: most pooled correlations were negligible to low, all pooled relationships were statistically non-significant, and all LOOCV models yielded negative predicted R^2 values. Taken together, these results indicate that SR captures only a limited portion of the variation in influent quality and should therefore be interpreted as a service coverage indicator with only modest explanatory power, rather than as a robust predictive variable.

The substantial interannual variability observed for COD and ammonia further supports this interpretation. Although SR increased gradually during 2022-2025, COD increased sharply in 2025, whereas ammonia decreased markedly in the same year. This opposite direction of change suggests that influent wastewater characteristics were not controlled simply by the number of connected households. COD reflects oxygen demand from organic and other oxidizable substances in wastewater, whereas ammonia is more closely associated with the nitrogen fraction of domestic sewage and may also be affected by dilution, transformation, and sampling conditions (Macura et al., 2024; Office of Water US EPA, 1978; Sururi et al., 2023; Watanabe et al., 2021). Based on the available records, no specific management intervention or change in laboratory analytical method was identified during the period. Therefore, the COD and ammonia patterns observed in this study should not be interpreted as evidence of a causal SR effect, but rather as an indication that source composition and operational conditions varied across years.

This weak relationship is technically plausible. Domestic wastewater characteristics are influenced not only by the number of connected households but also by water-use behavior, the relative contribution of greywater and blackwater, and the composition of activities contributing to the sanitary flow. In Indonesia, greywater volumes have been reported to be one to four times higher than blackwater volumes, and untreated greywater remains a major contributor to domestic wastewater pollution, which means that changes in water use and source composition can alter wastewater strength even without a proportional change in household numbers (Widyarani et al., 2022). Greywater generation and quality are also known to vary with lifestyle, fixtures, household practices, and climatic conditions (Oteng-Peprah et al., 2018).

At the sewer and collection level, additional variability can arise before wastewater even reaches the WWTP. An Indonesian study on a cluster-scale decentralized domestic wastewater system found that wastewater characteristics changed along the sewer line and differed between peak and non-peak periods, indicating that conveyance conditions and sampling timing can influence inlet measurements (Sururi et al., 2023). Infiltration and inflow can further alter both flow and concentration. A study showed that infiltration substantially increased sewage flow and decreased sewage concentration in a residential sewer catchment (Watanabe et al., 2021), while likewise reported that incidental stormwater inflow affected both the volume and quality of influent wastewater (Bugajski et al., 2025).

These mechanisms are especially relevant to the Tidung Besar case. Tidung Besar is a densely populated, tourism-based small island with four decentralized WWTP service zones and documented operational challenges, including uneven service pressure, parameter exceedances, and equipment-related issues (Prayoga et al., 2025). In such a setting, variation in influent quality may reasonably reflect heterogeneity in domestic wastewater sources across households, schools, government offices, dormitories, and small- to medium-scale accommodation facilities, rather than households alone. This interpretation is also consistent with island-based evidence showing that tourism pressure and seasonal population dynamics can influence wastewater quality in WWTP systems (Rodríguez-Alcántara et al., 2024).

Local studies on Tidung Island further support the relevance of tourism-related pressure in interpreting wastewater variability. Tourism development in Tidung has been associated with the growth of amenities such as homestay accommodation, boat rental, and catering businesses, as well as environmental impacts including waste accumulation, changes in water quality, and coral reef disturbance (Khrisnamurti et al., 2016). Hayati et al. (2020) also reported that tourism was one of the largest waste-generating sources on the Tidung Islands, indicating that visitor-related activities can materially affect local environmental pressure. From a broader small-island tourism perspective, (Adrianto et al., 2021) emphasized that carrying capacity is an essential consideration for reducing tourism impacts in Tidung Islands, while Arijuddin (2024) highlighted the importance of balancing domestic and tourism water demand in the same island context. These studies do not directly quantify influent wastewater chemistry, but they provide a contextual basis for considering tourism intensity, occupancy, water use, and facility composition as plausible explanatory variables beyond SR.

The secondary and exploratory parameters further show how monitoring continuity affects inferential strength. Discharge had the highest pooled correlation in the study, but confidence in that result is weakened by the documented uncertainty in the flow-meter record. Similarly, MBAS, free chlorine, and fecal coliform produced moderate correlations, but these were based on only four observations from 2025 and therefore cannot be treated as stable relationships. Total coliform, in contrast, had a longer series but still showed only a negligible pooled relationship. Thus, part of the statistical weakness observed in this study reflects not only environmental complexity but also uneven temporal coverage across parameters following regulatory changes in monitoring requirements.

From a practical perspective, the regression equations derived here are best interpreted as empirical source-characterization relationships rather than as design equations or decision rules. For decentralized small-island systems, relying on SR alone could oversimplify actual influent loading conditions and lead to misleading expectations regarding concentration trends. Given the weak SR-based relationships observed in this study, the contrasting interannual patterns of influent parameters, and the documented tourism-related environmental pressures on Tidung Island, future analyses should be strengthened by including additional predictors such as occupancy, tourism intensity, per-capita water use, rainfall, infiltration/inflow indicators, and facility-type composition within each service zone. With the current annual dataset, the most defensible conclusion is that SR provides only limited directional information and does not yet function as a robust standalone predictor of influent domestic wastewater characteristics.

3. Conclusion

This study evaluated whether household connections (SR) can explain influent domestic wastewater characteristics in four decentralized WWTP service zones on Tidung Besar Island during 2022-2025. Although SR increased gradually across the study period, the influent parameters did not show a consistent or proportional response. The pooled correlation and regression results indicate that SR had only weak explanatory power, while the LOOCV results and year-specific correlations further showed limited predictive stability and temporal consistency.

The substantial interannual variation observed for COD and ammonia reinforces the conclusion that influent wastewater quality is shaped by broader source-side and system-level variability rather than by household connection numbers alone. Therefore, the regression equations derived in this study should be interpreted as empirical descriptive relationships, not as robust predictive or design equations. Future assessments in decentralized small-island WWTP systems should combine SR with additional explanatory variables, such as occupancy, tourism intensity, per-capita water use, rainfall, infiltration/inflow, and the composition of household and non-household facilities within each service zone. Operationally, influent wastewater quality should be interpreted together with broader source and system characteristics, rather than relying on SR alone.

Acknowledgment

The author would like to express sincere gratitude to collaboration agency of AIS Forum, UNDP, and Department of Aquatic Resources Management, IPB University which provided full financial support for this study through the AIS Innovator Scholarship Program. The author also extends appreciation to all parties who contributed to the research activities, including stakeholders (Seribu Islands Water Resources Sub-Agency, Tidung Island Village office and residents), and research team colleagues (Center for Environmental Research, IPB University).

Reference

Adrianto, L., Kurniawan, F., Romadhon, A., Bengen, D. G., Sjafrie, N. D. M., Damar, A., & Kleinertz, S. (2021). Assessing social-ecological system carrying capacity for urban small island tourism: The case of Tidung Islands, Jakarta Capital Province, Indonesia. *Ocean & Coastal Management*, *212*, 105844. https://doi.org/10.1016/J.OCECOAMAN.2021.105844

Arijuddin, B. I. (2024). Water Carrying Capacity on Small Islands Tourism: Case Study of the Tidung Islands, Jakarta. *Jurnal Syntax Admiration*, *5*(6), 2350–2358. https://doi.org/10.46799/JSA.V5I6.1214

Bugajski, P., Mucha, Z., Kois, B., Zakład, M., Agów I Kanalizacji, W. ˛, & Targu, N. (2025). The Influence of Rainwater and

Snowmelt (Inflow and Infiltration) on the Performance of Wastewater Treatment in a Plant Using Membrane Bioreactors (MBR). *Applied Sciences 2025, Vol. 15, Page 10608*, *15*(19), 10608. https://doi.org/10.3390/app151910608

Eales, K., Blackett, I., Siregar, R., & Febriani, E. (2013). Review of Community-Managed Decentralized Wastewater Treatment Systems in Indonesia. www.worldbank.org

Hayati, Y., Adrianto, L., Krisanti, M., Pranowo, W. S., & Kurniawan, F. (2020). Magnitudes and tourist perception of marine debris on small tourism island: Assessment of Tidung Island, Jakarta, Indonesia. *Marine Pollution Bulletin*, *158*. https://doi.org/10.1016/j.marpolbul.2020.111393

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). Houghton Mifflin. https://archive.org/details/appliedstatistic0000hink_o7h5/page/n7/mode/1up

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R* (2nd ed.). Springer US. https://doi.org/10.1007/978-1-0716-1418-1

Khrisnamurti, Utami, H., & Darmawan, R. (2016). THE IMPACTS OF TOURISM ACTIVITIES ON THE ENVIRONMENT IN TIDUNG ISLAND, KEPULAUAN SERIBU. http://jejakwisata.

Macura, B., Metson, G. S., McConville, J. R., & Harder, R. (2024). Recovery of plant nutrients from human excreta and domestic wastewater for reuse in agriculture: a systematic map and evidence platform. *Environmental Evidence 2024 13:1*, *13*(1), 21-. https://doi.org/10.1186/S13750-024-00342-5

Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*, *24*(3), 69. https://pmc.ncbi.nlm.nih.gov/articles/PMC3576830/

Office of Water US EPA. (1978). *Method 410.3: Chemical Oxygen Demand (Titrimetric, High Level for Saline Waters) by Titration*. https://www.epa.gov/sites/default/files/2015-08/documents/method_410-3_1978.pdf

Oteng-Peprah, M., Acheampong, M. A., & deVries, N. K. (2018). Greywater Characteristics, Treatment Systems, Reuse Strategies and User Perception—a Review. *Water, Air, & Soil Pollution 2018 229:8*, *229*(8), 255-. https://doi.org/10.1007/s11270-018-3909-8

Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, *79*(387), 575–583. https://doi.org/10.1080/01621459.1984.10478083

Prayoga, G., Yulianto, G., & Effendi, H. (2025). Assessing wastewater treatment performance in small island contexts: a multi-aspect evaluation of Tidung Besar Island, Indonesia. *Jurnal Pengelolaan Lingkungan Berkelanjutan (Journal of Environmental Sustainability Management)*, 265–289. https://doi.org/10.36813/jplb.9.3.265-289

Regulation of the Minister of Environment and Forestry No P.68 Year 2016 on Domestic Wastewater Quality Standard (2016).

Regulation of the Minister of Environment/Environmental Management Agency No. 11 of 2025 on Wastewater Quality Standards and Wastewater Treatment Technology Standards for Domestic Wastewater (2025).

Rodríguez-Alcántara, J. S., Cruz-Pérez, N., Rodríguez-Martín, J., García-Gil, A., & Santamarta, J. C. (2024). Effect of tourist activity on wastewater quality in selected wastewater treatment plants in the Balearic Islands (Spain). *Environmental Science and Pollution Research*, *31*(10), 15172–15185.

https://doi.org/10.1007/S11356-024-32173-9/FIGURES/10

Rouse, J. D. (2015). Development of Environmentally Sustainable Methods for Treatment of Domestic Wastewater and Handling of Sewage Sludge on Yap Island. *Sustainability 2015, Vol. 7, Pages 12452-12464*, *7*(9), 12452–12464. https://doi.org/10.3390/su70912452

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(2), 111–147. https://sites.stat.washington.edu/courses/stat527/s13/readings/Stone1974.pdf?utm_source=chatgpt.com

Sururi, M. R., Dirgawati, M., Wiliana, W., Fadlurrohman, F., Hardika, & Widiyati, N. (2023). Performance evaluation of domestic waste water treatment system in urban Indonesia. *Case Studies in Chemical and Environmental Engineering*, *8*(2), 100507. https://doi.org/10.1016/j.cscee.2023.100507

Watanabe, R., Harada, H., Yasui, H., Le, T. Van, & Fujii, S. (2021). Exfiltration and infiltration effect on sewage flow and quality: a case study of Hue, Vietnam. *Environmental Technology (United Kingdom)*, *42*(11), 1747–1757. https://doi.org/10.1080/09593330.2019.1680739

Widyarani, Wulan, D. R., Hamidah, U., Komarulzaman, A., Rosmalina, R. T., & Sintawardani, N. (2022). Domestic wastewater in Indonesia: generation, characteristics and treatment. *Environmental Science and Pollution Research*, *29*(22), 32397–32414. https://doi.org/10.1007/s11356-022-19057-6

World Bank. (2013). East Asia Pasific Region Urban Sanitation Review : Indonesia Country Study. In *Urban Sanitation Review : Indonesia Country Study*. Washington, DC. https://doi.org/https://doi.org/10.1596/17614