



Analysing the influence of climatic variables on air conditioning condensate water production

Danilo Carvalho¹ · Daniel Sant'Ana¹ · Arthur Tavares Schleicher² · Júlia Garcia Ribeiro³

Received: 11 July 2025 / Accepted: 11 May 2026
© The Author(s) 2026

Abstract

Air conditioning condensate recovery systems have potential applicability in buildings that make intensive use of air conditioning. However, little is known about the volume of condensate water generated by air conditioning systems for use in non-potable applications. This study aims to elucidate the influence of climatic variables on condensate water production. Data collection regarding the condensate water production of air conditioning system was carried out in a 15 m², totalling 968 h of operation during the rainy season in Brasilia, Brazil. Overall, relative air humidity and dew point showed positive correlations with the volume of condensate water, indicating that more humid conditions favour condensation. Conversely, higher temperatures, both external and internal, correlated negatively with the volume of condensate water. In this way, a generalized linear model was proposed, adjusted using the explanatory variables of external temperature and relative air humidity to predict the volume of condensate water produced. The model was not able to capture the variability of the response variable well. However, this study represents an advance regarding the development of a model capable of estimating the supply of air conditioning condensate water.

Keywords Air conditioning · Condensed water · Clear water · Forecasting model · Recovery

Introduction

Sustainable water resource management in urban environments is a growing challenge. Extreme climatic conditions and the increase in urban demand for potable water can result in a new context of water crisis in Brazil (Sant'Ana and Lima 2021). The supply of water in urban areas, which represents approximately 25% of the total water consumed in the country, has increased the pressure on the systems responsible for water production (ANA 2020). Consequently, about 43 million people were affected by extreme

drought or dry spells in 2018 (ANA 2019). The availability of water resources, especially in areas with hot and dry climates, is becoming increasingly problematic (Congedo et al. 2021). Ensuring a safe water supply for the expanding global population is a challenge that becomes even more complex in areas facing water scarcity, characterized by hot climates and an increasing per capita water demand (Luqman and Al-Ansari 2021).

In this context, there is a need to implement actions aimed at water conservation, such as conscious use and the incorporation of alternative water sources in buildings. Non-potable building water systems are considered a viable option to significantly reduce the demand for potable water in urban areas. In recent decades, there has been an increase in the installation of non-potable water systems in Brazil, such as rainwater harvesting and greywater reuse. The use of alternative water sources in buildings is an effective strategy to reduce potable water consumption, ensure a continuous supply, and minimize the impacts of water resource exploitation (Sant'Ana 2017).

Even though condensate water generated during the cooling and dehumidification process of air conditioning systems is usually discarded as waste, it could be considered

Communicated by Francesca Alessandro.

✉ Daniel Sant'Ana
dsantana@unb.br

¹ Water and Built Environment Research Laboratory, University of Brasilia, Brasilia, Federal District, Brazil

² Hydrology Laboratory, Faculty of Technology, University of Brasilia, Brasilia, Federal District, Brazil

³ Department of Statistics, University of Brasilia, Brasilia, Federal District, Brazil

as an alternative source of water for non-potable uses in buildings. The use of air conditioning has increased significantly in recent years, especially in industries such as food processing, healthcare facilities, retail and commercial buildings, providing support for essential tasks such as data management, food preservation, and creating a comfortable working environment (Dhamodharan et al. 2024a). Buildings that contain high demand for cooling could benefit from the use of condensate water produced by air-conditioning systems (Cichoń and Worek 2021). The search for sustainable resources drives research and development of efficient condensate water recovery systems in buildings and built environments. Such systems, when planned in the design phase, allow for the collection, storage, and utilization of condensate water in an optimized way (Algarni et al. 2018).

Condensate water production is significantly affected by humidity and temperature, which vary according to geographic location, in addition to the air management mode in the environments (fresh or recirculated) as well as the use and occupation of buildings, cooling capacity, operating hours, and internal climatic data of the built environment (Argani et al. 2018; Dhamodharan et al. 2024b; Zolfagharkhani et al. 2018; Eades 2018; Poredos et al. 2021). It is important to note that large air conditioning systems have a greater potential to produce condensate water compared to smaller equipment, such as split systems. Theoretical predictions of the volume of condensate generated often exceed experimental results due to ideal assumptions in the analyses, which highlights the need for adjustments in these approaches to increase accuracy (Argani et al. 2018).

In a hotel located in Villahermosa, Mexico, a system was installed to extract water from the air using a reverse compression cycle, which is used in heat pump systems. This system not only treated the water to make it suitable for human consumption but also provided water heating to the existing heating plant, as well as to the hotel's kitchen and laundry. The test period, which lasted approximately one month, confirmed that the water quality was in compliance with local and international standards (Cattani et al. 2018).

Another study analysed the daily production of condensed water generated by the operation of 113 air conditioning units at the Faculty of Engineering, Horus University, located in New Damietta, a coastal city in northern Egypt. The data revealed that during the peak summer months (June, July, and August), the cumulative daily production was 51.24 L, 51.56 L, and 55.65 L, respectively. In the autumn months (September, October, and November), the recorded values were 53.32 L, 44.13 L, and 44.53 L, respectively (Elbrashy et al. 2025).

Cattani et al. (2023) carried out a case study on a large-scale atmospheric water generator to extract water from humid ambient air in a workers' village in Dubai, a region

where the supply of potable water is a constant concern, measured a condensate water production that exceeded 1000 L/day. According to a recent review, atmospheric water generators are classified into low, medium, and high-capacity units, with production rates ranging from 20 L/day to 10,000 L/day (Beithou et al. 2024).

Chantasiriwan (2024) evaluated an air-cooling system used as an atmospheric water generator in a combined cycle power plant. This system generates saturated air by mixing ambient air with water and uses a mechanical cooler to reduce both the temperature and humidity of the air, resulting in the conversion of air humidity into water. The simulation results showed that, under standard conditions of 15 °C and 60% relative air humidity, 0.216 kg/s of condensate water was produced by the air-cooling system. Therefore, the system functioned effectively as an atmospheric water generator in humid climates.

Magrini et al. (2017) analysed an integrated air conditioning system that combines the function of air conditioning with water production in a hotel was discussed. For this purpose, the climatic conditions of the coast of the United Arab Emirates were considered for the case study analysis. The findings suggest that, under the evaluated climatic conditions, such an air conditioning system can produce a significant amount of condensate water, sufficient to meet a relevant portion of the hotel's water demands. A recent investigation has shown that condensate water production ranged between 23 L/h and 35 L/h with verified temperatures of 15 °C and 10 °C, in equipment with cooling capacities of 115 kW and 175 kW (Dhamodharan et al. 2024a, b).

It is observed that previous studies report that the integration of air conditioning technologies that produce condensate water has significant potential to supplement local water supplies. Even though different studies have explored condensate water production, there is still a limited body of research focusing on the quantitative modelling of condensate water production, under real operating conditions, particularly for small-scale air-conditioning systems. This limitation constrains the development of reliable predictive models and highlights the need for empirical studies that address this gap.

This study aims to analyse the influence of climatic variables on the production of condensate water in air conditioning systems, taking into account the local climate, which includes external temperature and relative air humidity, in addition to the conditions of thermal comfort in internal environments, such as temperature and relative air humidity.

Materials and methods

For the case study environment, a faculty room located in the University of Brasilia, Brazil, was selected. The room, approximately 15 m², was occupied by up to two people

and climatized by a 9000 BTU/h Hi-Wall split-type air conditioner. This typology has representative potential for a variety of built environments, such as other faculty rooms, offices, apartments, and residential environments. In addition to ease of access, the University Campus has significant potential for condensate water production, considering the number of building users, climatized areas, and building usage periods.

For the conception of the experimental setup, the external climatic environmental conditions and internal conditions of the selected built environment were defined as explanatory variables. The data on the volume of condensate water produced by the air conditioning equipment each hour and selected as the response variable. Data collection took place in the months of November 2023 and January 2024, with the air conditioner configured to climatize the environment at a temperature of 22 °C, obtaining 968 h of operation of the experimental setup.

An experimental apparatus was developed with a focus on fundamental aspects such as high precision, data integrity, and process automation, enabling data collection over extended periods and, consequently, a greater variability of data. The system was designed to record, on an hourly basis, climatic data from both the external and internal environments of the building, as well as the volume of condensed water produced by the air conditioning system. To achieve this, electronic components such as Arduino®, transducers, electronic modules, and a thermohygrometer data logger were utilized.

After data extraction, a procedure of organization and tabulation was carried out, resulting in the following tested variables:

- $\Delta Volume$, delta volume, dependent variable corresponding to the volume of condensed water supply per hour (L/h);
- T_{ext} , external temperature, explanatory variable corresponding to the dry-bulb temperature measured in the external environment of the building (°C);
- RH_{ext} , external relative air humidity, explanatory variable corresponding to the relative humidity measured in the external environment of the building (%);
- T_{int} , internal temperature, explanatory variable corresponding to the dry-bulb temperature measured in the air-conditioned environment (°C);
- RH_{int} , internal relative air humidity, explanatory variable corresponding to the relative humidity measured in the air-conditioned environment (%); and
- P_{int} , internal dew point, explanatory variable corresponding to the dew point temperature measured in the air-conditioned environment (°C).

Subsequently, a descriptive statistical analysis of the results was conducted, exploring the collected data to verify normality regarding aspects of central tendency, dispersion, distribution, and results of the normality test of the variables. All analyses were performed using R software, version 4.3.0. The use of this statistical environment enabled the construction of a descriptive statistics table, allowing for a detailed examination of each variable, including measures of central tendency (median, mean), dispersion and distribution (standard deviation, skewness).

Furthermore, a preliminary analysis of the boxplot diagram allowed for the identification of outliers, especially in the variables $\Delta Volume$ and external relative air humidity. The presence of negative values of $\Delta Volume$ was verified, possibly due to subtle disturbances on the water surface or even resulting from the precision level of the electronic components. As these negative values do not reflect the reality of the experiment, and considering the tightness of the experimental setup, it was decided to assign the value zero to these occurrences. Regarding external relative air humidity, values higher than 100% were recorded, possibly due to the margin of error of the measuring device. Given that these values also do not correspond to reality, they were replaced by the value of 100%. After the adjustments were made, the data were subjected to a new descriptive analysis of the results separately, exploring the collected data to verify aspects of central tendency, dispersion, distribution, and results of the normality test of the variables.

When performing the normality analysis using the Shapiro-Wilk test, strong indications of non-conformity with the normal distribution were found for all analysed variables ($p\text{-value} < 0.001$). Given the lack of normality of the variables, it was decided to attempt a logarithmic transformation of the data, a common method to approximate the data distribution to normality. Unfortunately, this approach was unsuccessful in correcting the distributions, and the issue of normality remained.

The Gamma model assumes that the response variable ($\Delta Volume$) is continuous, strictly positive, and right-skewed. In addition, the normality assumption was explicitly assessed using the Shapiro-Wilk test, which provided strong evidence against conformity to the Normal distribution for all variables analysed ($p\text{-value} < 0.001$), and even after applying logarithmic transformations normality could not be achieved. The choice of fitting a Generalized Linear Model (GLM) with a Gamma distribution is therefore grounded in the empirical characteristics of this variable and in the limitations of the classical linear model. Ordinary linear regression assumes normally distributed, symmetric errors with support on the entire real line and constant variance, an assumption that is incompatible with a variable such as $\Delta Volume$, which is strictly positive, asymmetric,

and exhibits increasing variability as the mean level rises. The Gamma family, in contrast, is appropriate for continuous, positive, and skewed data and belongs to the exponential family, which allows its use within the Generalized Linear Models framework (Dobson and Barnett 2008). Thus, by adopting a Gamma GLM, is obtained a model that respects the support of the response variable.

The sample used in this study was segmented into two subsets: one for training and one for validation. The training subset, composed of 80% of the data, was used in the construction of the generalized linear model, adjusting the model parameters based on the observed data. The sample selection was performed randomly, ensuring that both subsets adequately represented the original distribution of the data.

The validation subset was reserved to test the performance of the model with data not used during the training phase. This process is essential to verify the generalization capacity of the model, its ability to make good predictions on new and unknown data. Validation plays an important role in mitigating problems such as overfitting, which occurs when the model excessively adapts to the training data, becoming highly specific to those data and, therefore, presenting unsatisfactory performance on previously unseen data.

To fit the data to a Gamma distribution, it is necessary that all values be strictly positive, since the probability density function of the Gamma distribution is not defined for values equal to zero. Therefore, in cases where the $\Delta Volume$ resulted in zero, a very small value, such as 0.0001, was adopted to enable the adjustment of the distribution.

This data treatment made it possible to preserve the nature of the observed values, ensuring that the statistical model could be applied appropriately, without violating the assumptions of the adopted distribution. The choice of the value 0.0001 was made based on the need to maintain a sufficiently small value, so as not to significantly impact the behaviour of the model, but sufficient to allow the estimation of the parameters within the structure of the Gamma distribution.

The Pearson Correlation Coefficient measures the intensity and direction of the linear relationship between two quantitative variables. It varies between -1 and 1 , where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates the absence of linear correlation. For the interpretation of the results, the following criteria were adopted: weak correlation (0.0 to 0.3), moderate correlation (0.3 to 0.7), and strong correlation (0.7 to 1.0).

Pearson's chi-square (χ^2) statistic was used to assess the goodness-of-fit of the fitted Gamma GLM, comparing the model-based expected values with the observed data. The

test was performed at a 5% significance level, and non-significant p-values ($p > 0.05$) were interpreted as evidence that the model provides an adequate fit.

For the evaluation of adjusted models, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used. Both were widely used for model comparison, penalizing model complexity and helping to avoid overfitting.

AIC favours models that best explain the data, while penalizing the inclusion of additional parameters. The model with the lowest AIC value is considered the best among the compared models. BIC, in turn, imposes a greater penalty for more complex models, especially when the sample size is large. The best model is the one with the lowest BIC value.

In addition to these information criteria, Nagelkerke's R^2 was adopted. This index is based on the ratio between the likelihood of the fitted model and that of the null (with only intercept) model and can be interpreted as an indication of the proportion of variation in $\Delta Volume$ accounted for by the covariates. Higher values of Nagelkerke's R^2 suggest better overall model fit, although it should be viewed as a descriptive measure rather than a direct analogue of the R^2 from linear regression.

Residual analysis is a crucial step in the evaluation of regression models, including Generalized Linear Models (GLM). The response residuals are the difference between the observed values and the values adjusted by the model, reflecting the adequacy of the model and helping to identify patterns not captured. The main assumption is that the residuals should be random and should not exhibit systematic patterns when plotted against the adjusted values.

Working residuals are a modification of the response residuals, used specifically in iterative adjustment algorithms, such as the Newton-Raphson method. They are calculated from the response residuals but weighted by the estimated variance. These working residuals are essential as they aid in the convergence of iterative methods, such as the Newton-Raphson method. In addition, it is important that they are random in their distribution to ensure the effectiveness of the model adjustment.

Pearson residuals are defined as the ratio between the response residuals and the square root of the model variance. Pearson residuals are widely used to evaluate the adequacy of the adjusted model, with values close to zero expected if the model is adequately adjusted to the data.

Deviance residuals are a measure of the discrepancy between the adjusted model and a saturated model (which fits the data perfectly), calculated from the difference between the log-likelihood of the saturated model and the adjusted model. Like other types of residuals, deviance residuals are assumed to be random. Residuals with significantly high values may indicate that the model does not

adequately capture the variation in the data, suggesting the need for an alternative or better-specified model.

The analysis of residuals in a General Linear Model is essential to ensure the validity of the results. The identification of patterns in the residuals may indicate the need for adjustments to the model or the exploration of additional variables.

To evaluate the adequacy of the model, one of the main graphical tools used is the plot of deviance residuals versus the adjusted values. This graph is essential for identifying heteroscedasticity problems, which may indicate that the model is not adequately capturing the variation of the data.

In a well-adjusted model, the residuals are expected to be randomly distributed around zero, without presenting clear patterns. If there is an increasing or decreasing pattern in the residuals, this may suggest that the variance is not being modelled correctly. This generally indicates the need to adjust the model or reconsider the variables included to better capture the structure of the data.

Influence analysis is an important part of the evaluation of regression models, as it identifies observations that have a disproportionate impact on the results. In this context, two measures stand out: Cook's distance and the influence observed by the h_{ii} and η_i indices.

Cook's distance assesses the influence of an observation on the fit of the model, quantifying the effect that excluding a data point would have on the estimation of the model parameters. D_i (Cook's Distance) values greater than 0.8, or more conservatively, $4/n$, where (n) is the total number of observations, indicate that the observation is potentially influential. This means that this observation has a significant impact on the estimated parameters of the model and may be a point of attention when evaluating the robustness of the model fit. It is recommended to analyse such observations carefully to consider the possibility of outliers or errors in the data.

The h_{ii} index represents the diagonal value of the projection matrix, indicating the influence that each observation has on its own predictions. Values close to 0.8 suggest great influence and may significantly affect the estimated coefficients, while values close to 0 indicate little or no influence.

The η_i index is the ratio between the variation in the prediction of a model with the inclusion of observation (i) and the variation in the prediction without it. This index helps to measure the specific impact that an individual observation has on the stability of the model's predictions, indicating

whether the inclusion of that observation results in significant changes in the projected results. When η_i presents high values, this may signal that observation i exerts considerable influence, potentially affecting the robustness of the model's conclusions.

Thus, the relationship between h_{ii} and η_i can be interpreted as follows: high values of both indicate that the observation has a significant influence on the model's predictions, while low values suggest that the observation has little or no influence. The analysis of these indices, together with Cook's distance, offers a comprehensive understanding of the influence of observations on a regression model, making it essential to ensure the robustness and validity of the results.

Results

Regarding the behaviour of $\Delta Volume$, it was observed that the amount of liquid added each hour was approximately 0.85 L, with a slightly lower median of 0.41 L and a standard deviation of 1.10 L. Additionally, it exhibited a skewness coefficient of approximately 1.50, indicating a higher concentration of lower $\Delta Volume$ values. Concerning external temperature, the recorded median was 27.60 °C, indicating that external temperature values were predominantly high in many measurements. With respect to external relative humidity, the observed median was 85.55%, demonstrating that external relative humidity levels were high in half of the measurements. Internal relative humidity showed a median of 58.20%, suggesting a balance of levels, with a tendency toward moderately high values. The descriptive statistics for these variables are summarized in Table 1.

To preliminarily explore the linear association between the dependent variable $\Delta Volume$ and the explanatory variables, Pearson's correlation coefficients were calculated. Overall, weak correlations were observed, with values close to zero for all covariates, as shown in Table 2.

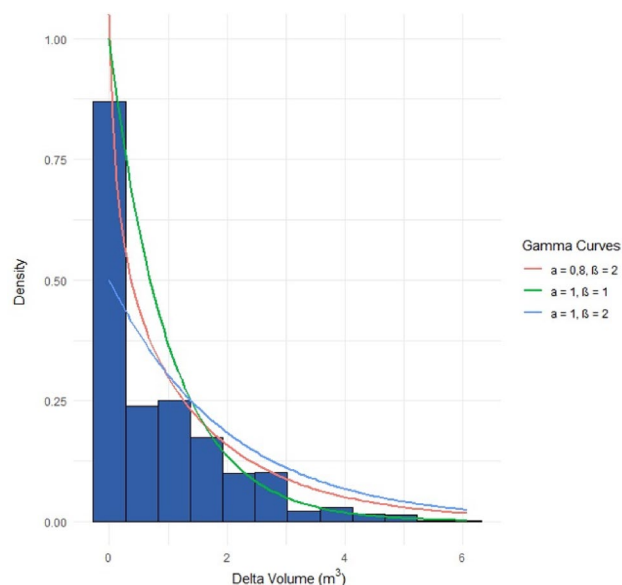
To support the use of a Gamma Generalized Linear Model (GLM) for modelling the response variable, a histogram was generated to assess whether the distribution of the observed data was compatible with the shapes commonly assumed by the Gamma family (Fig. 1). It was observed that the data of the dependent variable $\Delta Volume$ are right-skewed with rapid decay, and the simulated Gamma curves

Table 1 Descriptive statistics for $\Delta Volume$ and climatic explanatory variables

Statistic	$\Delta Volume$	T_{ext}	RH_{ext} (%)	RH_{int} (%)	T_{int} (°C)	P_{int} (°C)
Mean	0.85	28.29	81.09	57.69	21.42	12.46
Median	0.41	27.60	85.55	58.20	21.28	12.79
Standard deviation	1.10	3.84	17.02	11.52	0.58	2.86
Skewness coefficient	1.50	0.57	−0.90	0.02	4.02	−0.13

Table 2 Pearson correlation between Δ volume and explanatory variables

Explanatory variables	$\Delta Volume$
T_{ext} (°C)	-0.04
RH_{ext} (%)	0.10
T_{int} (°C)	0.01
RH_{int} (%)	0.00
P_{int} (°C)	0.00

**Fig. 1** Histogram of the $\Delta Volume$ variable and gamma distribution**Table 3** Candidate Gamma GLM models for $\Delta Volume$ and their AIC and BIC values

Models	AIC	BIC
Intercept + RH_{int}	-2967	-2952
Intercept + T_{ext}	-2965	-2951
Intercept + RH_{ext}	-2964	-2950
Intercept + P_{int}	-2967	-2943
Intercept + RH_{ext} + T_{ext}	-2967	-2943
Intercept + RH_{ext} + T_{ext} + Interaction	-2967	-2943

with varying α and β parameters indicate that the Gamma distribution fits well to the data.

The next step was to verify candidate models in which the variables are considered significant and with the lowest AIC and BIC values. Individually, the variables external relative humidity, external temperature, internal relative humidity, and dew point were significant, but the statistical significance of all variables appears only in the model with intercept, external relative humidity, external temperature, and an interaction between these variables. As all models are close in terms of AIC and BIC, the most comprehensive model was selected, and a summary of the candidate

Table 4 Estimated regression coefficients for Δ volume

Coefficients	Estimate	Std. error	t value	Pr(> t)
Intercept	3.307	1.0841	3.051	<0.01
T_{ext}	-0.078	0.0255	-3.051	<0.01
RH_{ext}	-0.038	0.0093	-4.160	<0.01
$T_{ext} \cdot RH_{ext}$	0.001	0.0002	6.259	<0.01

Gamma GLM specifications for $\Delta Volume$ and their corresponding AIC and BIC values is provided in Table 3.

In this perspective, the mean of the dependent variable $\Delta Volume$ is modelled according to Table 4 and represented in Eq. (1):

$$\Delta Volume = 3.307 - 0.078 T_{ext} - 0.038 RH_{ext} + 0.001 (T_{ext} \cdot RH_{ext}) \quad (1)$$

This indicates that both the external temperature and the external relative humidity exert a negative influence on the average variation of the condensate water volume, which does not reflect the physical phenomenon. Furthermore, all coefficients in the model are statistically significant, suggesting that the external temperature, the external relative humidity, and their interaction have a relevant impact on the analysed dependent variable. In this way, when considering Eq. (1), the effect of an external temperature of 30 °C and a relative humidity of 60% on the dependent variable was verified, and an estimated value of the dependent variable of approximately 0.80 L/h was observed.

In generalized linear models, the Deviance measure follows a chi-square distribution under certain conditions. When the null hypothesis is true, that is, when the fitted model is correct, the test statistic follows a chi-square distribution. In the analysed case, the null hypothesis was rejected, with a deviance of 7414.3 (p-value < 0.01).

In general, the smaller this measure, the smaller the uncertainty about the variable $\Delta Volume$, indicating that the explanatory variables are very important to understand the changes in volumes. However, when calculating Nagelkerke's R^2 , an indicator similar to the traditional coefficient of determination, it is observed that the model explains only 4% of the variation in the variable $\Delta Volume$.

Another important measure evaluated was the sum of Pearson squares, also known as Pearson's chi-square. In generalized linear models (GLM), the sum of Pearson squares is a statistical measure that, through the chi-square distribution, helps to test hypotheses about the discrepancy between the expected values and the observed values in a model. The test statistic was 1,304.95 (p-value < 0.01). In this sense, the null hypothesis that the model fits the data well is rejected.

The next step was to analyse the residuals of the model (Fig. 2), focusing on three aspects: independence,

homoscedasticity, and randomness of errors. Thus, it was observed that the residuals in relation to the values predicted by the model did not present sequential pattern or signs of dependence over time. Regarding homoscedasticity, the variance of the residuals should be constant at all levels of prediction. A uniform dispersion was verified, which suggests that the variance does not change with the values of the linear predictor. In addition, the analysis of randomness observed that there are no visible patterns in the data, such as curves, indicating that the data are random, which suggests that there is a significant level of error and that the model does not fit the data well.

Finally, an influence analysis was performed to identify observations that have a disproportionate impact on the results. The calculation of h_{ii} was approximately 0.015, which measures the effect that the observation has on the predicted values. Few observations exceeded this threshold; and these were removed, thus the model was re-estimated. However, this removal did not result in major differences in the results. Considering Cook's distance, (Fig. 3), only one observation exceeded the value of 0.8. This observation was removed, and the model was estimated again, but, as before, without major changes in the results.

Fig. 3 Influence via h_{ii} and Cook's distance

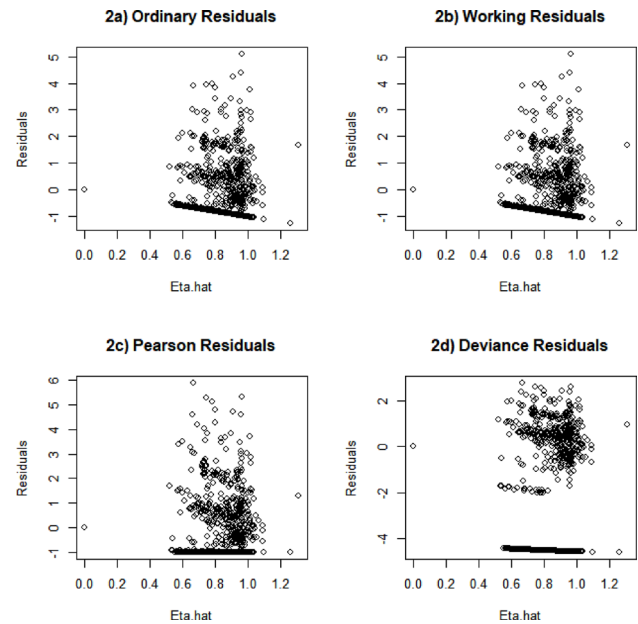
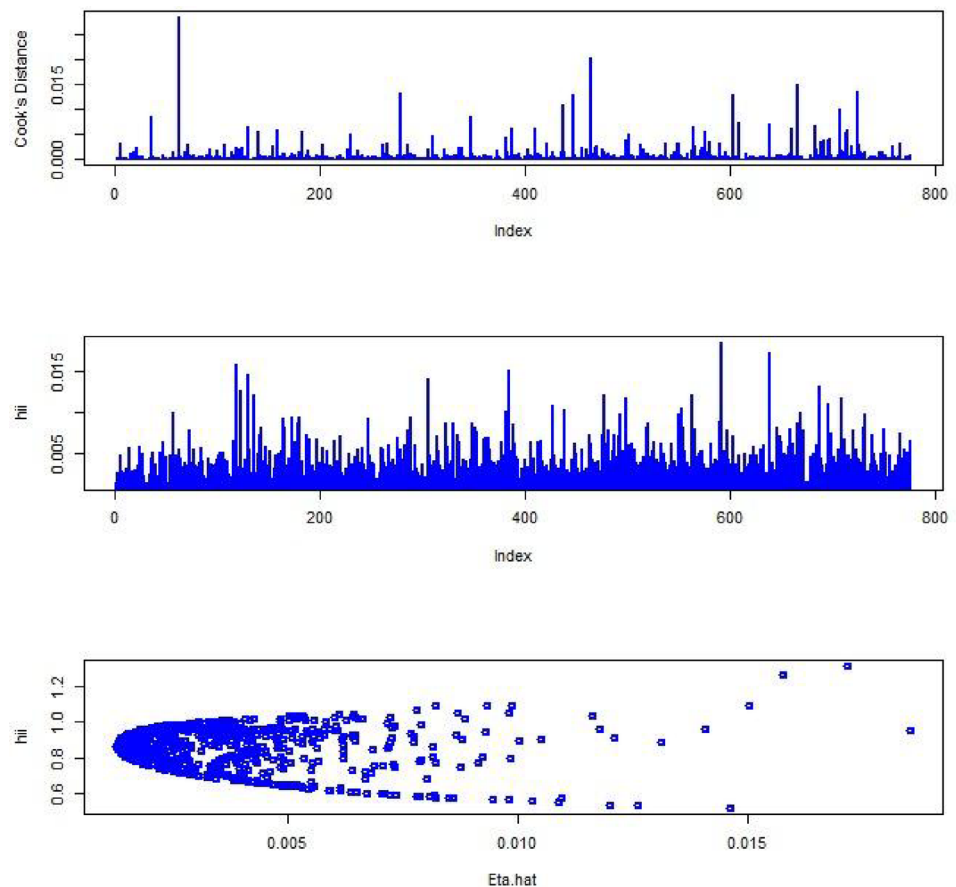


Fig. 2 Four types of model residuals

Discussion

The results indicate that, although external temperature and relative humidity were statistically significant predictors

in the Gamma GLM, their explanatory power was limited, as evidenced by the low Nagelkerke's R^2 (4%) and poor goodness-of-fit indicators. This suggests that condensate water production cannot be adequately explained by simple climatic variables alone, particularly in small-scale split air-conditioning systems. This might explain the weak correlations observed between $\Delta Volume$ and the analysed variables. While previous studies have reported the influence of temperature and humidity on condensate generation (e.g., Algarni et al. 2018; Dhamodharan et al. 2024a), such relationships are often derived under controlled or large-scale system conditions. In contrast, the present study highlights the high variability and stochastic behaviour of condensate production in real operating environments.

Several models with different variables were tested aiming for the best data fitting, but this attempt was unsuccessful for the analysed sample. The difficulty in developing a robust model stem from the lack of significant patterns between the dependent variable and the explanatory variables in the available experimental data. It is important to emphasize that these findings are preliminary and derived from a specific case study. Furthermore, the present experimental study dealt with primary data. Algarni et al. (2018) noted that theoretical models often overestimate condensate generation compared to experimental data due to idealized assumptions embedded in these models.

The production of condensed water in air conditioning systems is intrinsically linked to the compression cycle, which is marked by the successive activation and deactivation of the equipment. In oversized systems, shorter operational cycles are often observed, potentially leading to insufficient time for the evaporator coil to reach its minimum stable temperature. Additionally, setpoint temperatures set in close proximity to the ambient temperature can reduce the frequency of compressor operation, thereby limiting the generation of condensate. Conversely, in hot climates, the compressor tends to operate more frequently and for longer cycles, which may favour an increased production of condensed water.

Poredos et al. (2021) emphasized the critical role of coil geometry in influencing system performance and condensate production. While the current study did not explicitly evaluate geometric parameters, this omission may have contributed to the difficulty in establishing a stable and statistically significant predictive model. Furthermore, Algarni et al. (2018) noted that theoretical models often overestimate condensate generation compared to experimental data due to idealized assumptions embedded in these models.

Results from this study may complement existing knowledge by demonstrating the limitations of climate-based predictive approaches when applied to small-scale systems. A key implication of this study is that robust predictive

models for condensate generation will likely require hybrid approaches, integrating climatic data with operational parameters and possibly data-driven techniques such as machine learning.

Limitations of this study must be acknowledged. Data were collected from a single air-conditioning unit over a limited time period, restricting variability and generalisability. Expanding the dataset to include different system types, capacities, and seasonal conditions is essential for improving model performance and reliability. Despite the limitations faced, this research created important opportunities for future studies that can collect a greater variability of data for a more prolonged period. This can help capture the temporal variability and long-term patterns present in the data, offering a more complete view of how different climatic conditions affect the performance of air conditioning systems. In addition, it is crucial to verify and adjust the seasonal patterns, considering the changes in the data due to climatic variations. Another important point is to increase the number of samples of air conditioning equipment considered. The inclusion of different models and cooling capacities can provide important information about the performance in a wider range of operating conditions. This is especially important to understand how different air conditioning systems respond to a diversity of typologies of built and climatized environments.

Conclusions

This study investigated the influence of climatic variables on condensate generation from a split air-conditioning system under real operating conditions, considering internal and external temperature, relative humidity, and dew point as explanatory variables. The experimental setup operated for 968 h, enabling the development of a preliminary statistical model. Relative humidity (internal and external) showed positive correlations with condensate production, whereas temperature exhibited negative correlations. Dew point also presented a positive association, indicating that more humid conditions favour condensation. Overall, findings suggest that:

- A 9000 BTU/h hi-wall split-type air conditioner generated, on average, 0.85 L/h of condensate, indicating its potential as an alternative non-potable water source in buildings.
- External temperature and relative humidity were statistically significant predictors; however, their explanatory power was limited (Nagelkerke's $R^2 = 0.04$).
- Weak correlations indicate that condensate production is not primarily governed by climatic variables alone.

- The results highlight the limitations of simplified predictive models and the need for integrated approaches incorporating both climatic and operational variables.
- Despite its limitations, this study provides empirical evidence supporting the development of more robust estimation methods for condensate water recovery in buildings.

These findings are preliminary and derived from a single case study. Nevertheless, they demonstrate the potential for advancing statistical modelling of condensate production and improving the understanding of environmental influences on water generation. Future research should focus on expanding datasets across different climatic conditions, system types, and operational regimes to enhance predictive accuracy.

Author contributions Conceptualization, D.C. and D.S.; methodology, D.C., D.S., A.T.S. and J.G.R.; validation, D.C., D.S. and J.G.R.; formal analysis, D.C. and J.G.R.; investigation, D.C. and A.T.S.; resources, D.S.; data curation, D.C. and J.G.R.; writing - original draft preparation, D.C.; writing - review and editing, D.C., D.S., A.T.S.; visualization, D.C. and J.G.R.; supervision, D.S.; project administration, D.S.; funding acquisition, D.S. All authors have read and agreed to the submitted version of the manuscript.

Funding This research was funded by *Fundação de Apoio à Pesquisa do Distrito Federal* (FAPDF), with the grant number 00193-00001811/2023-76 awarded to D.S. and by the University of Brasília (UnB), with the grant number 23106.049861/2026-51 awarded to D.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. FAPDF is a public foundation dedicated to promoting scientific, technological, and innovation-related development within the Federal District of Brazil. It plays a key role in funding academic and institutional research by offering grants, scholarships, and support programs aimed at strengthening research capacity in universities and research centers.

Declarations

Conflict of interest The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Algarni S, Saleel CA, Mujeebu MA (2018) Air-conditioning condensate recovery and applications—current developments and challenges ahead. *Sustain Cities Soc* 37:263–274. <https://doi.org/10.1016/j.scs.2017.11.032>
- ANA (2019) Conjuntura dos recursos hídricos no Brasil 2019: informe anual. Agência Nacional de Águas, Brasília
- ANA (2020) Conjuntura dos recursos hídricos no Brasil 2020: informe anual. Agência Nacional de Águas, Brasília
- Beithou N, Khalid MB, As'ad S, Alsaqoor S, Borowski G, Alshabat N, Andruszkiewicz A (2024) Atmospheric water harvesting technology: review and future prospects. *J Ecol Eng* 25:291–302. <https://doi.org/10.12911/22998993/182814>
- Cattani L, Magrini A, Cattani P (2018) Water extraction from air by refrigeration: experimental results from an integrated system application. *Appl Sci* 8:2262. <https://doi.org/10.3390/app8112262>
- Cattani L, Cattani P, Magrini A (2023) Air to water generator integrated system real application: a study case in a worker village in United Arab Emirates. *Appl Sci* 13:3094. <https://doi.org/10.3390/app13053094>
- Chantasiriwan S (2024) The use of air cooling system in combined cycle power plant as atmospheric water generator. *Eng* 5:1850–1862. <https://doi.org/10.3390/eng5030099>
- Cichoń A, Worek W (2021) Analytical investigation of a novel system for combined dew point cooling and water recovery. *Appl Sci* 11:1–18. <https://doi.org/10.3390/app11041481>
- Congedo PM, Baglivo C, Negro G (2021) A new device hypothesis for water extraction from air and basic air condition system in developing countries. *Energies* 14:4507. <https://doi.org/10.3390/en14154507>
- Dhamodharan P, Ayalur BK, Annamalai SK, Prabakaran R, Kim SC (2024a) Development and analysis of air-conditioning condensate assisted compact cooler unit: a novel approach in condensate recovery. *J Therm Anal Calorim* 149:3303–3316. <https://doi.org/10.1007/s10973-024-12923-0>
- Dhamodharan P, Ayalur BK, Prabakaran JR, Kim SC (2024b) Energy saving potential in radiant cooling system by utilizing air-conditioning condensate: a strategy for green building rating. *Appl Therm Eng* 236:121492. <https://doi.org/10.1016/j.applthermaleng.2023.121492>
- Dobson AJ, Barnett AG (2008) An introduction to generalized linear models. CRC
- Eades WG (2018) Energy and water recovery using air-handling unit condensate from laboratory HVAC systems. *Sustain Cities Soc* 42:162–175. <https://doi.org/10.1016/j.scs.2018.07.006>
- Elbrashy A, Vafai K, Elshennawy A, Ayman A, Elgebaly A, Rashad A (2025) Harvesting of condensate water from air conditioners in large institutions as a sustainable resource. *Water Resour Manag* 39:5289–5312. <https://doi.org/10.1007/s11269-025-04203-9>
- Luqman M, Al-Ansari T (2021) A novel integrated wastewater recovery, clean water production and air-conditioning system. *Energy Convers Manag* 244:114525. <https://doi.org/10.1016/j.enconman.2021.114525>
- Magrini A, Cattani L, Cattani L, Cartesegna M, Magnani L (2017) Water production from air conditioning systems: some evaluations about a sustainable use of resources. *Sustainability* 9:1309. <https://doi.org/10.3390/su9081309>
- Poredos P, Petelin L, Žel T, Vidrih B, Gatařić P, Kitanovski A (2021) Performance of the condensation process for water vapour in the presence of a non-condensable gas on vertical plates and horizontal tubes. *Energies* 14:2291. <https://doi.org/10.3390/en14082291>
- Sant'Ana D (2017) Reúso-DF: viabilidade técnica e operacional do aproveitamento de águas pluviais e do reúso de águas cinzas em edificações não-residenciais do Distrito Federal: relatório técnico

- 5/2017. Universidade de Brasília. <http://repositorio.unb.br/handle/10482/40689>. Accessed 25 June 2025
- Sant'Ana D, Lima T (2021) Mudanças climáticas, mercantilização das águas e sistemas de água não potável. In: Moraes GGB, Montezuma TFPF, Ferraço AAG (eds) (orgs) Estudos de direito das águas: desafios jurídicos, sociais e agravantes climáticas, 1 edn. Lumen Juris, Rio de Janeiro, pp 229–255
- Zolfagharkhani S, Zamen M, Shahmardan MM (2018) Thermodynamic analysis and evaluation of a gas compression refrigeration cycle for fresh water production from atmospheric air. *Energy Convers Manag* 170:97–107. <https://doi.org/10.1016/j.enconman.2018.05.016>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.