Received: 22 November 2024 | Revised: 8 February 2025 | Accepted: 27 March 2025

DOI: 10.1111/1556-4029.70046

JOURNAL OF FORENSIC SCIENCES AAFS

ORIGINAL PAPER

Anthropology

Identifying factors that help improve existing decomposition-based PMI estimation methods

Anna-Maria Nau PhD[1,2] | Phillip Ditto BA[3] | Dawnie Wolfe Steadman PhD[3] | Audris Mockus PhD[1]

[1]Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, Tennessee, USA

[2]The Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee, Knoxville, Tennessee, USA

[3]Department of Anthropology, The University of Tennessee, Knoxville, Tennessee, USA

Correspondence
Anna-Maria Nau, Department of Electrical Engineering and Computer Science, The University of Tennessee, 1520 Middle Dr, Knoxville, TN 37996, USA.
Email: annanau123@gmail.com

Funding information
National Institute of Justice, Grant/Award Number: 15PNIJ-21-GG-04161-SLFO and 2018-DU-BX-0181

Abstract

Accurately assessing the postmortem interval (PMI) remains a challenging task in forensic science. Existing regression models use the decomposition score to predict the PMI or accumulated degree days (ADD) but are often imprecise and rely on small sample sizes. This study explores if we can construct more accurate outdoor PMI estimation models using (a) a larger sample, (b) more sophisticated statistical models, and (c) additional predictors derived from demographic and environmental factors. Using a sample of 213 human subjects from a human decomposition photographic dataset collected at the [removed for double-blind review], we evaluated existing outdoor PMI and ADD formulae developed by Gelderman et al. [Int J Legal Med, 2018, 132, 863] and also developed more sophisticated models that incorporate additional predictors. Models using the total decomposition score (TDS), demographic factors (age, sex, and BMI), and weather-related factors (season and humidity history) reduced PMI and ADD prediction errors by over 50%. The best PMI model, incorporating TDS, demographic, and weather predictors, achieved an adjusted *R*-squared of 0.42 and an RMSE of 0.76. It had a 15% lower RMSE than the TDS-only model to predict PMI and a 54% lower RMSE than the pre-existing PMI formula. Similarly, the best ADD model, using the same predictors, achieved an adjusted *R*-squared of 0.54 and an RMSE of 0.73. It had a 10% lower RMSE than the TDS-only model to predict the ADD and a 55% lower RMSE than the pre-existing ADD formula. These results demonstrate that significant improvements in accuracy can be achieved using readily available predictors.

KEYWORDS
accumulated degree days, decomposition, forensic anthropology, linear regression, PMI, total decomposition score

Highlights

- Complex modeling designed to enhance decomposition-based PMI estimation.
- Modeling approach with 213 human donors using demographic and environmental data.
- Development of regression models to predict time since death in outdoor decomposition cases.
- Incorporating diverse factors improved PMI and ADD prediction accuracy.

1 | INTRODUCTION

Determining the postmortem interval (PMI), or time since death, is an important task in human remains cases. An accurate estimation of the PMI can substantially narrow down the list of potential decedents the remains might correspond to, facilitating the eventual identification of the individual. In cases involving homicide, law enforcement professionals can utilize the PMI to eliminate potential suspects and validate witness accounts. Additionally, having knowledge of the PMI aids in establishing the spectrum of natural occurrences and environmental influences that impacted the remains, thereby enabling a more comprehensive analysis.

Human decomposition is a natural and continuous process involving the breakdown of tissues after death that can span from weeks to years, contingent upon a range of biological and environmental factors. Comprehending how these factors influence the rate of human decomposition is pivotal in ascertaining the PMI [1–3].

Over the years, researchers in the field of forensic anthropology have developed methods that divide the human decomposition process into high-level categories or stages of decay. Each stage is characterized by distinct decomposition phenomena, aiding in the estimation of the PMI [1, 4]. One such method was developed by Gelderman et al. [5], a simplified version of Megyesi et al.'s [2] method, which involves a human decomposition scoring method and regression formulae to estimate the PMI and the accumulated degree days (ADD). The scoring method separates the human body into distinct anatomical regions, which are each assigned a score reflecting the amount of decomposition that has occurred. The individual scores are summed to obtain the total decomposition score (TDS), which is then used to estimate the PMI and ADD using their developed formulae. The ADD is the sum of the average daily ambient temperatures (rather than calendar days) between the date of death and discovery. Specifically, it represents the heat energy units required for the chemical and biological reactions that decompose a body. Incorporating temperature data and treating decomposition as a semi-continuous variable enabled the measurement of the impact of temperature on decomposition rates. This approach improved PMI estimates by modeling decomposition as dependent on both time and accumulated temperature [2, 5, 6]. In their work, Gelderman et al. [5] developed two PMI estimation formulae for both indoor and outdoor cases: one for directly estimating the PMI and another for estimating the ADD, from which the PMI could later be derived. For this study, we will focus exclusively on Gelderman et al.'s [5] outdoor PMI and ADD estimation formulae.

Gelderman et al.'s [5] outdoor PMI and ADD estimation formulae only considered the TDS as a predictor variable, but what about other factors, such as decedent demographics and environmental information? It is known from existing literature that a wide variety of factors, including body mass [7–9], temperature [3, 10], humidity [3, 11], and insect activity [6, 12] also affect the decomposition process and should be taken into consideration when determining the PMI. Therefore, expanding the TDS-only PMI and ADD formulae by including such factors could potentially improve their prediction accuracy. Additionally, Gelderman et al.'s [5] outdoor PMI and ADD estimation formulae were developed on a small sample of only 12 outdoor cases. In fact, sample size is a crucial factor that not only affects a study's statistical power and precision but also influences its capacity to draw meaningful conclusions and apply the results to a wider context [13, 14]. And finally, to date, no independent validation studies have assessed the reliability or generalizability of Gelderman et al.'s [5] outdoor PMI and ADD estimation formulae. This gap in the literature underscores the importance of the present study, which directly evaluates the applicability and accuracy of these formulae to advance human decomposition research and improve time since death estimation.

The aim of this study is to investigate ways to improve outdoor decomposition-based PMI prediction by (a) using a larger sample size, (b) employing more flexible and advanced linear models, and (c) enhancing the TDS models with demographic and environmental factors known to affect the human decay process. Specifically, various univariate and multiple linear regression analyses using a combination of predictor variables to estimate the PMI and ADD will be conducted and evaluated. Similar to Gelderman et al.'s [5] study, our approach will be based on archived (digital) photographs. Additionally, Gelderman et al.'s [5] decomposition scoring method will be used to measure decomposition and calculate the TDS, and their PMI and ADD estimation formulae will be evaluated.

2 | MATERIALS AND METHODS

2.1 | The human decomposition photographic collection

The human decomposition photographic collection is a large-scale image dataset of decomposing bodies donated to the Forensic Anthropology Center at The University of Tennessee, Knoxville. The Center houses the Anthropology Research Facility (ARF). Forensic anthropologists from the ARF captured these images at non-uniform intervals, with one or more days between each capture. The images, taken from various angles, depict different anatomical areas to illustrate the various stages of human decomposition. The image resolutions vary from 2400×1600 up to 4900×3200. The dataset covers the period from 2011 to 2023 and comprises over 1.5 million images contributed by more than 800 donors.

2.2 | The study sample

The subjects were chosen with respect to the following attributes:

- Biological sex: female or male.
- Age groups defined by subtracting the age of the youngest from the age of the oldest donor and dividing by three: age < 49 years (younger), 49 ≤ age < 72 (middle), age ≥ 72 years (older).

- Body mass index (BMI) groups defined by the Center for Disease Control and Prevention (CDC): underweight (BMI < 18.5), healthy (18.5 ≤ BMI < 25), overweight (25 ≤ BMI < 30), and obese (BMI ≥30).
- PMI <6 months at the time the photograph was taken. The decision to include only PMIs of <6 months was primarily driven by our human decomposition dataset. Specifically, the number of subjects with a PMI exceeding 6 months dropped rapidly. To ensure the reliability and robustness of our models, we focused on subjects with PMIs under 6 months.

By considering the above attributes, the objective was to create a study sample that was both diverse and balanced, meaning it included subjects across both sexes, different age and BMI groups, and PMIs. The final study sample consisted of 213 subjects, with the attribute distribution plots shown in Figure 1. Out of the 213 subjects, 108 were female and 105 were male, with ages ranging from 26 to 96 years.

All subjects decomposed outdoors unclothed and intact, meaning no missing body parts, so that decomposition could be scored for the entire body in a standardized fashion. Additionally, no buried, burned, or submerged remains were used in this study because they decompose differently [15]. Since all donors were placed within the ARF, they decomposed in a similar environmental setting: an open-wooded, mostly shaded area with the ground consisting of soil and dead/decaying plant matter.

2.3 | Decomposition measurement

The state of decomposition was measured for each subject using the human decomposition scoring method proposed by Gelderman et al. [5]. To account for the differential decomposition that occurs in different body segments, this scoring method separates the human body into three anatomical regions: (a) the head (including the neck), (b) the torso, and (c) the limbs (including the hands and feet). Based on the morphological features present, each anatomical region is categorized into six stadia or stages, with the lowest (i.e., score 1) indicating no visible changes, and the highest (i.e., score 6) indicating complete skeletonization as shown by Tables 1–3. Once each anatomical region is scored, the three decomposition scores are summed to obtain the TDS, which ranges from 3 to 18.

The decomposition scoring was performed by author PD, a forensic anthropologist who has extensive experience in Gelderman

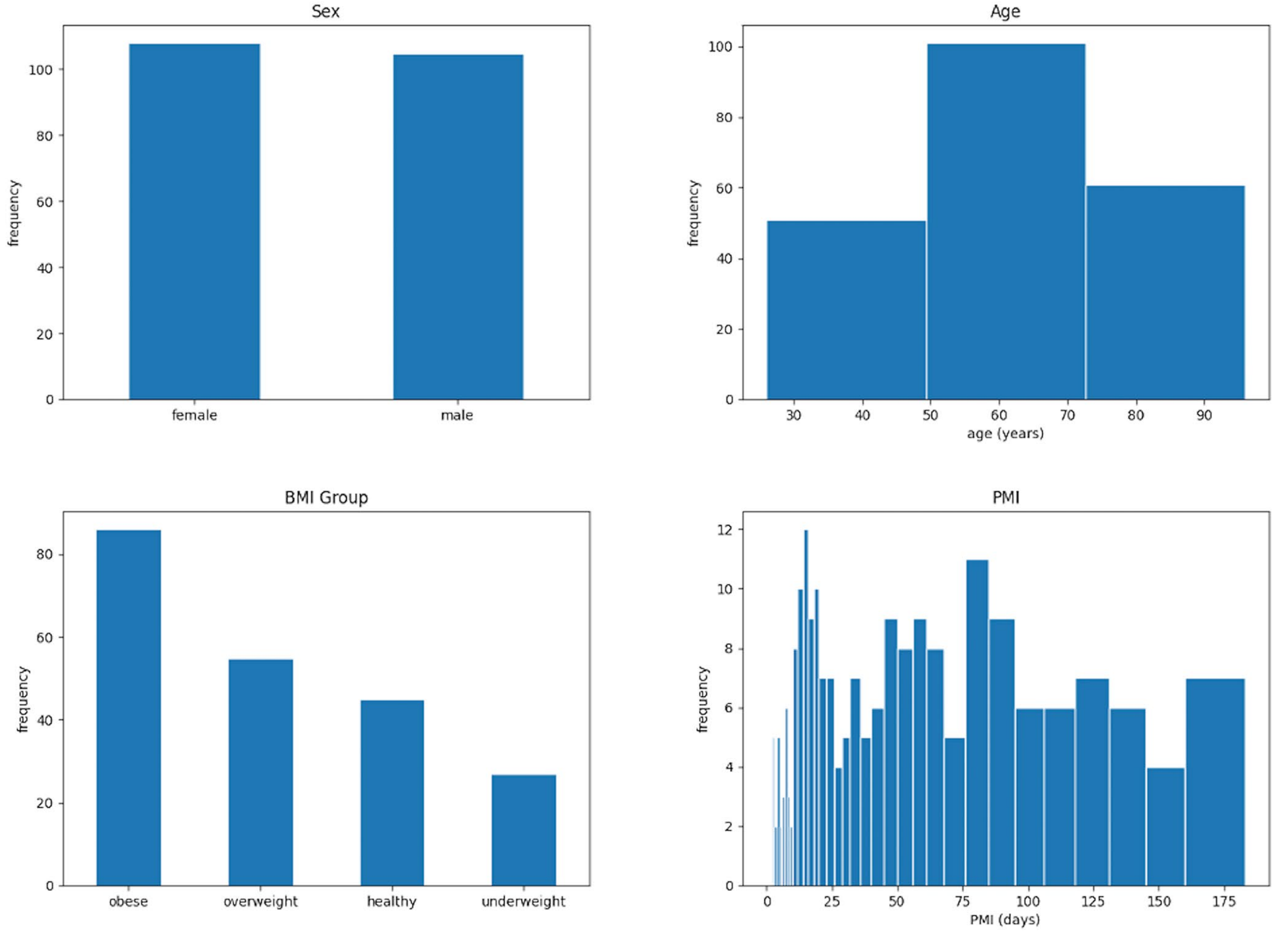

FIGURE 1 Study sample ($n = 213$) distribution plots by sex, age, BMI group, and PMI.

TABLE 1 Gelderman et al.'s [4] decomposition scoring method for the head. Point 1, point 2, etc. represent the different phenomena.

Region	Score	Description
Head	1	1.1. No visible changes
	2	2.1. Livor mortis and rigor mortis
		2.2. Eyes: cloudy and/or tache noir
		2.3. Discoloration: brownish shades particularly at the edges. Drying of nose, ears and lips
	3	3.1. Gray to green discoloration
		3.2. Bloating of neck and face is present and/or skin blisters, skin slippage and/or marbling
		3.3. Purging of decompositional fluids out of ears, nose and mouth and/or brown to black discoloration
	4	4.1. Caving in of the flesh and tissues of eyes and throat. Skin having a leathery appearance
		4.2. Partial skeletonization, joints still together
	5	5.1. Gross skeletonization, some joints disarticulated
	6	6.1. Complete skeletonization

TABLE 2 Gelderman et al.'s [4] decomposition scoring method for the torso. Point 1, point 2, etc. represent the different phenomena.

Region	Score	Description
Torso	1	1.1. No visible changes
	2	2.1. Livor mortis and rigor mortis
	3	3.1. Gray to green discoloration
		3.2. Bloating with green discoloration and/or skin blisters, skin slippage and/or marbling
		3.3. Rectal purging of decompositional fluids
		3.4. Post-bloating: release of abdominal gasses with discoloration changing from green to black
	4	4.1. Decomposition of tissue producing sagging of flesh. Caving in of the abdominal cavity
		4.2. Skin having a leathery appearance
		4.3. Partial skeletonization, joints still together
	5	5.1. Gross skeletonization, some joints disarticulated
	6	6.1. Complete skeletonization

et al.'s [5] scoring method, having used it in their own work at [removed for double-blind review] and throughout their graduate studies. This hands-on application of the method, combined with the

TABLE 3 Gelderman et al.'s [4] decomposition scoring method for the limbs. Point 1, point 2, etc. represent the different phenomena.

Region	Score	Description
Limbs	1	1.1. No visible changes
	2	2.1. Livor mortis and rigor mortis
		2.2. Discoloration: brownish shades particularly at the edges. Drying of fingers and toes
	3	3.1. Skin blisters and/or skin slippage and/or marbling
		3.2. Gray to green discoloration
		3.3. Brown to black discoloration
	4	4.1. Skin having a leathery appearance
		4.2. Partial skeletonization, joints and tendons still together
	5	5.1. Gross skeletonization, some joints disarticulated
	6	6.1. Complete skeletonization

anthropologist's overall expertise in forensic anthropology, ensures that the scoring method was applied with a high level of reliability. Specifically, three photos (i.e., one photo per anatomical region) for each subject in the study sample was assessed and scored on an in-house developed data visualization and annotation software called ICPUTRD [16]. The three photos per subject were semi-randomly selected by (a) randomly selecting one photo per anatomical region for each subject and (b) visually inspecting the randomly selected photos to ensure quality and clarity. If any photos were found to be of poor quality, they were replaced with another photo specific to the subject and anatomical region. This resulted in a total of 639 photos being scored – 213 subjects multiplied by three photos (i.e., one head, one torso, and one limb).

2.4 | Weather data and ADD calculation

Temperature data were collected from [removed for double-blind review], which is the closest National Weather Service station (approximately 10 miles away), in order to calculate the ADD. All temperature data was in the form of daily averages and recorded in degrees Celsius (C). Following Gelderman et al. [5], the ADD was calculated for each subject by adding the average daily temperatures above 0 degrees C (base temperature) from death until the photograph was taken.

2.5 | PMI and ADD estimation

For the PMI and ADD estimation the following were conducted: (a) evaluate Gelderman et al.'s [5] PMI and ADD formulae (1) and (2) on our larger study sample and (b) perform various univariate and multiple linear regression analyses using different combinations of the following input variables to predict the PMI and ADD:

- TDS
- Subject demographics, including biological sex, age at death, and BMI. These variables were tested for multicollinearity, which showed no strong linear relationships among them.
- Weather-related features, including the season of discovery (winter, spring, summer, or fall) and humidity history:

$$\mathrm{PMI} = 10^{(-0.93+(0.18\times\mathrm{TDS}))} \tag{1}$$

$$\mathrm{ADD} = 10^{(0.03+(0.19\times\mathrm{TDS}))} \tag{2}$$

In order to be used in the regression analysis, the categorical variables, sex and season of discovery, were converted to numeric by using one-hot encoding. Specifically, one-hot encoding creates a dummy variable for each category of the categorical variable with values 1 or 0 representing the presence or absence of the category. When including dummy variables in a regression analysis, one needs to be aware of the dummy variable trap, which is when two or more variables are highly correlated; in simple terms, one variable can be predicted from the others. For instance, knowing male = 0 implies female = 1, or knowing spring = 0, summer = 0, fall = 0, implies winter = 1, and vice versa. The solution to the dummy variable trap is to drop one of the categorical variables; that is, if there are m number of categories, use m−1 in the regression analysis. In doing so, the categorical sex variable was converted to one dummy variable with values 1 for male and 0 for not male (hence female). Similarly, the categorical season of discovery variable was converted to numeric by creating three dummy variables (i.e., spring, summer, and fall), each with values 1 and 0. Note, this dummy variable trap was automatically handled with Pandas, a Python package for data analysis and processing. The BMI was calculated from the estimated cadaver weight (in pounds) and height (in inches) measured at the time of intake.

To create the humidity history feature, humidity (relative) data were collected from the nearest National Weather Service station located at [removed for double-blind review]. Specifically, the humidity history feature was calculated as follows for each subject in the study sample: (a) obtain daily average humidity values for the 2 weeks prior to the date of discovery (or in this case, when the photograph was taken) and (b) calculate the average for the 2-week humidity history, resulting in the numeric variable called hum_hist. The 2-week humidity history was aggregated to produce a single summary value (i.e., the average humidity) to be used in the regression model. Future work should further evaluate this approach by using different humidity history ranges (e.g., 3 and 4 weeks) and various aggregation methods (e.g., median, mode, and standard deviation). Also, a 2-week temperature history (i.e., temp_hist) was initially included, but it was dropped from the regression models due to multicollinearity with other weather-based variables.

The statistical analysis was conducted using SPSS Statistics (version 29.0.2.0) and the Python package, statsmodels. Specifically, the following univariate and multiple linear regression analyses using the study sample of 213 subjects were conducted and compared:

- TDS to predict the PMI and ADD (univariate).
- TDS + demographic features (sex, age, and BMI) to predict the PMI and ADD (multiple).
- TDS + demographic features (sex, age, and BMI) + weather features (spring, summer, fall, and hum_hist) to predict the PMI and ADD (multiple).

Similar to Gelderman et al. [5], for the linear regression analysis, the PMI and ADD were naturally log-transformed to achieve a more linear relationship between the TDS and PMI, and the TDS and ADD as shown by Figures 2–5. When choosing the type of log transformation (e.g., natural logarithm with base e vs. base 10), the choice generally does not matter as both transformations will produce the same underlying linear relationship, only differing slightly in the scale of the y-axis due to the different bases. In most cases, the natural logarithm (base-e) is preferred due to its widespread use in mathematical and statistical applications.

All continuous independent variables, including the TDS, age, BMI, and hum_hist, were standardized to be on the same scale and thus contribute equally to the regression analysis. *Z*-score standardization was applied, rescaling the original variable to have a mean of zero and a standard deviation of one. Mathematically, this involves subtracting the mean of the original variable from the raw value and then dividing it by the standard deviation of the original variable, as shown in Equation (3):

$$\text{standardized } X = \frac{X - \text{mean}}{\text{standard deviation}} \tag{3}$$

The metrics used to evaluate the different regression analyses were the adjusted *R*-squared (adj. R^2) and root mean squared error (RMSE). The adj. R^2, as shown by Equation (5), is comparable to the R^2 or the coefficient of determination, as shown by Equation (4), as its value lies between 0 and 1 (the higher, the better) and explains the proportion of variance for a dependent variable with respect to independent variable(s) in the regression model [17]:

$$R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\sum \left(y_i - \hat{y}_i\right)^2}{\sum \left(y_i - \overline{y}\right)^2} \tag{4}$$

where RSS = residual sum of squares, TSS = total sum of squares, y_i = actual value of sample *i*, $\hat{y}_i$ = predicted value of sample *i*, $\overline{y}$ = mean value of *y*.

$$adj.\, R^2 = \frac{\left(1 - R^2\right)(n - 1)}{(n - k - 1)} \tag{5}$$

where n = number of samples, k = number of independent variables.

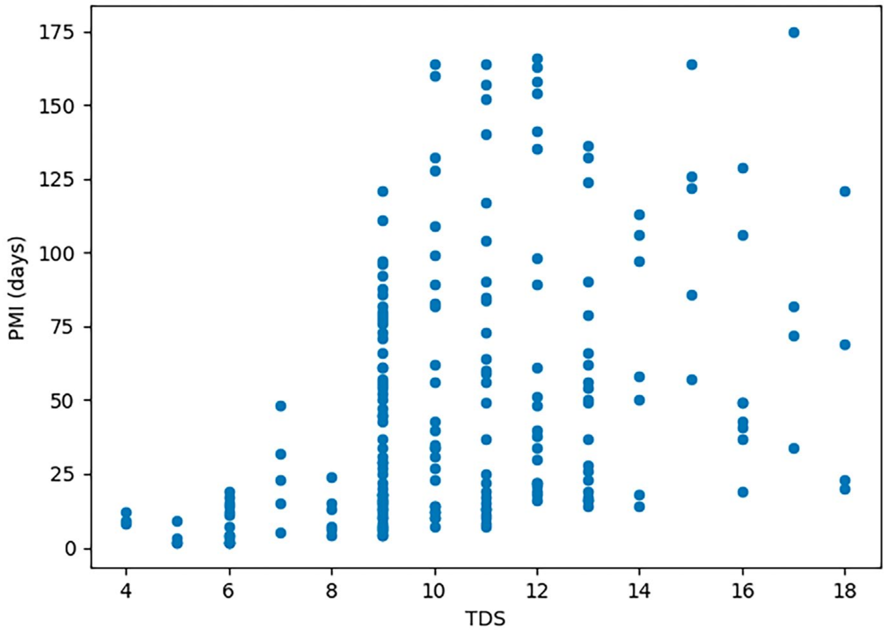

FIGURE 2 Plot of TDS versus PMI ($n = 213$).

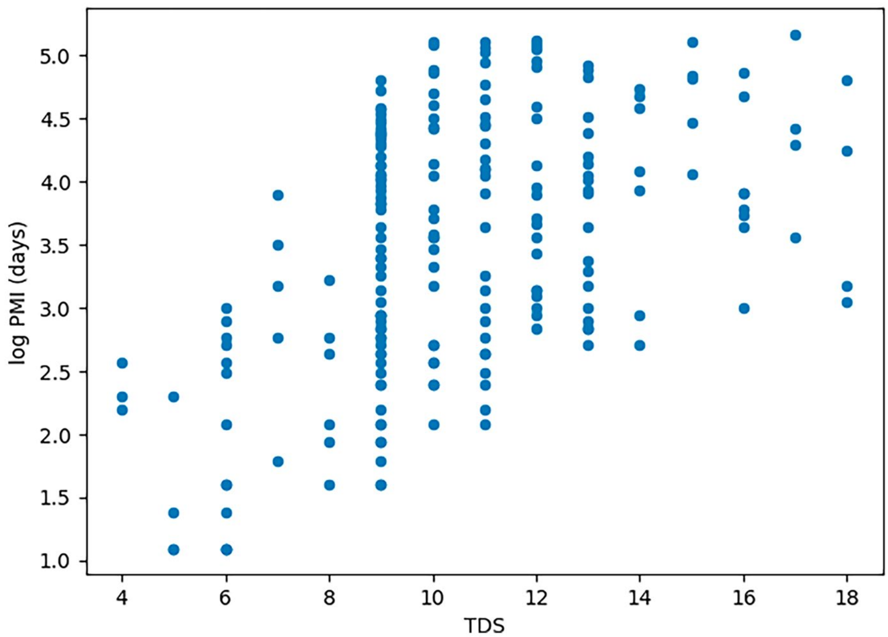

FIGURE 3 Plot of TDS versus naturally log-transformed PMI ($n = 213$).

The difference between R^2 and adj. R^2 is that R^2 assumes all independent variables considered affect the model's outcome, whereas adj. R^2 accounts only for those variables that actually impact the model's performance. Adj. R^2 penalizes the model for including unnecessary variables that do not improve its accuracy. For instance, when multiple linear regression models are built, such as in this study with the forward addition method, at each iteration independent variables are added, the R^2 will keep increasing, but the adj. R^2 will only increase when the variable actually affects the dependent variable. If a variable is non-significant, the R^2 value will still increase, but the adj. R^2 value will decrease at that point. The second model evaluation metric used was the RMSE, as shown by Equation (6), which is the standard deviation of the residuals, measuring the average difference between a model's predicted values and the actual values [18]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum \left(y_i - \hat{y}_i\right)^2}{n}} \tag{6}$$

where n = number of samples, y_i = actual value of sample i, $\hat{y}_i$ = predicted value of sample i.

Residuals represent the distance between the regression line and the data points. As the data points move closer to the regression line, the model has less error, lowering the RMSE and

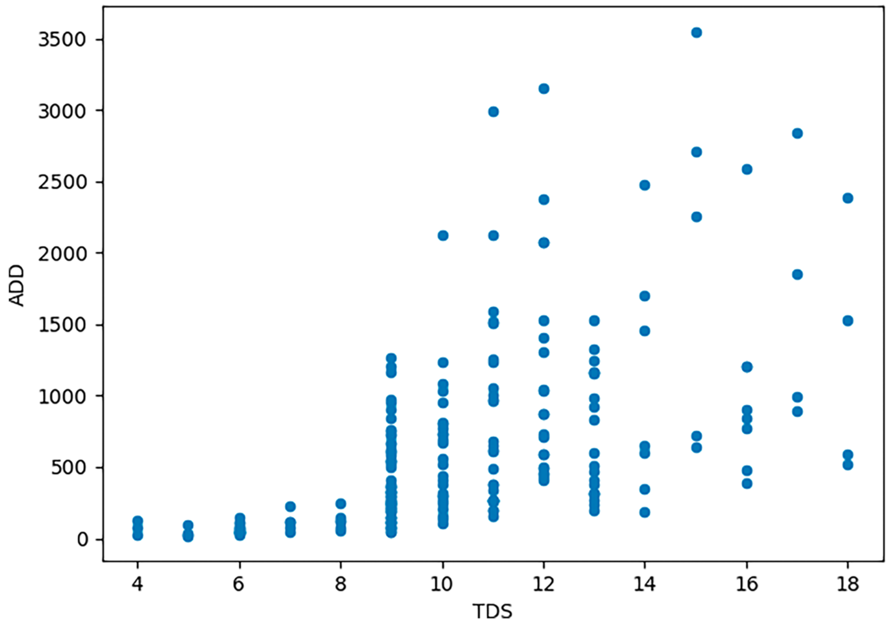

FIGURE 4 Plot of TDS versus ADD ($n=213$).

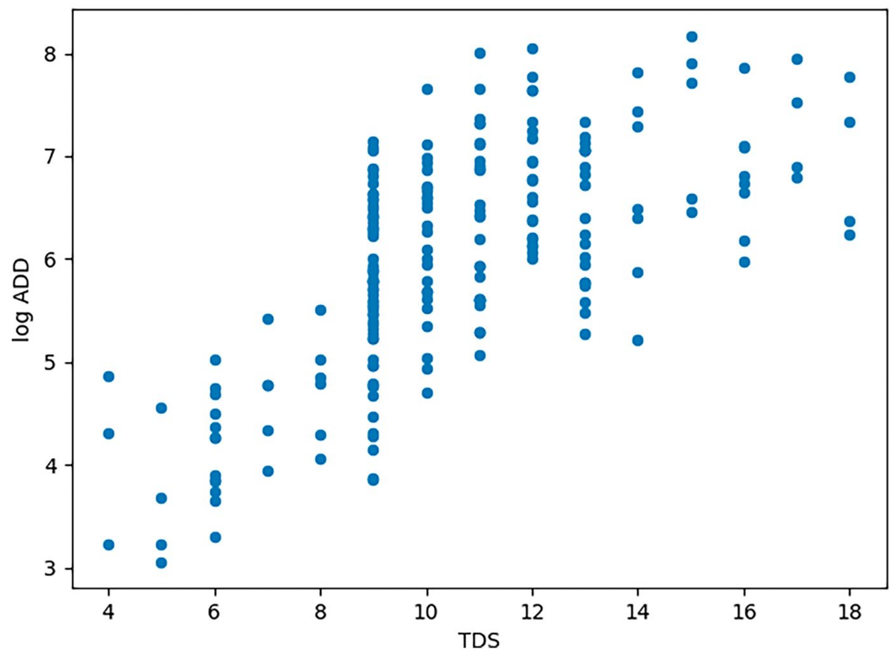

FIGURE 5 Plot of TDS versus naturally log-transformed ADD ($n=213$).

producing more precise predictions. The RMSE values range from 0 (perfect predictions) to positive infinity and are in the same units as the dependent variable. Note, the RMSE is almost identical to the standard error of the regression. One might wonder why RMSE was used instead of information criteria such as AIC or BIC for model evaluation. Our primary focus was on evaluating models based on their predictive accuracy (e.g., RMSE) rather than identifying statistically significant predictors. The regression coefficients and their p-values (with a significance level of 0.05) were calculated and reported. Since the predictor variables were standardized, the regression coefficients can be interpreted as the standardized regression coefficients. Finally, predicted versus actual scatter plots will be created to visualize and compare the fit of the different regression analyses. In these plots, the x-axis represents the actual values, and the y-axis represents the predicted values. Ideally, if the predictions are perfect, the points will lie along a straight line with a slope of 1.

3 | RESULTS

Table 4 shows the PMI and ADD estimation results. The first two rows (above the dashed line) give the results for Gelderman et al.'s [5] PMI and ADD formulae evaluated on our study sample. To ensure equal comparison among all regression experiments, the log-transformed versions of Gelderman et al.'s [5] PMI and ADD

TABLE 4 The univariate and multiple linear regression analysis results. Reported are the adj. R^2, the RMSE, and the back-transformed formula with the standardized regression coefficients.

Analysis (predictors, output)	Adj. R^2	RMSE	Formula (back-transformed)
TDS, PMI (Gelderman et al. [5])	-	1.67	See Equation (1)
TDS, ADD (Gelderman et al. [5])	-	1.62	See Equation (2)
TDS, PMI	0.23	0.89	$\text{PMI} = e^{(1.68+0.17\ \text{TDS})}$
TDS, ADD	0.45	0.81	$\text{ADD} = e^{(3.27+0.26\ \text{TDS})}$
TDS + demographics, PMI	0.25	0.87	$\text{PMI} = e^{(3.58+0.5\ \text{TDS}+0.14\ \text{BMI}+0.02\ \text{age}-0.17\ \text{sex})}$
TDS + demographics, ADD	0.47	0.78	$\text{ADD} = e^{(5.97+0.73\ \text{TDS}+0.19\ \text{BMI}+0.03\ \text{age}+0.04\ \text{sex})}$
TDS + demographics + weather, PMI	0.42	0.76	$\text{PMI} = e^{(3.79+0.63\ \text{TDS}+0.12\ \text{BMI}-0.01\ \text{age}-0.07\ \text{sex}+0.1\ \text{spring}-0.55\ \text{summer}-0.77\ \text{fall}\ -0.15\ \text{hum\_hist})}$
TDS + demographics + weather, ADD	0.54	0.73	$\text{ADD} = e^{(5.58+0.68\ \text{TDS}+0.17\ \text{BMI}+0.02\ \text{age}+0.06\ \text{sex}+0.54\ \text{spring}+0.62\ \text{summer}+0.16\ \text{fall}\ -0.12\ \text{hum\_hist})}$

formulae were used, which required applying the base 10 logarithm to both sides of (1) and (2). This resulted in an RMSE of 1.67 when estimating the PMI and an RMSE of 1.62 when estimating the ADD. Note, since adj. R^2 is specifically designed to assess the fit of a model to the data on which it was developed the adj. R^2 was not calculated here since these formulae were not fitted on our study sample. The remaining rows of Table 4 give the results for the univariate and multiple linear regression using different sets of input variables. For the univariate linear regression analysis using only the TDS as a predictor variable of PMI, an adj. R^2 of 0.23 and an RMSE of 0.89, and for ADD, an adj. R^2 of 0.45 and an RMSE of 0.81, were achieved. In both cases, the standardized coefficients were statistically significant with a p-value <0.05. For the multiple linear regression analysis using the TDS and demographic features as predictor variables of PMI, an adj. R^2 of 0.25 and an RMSE of 0.87, and for ADD, an adj. R^2 of 0.47 and an RMSE of 0.78, were achieved. In both cases, the standardized coefficients for BMI had a p-value <0.05, hence statistically significant, and a p-value ≥ 0.05 for sex and age, hence not statistically significant. Lastly, for the multiple linear regression analysis using the TDS, demographic, and weather features as predictor variables of PMI, an adj. R^2 of 0.42 and an RMSE of 0.76, and for ADD, an adj. R^2 of 0.54 and an RMSE of 0.73, were achieved. In both cases, the standardized coefficients of the weather features (i.e., spring, summer, fall, and hum_hist) and BMI had a p-value <0.05, hence statistically significant, and a p-value ≥ 0.05 for the sex and age, hence not statistically significant.

Figure 6 shows the predicted versus actual PMI scatter plots for the regression analyses with input variables: (a) TDS, (b) TDS + demographic features, (c) TDS + demographic features + weather features.

Similarly, Figure 7 shows the predicted versus actual ADD scatter plots for the regression analyses with input variables: (a) TDS, (b) TDS + demographic features, (c) TDS + demographic features + weather features. The red dashed line gives the reference line $x = y$ (i.e., actual = predicted) with a slope 1.

4 | DISCUSSION

The objective of this study was to investigate ways to improve outdoor decomposition-based PMI estimation from human remains. Specifically, pre-existing outdoor PMI and ADD formulae proposed by Gelderman et al. [5] were evaluated on the substantially larger sample size. Additionally, various univariate and multiple linear regression analyses using a variety of predictor variables to estimate the PMI and ADD were conducted and evaluated. Since these predictor variables were easily accessible, our approach has the potential to eliminate the need for difficult to obtain or inaccessible variables that do not improve predictive performance.

When evaluating Gelderman et al.'s [5] outdoor PMI and ADD formulae on our study sample, their performance was substantially worse relative to the models reported in this study. This is not surprising given that the formulae were developed from a very small sample size (i.e., 12 outdoor cases) compared to the 213 outdoor cases of this study. In fact, Gelderman et al.'s [5] PMI formula resulted in an RMSE that was 54% higher than the RMSE of the PMI model that considered the TDS, demographic, and weather-related features as predictor variables. Similarly, Gelderman et al.'s [5] ADD formula resulted in an RMSE that was 55% higher than the RMSE of the ADD model that considered the TDS, demographic, and weather-related features as predictor variables. As done by Megyesi et al. [2], the RMSE can be doubled to approximate the 95% prediction interval (i.e., instead of $\pm 1.96 \times \text{RMSE}$, $\pm 2 \times \text{RMSE}$ is used because $1.96 \approx 2$). For a predicted PMI of 10 days (naturally log-transformed PMI = 2.3) and ADD of 200 (naturally log-transformed ADD = 5.3), can be applied. For Gelderman et al.'s [5] PMI formula applied to our study sample, the 95% prediction interval ranges from 0 to 281 days, compared to 2 to 46 days for the best PMI formula in this study (i.e., the model incorporating TDS, demographic features, and weather features). Similarly, for Gelderman et al.'s [5] ADD formula applied to our sample, the 95% prediction interval ranges from 8 to 5115 ADDs, compared to 47 to 863 ADDs for the best ADD formula in this study (i.e.,

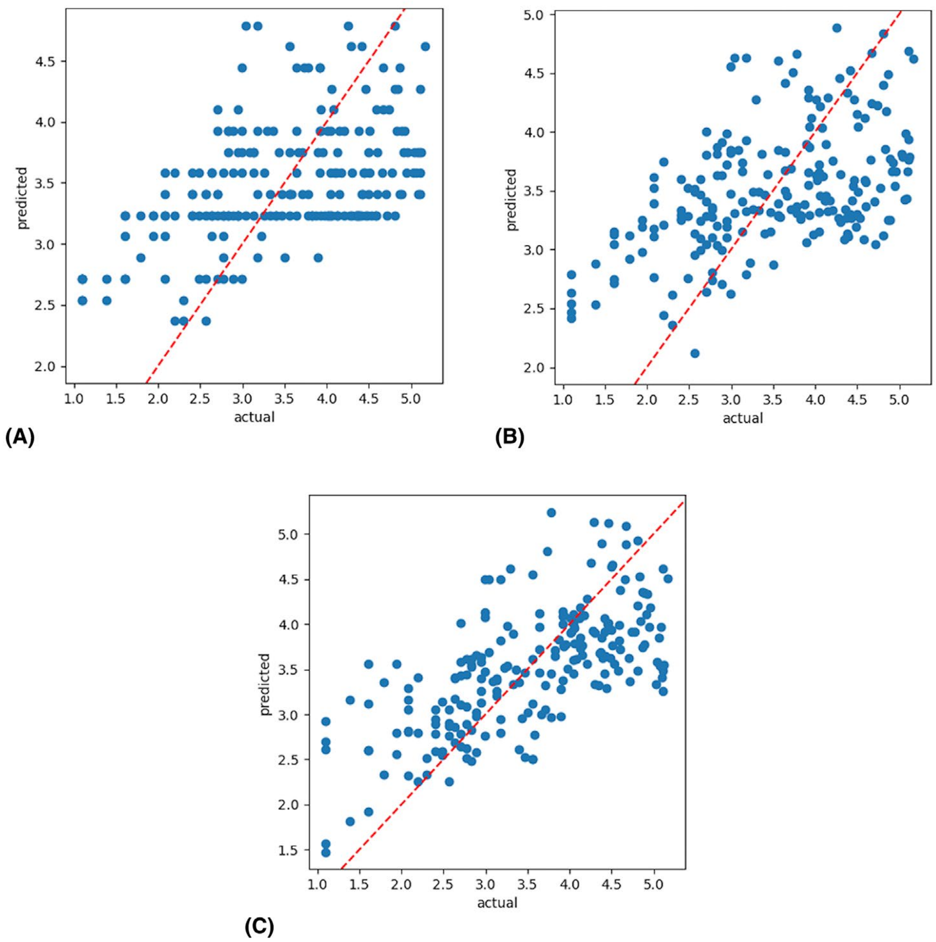

FIGURE 6 Predicted versus actual PMI scatter plots for the different regression analyses: (A) TDS, (B) TDS + demographic features, and (C) TDS + demographic features + weather features.

the model incorporating TDS, demographic features, and weather features). For both the PMI and ADD cases, the 95% prediction interval is significantly narrower using this study's best formulae. Models trained on small sample sizes, such as the outdoor formulae proposed by Gelderman et al.'s [5], typically have wider prediction intervals because (a) the model does not have enough information to accurately estimate the relationship between the input and outcome variable(s), leading to greater uncertainty in the predictions, (b) a smaller dataset can result in higher variance in the model's predictions, especially in regions of the input space where the model has less data, and (c) with small datasets, models may overfit and not generalize well to new data, resulting in unreliable predictions, as shown in this study [19]. On the other hand, a larger sample size provides the model with more accurate and stable regression coefficients, improves model generalizability, and allows for more confident predictions, which leads to narrower prediction intervals.

The univariate and multiple linear regression analysis results showed that including additional factors, such as demographic and weather-related information, in addition to the TDS, improves PMI and ADD prediction performance. When estimating the PMI, the adjusted R^2 was 45% higher and the RMSE was 15% lower when using the TDS, demographic, and weather-related features compared to only using the TDS as predictor variables. Similarly, when estimating the ADD, the adjusted R^2 was 17% higher and the RMSE was 10% lower when using the TDS, demographic, and weather-related features compared to only using the TDS as predictor variables. This increase in prediction performance is also seen in the predicted versus actual scatter plots (see Figures 6 and 7) as more predictor variables are considered, the points are closer to the red-dashed line (reference line), indicating a better fit and more accurate predictions. When assessing variable importance, the standardized coefficient of the TDS was statistically significant for the TDS-only PMI and ADD models. When considering the demographic features (i.e., age, sex, and BMI) in addition to the TDS, the standardized coefficients for TDS and BMI were statistically significant, but not for the age and sex variables when predicting the PMI and ADD. Lastly, when also considering the weather-related features (i.e., season of discovery and humidity history), the TDS, BMI, and all weather-related features had statistically significant standardized coefficients, but not the age and sex. This suggests that the

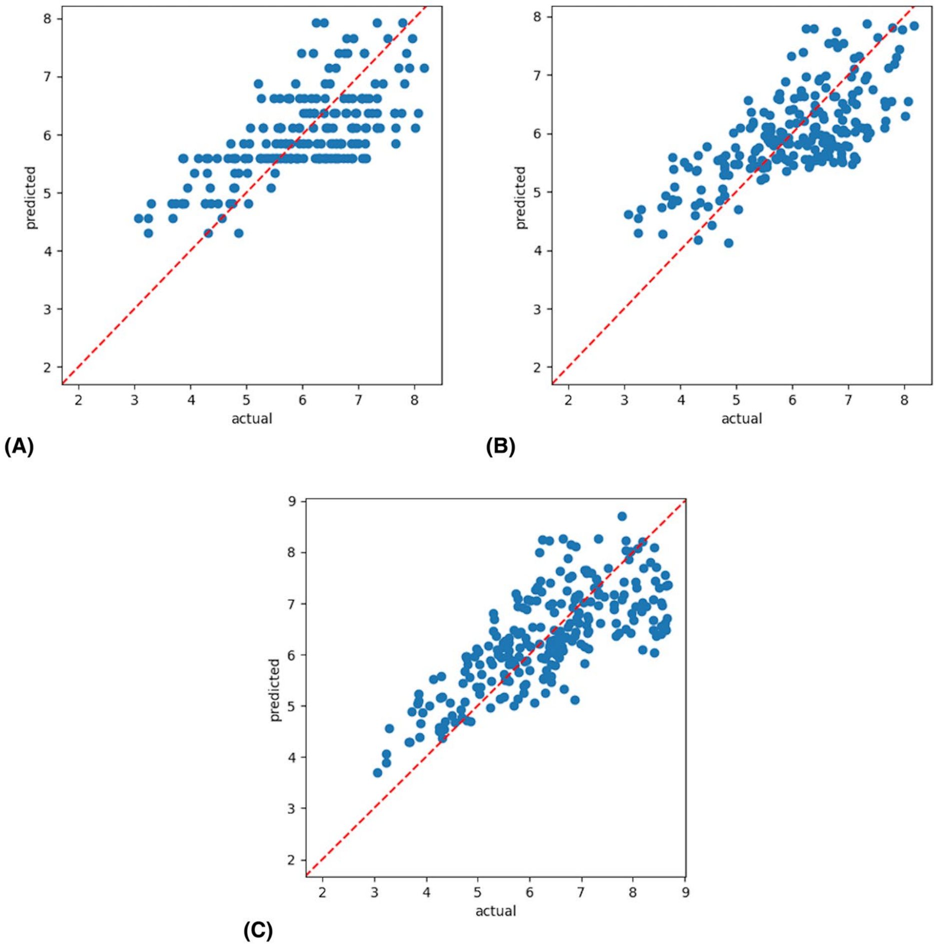

FIGURE 7 Predicted versus actual ADD scatter plots for the different regression analyses with input variables: (A) TDS, (B) TDS + demographic features, and (C) TDS + demographic features + weather features.

demographic features, sex and age, may not have a significant effect on the dependent variables, PMI and ADD, and could thus be removed from the model.

While the methods described in this research build upon and provide improvements to previous methods, some limitations apply. Our study is environmental- and climate-specific. All study subjects decomposed outdoors in an open-wooded decomposition environment within a humid subtropical climate, characterized by hot summers and moderate winters. As a result, instances of decomposition in extremely dry or cold environments are not represented. Furthermore, factors such as soil quality, the extent of scavenging (e.g., fenced enclosures or nets may have prevented some scavenging and scattering), and the absence of clothing may not represent all types of real-world conditions. Our weather data was retrieved from a station approximately 10 miles away, and microclimatic variations could affect the calculation of ADD, and consequently, the fit of our models. Also, the BMI was calculated based on the donor's height and weight at the time of intake, which may differ from the pre-mortem BMI as in, for example, medical records, though it better reflects the actual BMI at the time of decomposition. Finally, TDS scoring of the study sample was conducted by a single forensic expert, potentially introducing labeling bias and errors [20, 21]. However, this concern is mediated by Gelderman et al. [5], who found high inter-observer agreement. Moreover, our forensic expert had a Fleiss' Kappa inter-observer agreement of 0.6 (moderate-to-substantial) with two other experts when scoring TDS on photos similar to those used in this study.

Finally, the PMI and ADD formulae reported in this study should be used with caution. The aim of this work was not to create new PMI and ADD estimation formulae but to investigate ways to improve decomposition-based PMI estimation methods, such as the one proposed by Gelderman et al. [5], by using a larger sample size and including different factors known to affect the human decay process. Researchers are encouraged to evaluate and build upon our method to further assess its accuracy. Other existing decomposition-based PMI estimation formulae, such as those developed by Megyesi et al. [2] and Moffat et al. [22], could also be modified and evaluated using additional variables as proposed by this work.

As a future work, many of the limitations could be addressed by creating a benchmark dataset that includes images from different climates and from actual casework, recordings of microclimate temperature, humidity, and soil acidity as well as using multiple raters to score TDS. Given the limited accuracy observed even with improved models, future research should also explore alternative predictors of PMI, such as entomological or microbial data. To enable reproducibility of our results, we share the data used in this study as supplemental information.

5 | CONCLUSION

In conclusion, this study underscores the importance of leveraging larger datasets and incorporating a broader range of predictive factors to enhance PMI and ADD estimation models. By integrating demographic and environmental variables with the TDS, the accuracy of PMI and ADD predictions compared to traditional models improved. The results suggest that a more comprehensive approach—one that accounts for various environmental and biological factors such as age, biological sex, BMI, season, temperature, and humidity—can better capture the complexity of human decomposition. Although sex and BMI were not significant in our study, they remain biologically relevant factors that may influence decomposition under different conditions and should be further investigated in future research. Nevertheless, this study not only highlights the potential for more accurate postmortem interval assessments but also lays the groundwork for future research to refine and validate these models.

FUNDING INFORMATION

The work has been supported by the grants 2018-DU-BX-0181 and 15PNIJ-21-GG-04161-SLFO from the National Institute of Justice to Dr. Audris Mockus and Dr. Dawnie Wolfe Steadman.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

ORCID

Anna-Maria Nau https://orcid.org/0000-0002-6319-5947
Dawnie Wolfe Steadman https://orcid.org/0000-0003-0812-0739

REFERENCES

1. Galloway A, Birkby WH, Jones AM, Henry TE, Parks BO. Decay rates of human remains in an arid environment. J Forensic Sci. 1989;34(3):607–16. https://doi.org/10.1520/JFS12680J
2. Megyesi MS, Nawrocki SP, Haskell NH. Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. J Forensic Sci. 2005;50(3):618–26. https://doi.org/10.1520/JFS2004017
3. Vass AA. The elusive universal post-mortem interval formula. Forensic Sci Int. 2011;204(1–3):34–40. https://doi.org/10.1016/j.forsciint.2010.04.052
4. Rodriguez WC, Bass WM. Insect activity and its relationship to decay rates of human cadavers in East Tennessee. J Forensic Sci. 1983;28(2):423–32.
5. Gelderman H, Boer L, Naujocks T, IJzermans A, Duijst W. The development of a post-mortem interval estimation for human remains found on land in the Netherlands. Int J Legal Med. 2018;132(3):863–73. https://doi.org/10.1007/s00414-017-1700-9
6. Simmons T, Adlam RE, Moffatt C. Debugging decomposition data – comparative taphonomic studies and the influence of insects and carcass size on decomposition rate. J Forensic Sci. 2010;55(1):8–13. https://doi.org/10.1111/j.1556-4029.2009.01206.x
7. Matuszewski S, Konwerski S, Frątczak K, Szafałowicz M. Effect of body mass and clothing on decomposition of pig carcasses. Int J Leg Med. 2014;128:1039–48. https://doi.org/10.1007/s00414-014-0965-5
8. Sutherland A, Myburgh J, Steyn M, Becker PJ. The effect of body size on the rate of decomposition in a temperate region of South Africa. Forensic Sci Int. 2013;231(1–3):257–62. https://doi.org/10.1016/j.forsciint.2013.05.035
9. Mason AR, McKee-Zech HS, Hoeland KM, Davis MC, Campagna SR, Steadman DW, et al. Body mass index (BMI) impacts soil chemical and microbial response to human decomposition. mMsphere. 2022;7(5):e00325-22. https://doi.org/10.1128/msphere.00325-22
10. Johnson AP, Mikac KM, Wallman JF. Thermogenesis in decomposing carcasses. Forensic Sci Int. 2013;231(1–3):271–7. https://doi.org/10.1016/j.forsciint.2013.05.031
11. Cockle DL, Bell LS. Human decomposition and the reliability of a 'universal' model for post mortem interval estimations. Forensic Sci Int. 2015;253:136–9. https://doi.org/10.1016/j.forsciint.2015.05.018
12. Payne JA. A summer carrion study of the baby pig *Sus scrofa* Linnaeus. Ecology. 1965;46(5):592–602. https://doi.org/10.2307/1934999
13. Faber J, Fonseca LM. How sample size influences research outcomes. Dental Press J Orthod. 2014;19(4):27–9. https://doi.org/10.1590/2176-9451.19.4.027-029.ebo
14. Sullivan LM, Weinberg J, Keaney JF. Common statistical pitfalls in basic science research. J Am Heart Assoc. 2016;5(10):e004142. https://doi.org/10.1161/JAHA.116.004142
15. van Daalen MA, de Kat DS, Oude Grotebevelsborg BF, de Leeuwe R, Warnaar J, Oostra RJ, et al. An aquatic decomposition scoring method to potentially predict the postmortem submersion interval of bodies recovered from the North Sea. J Forensic Sci. 2017;62(2):369–73. https://doi.org/10.1111/1556-4029.13258
16. Nau AM, Mousavi S, Lee D, Hossain R, Griffin T, Steadman DW, et al. ICPUTRD: image cloud platform for use in tagging and research on decomposition. J Forensic Sci. 2024;69(5):1671–80. https://doi.org/10.1111/1556-4029.15448
17. Srivastava AK, Srivastava VK, Ullah A. The coefficient of determination and its adjusted version in linear regression models. Econom Rev. 1995;14(2):229–40. https://doi.org/10.1080/07474939508800317
18. Hodson TO. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci Model Dev. 2022;15(14):5481–7. https://doi.org/10.5194/gmd-15-5481-2022
19. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open. 2016;6(7):e010247. https://doi.org/10.1136/bmjopen-2015-010247
20. Jiang H, Nachum O. Identifying and correcting label bias in machine learning. In: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics; 2020 Aug 26–28; held virtually. Philadelphia, PA: mlr Press; 2020. p. 702–12. https://doi.org/10.48550/arXiv.1901.04966
21. Nau A-M, Ditto P, Steadman DW, Mockus A. Towards automation of human stage of decay identification: an artificial intelligence approach. arXiv preprint. 2024. https://doi.org/10.48550/arXiv.2408.10414
22. Moffatt C, Simmons T, Lynch-Aird J. An improved equation for TBS and ADD: establishing a reliable postmortem interval

framework for casework and experimental studies. J Forensic Sci. 2016;61(S1):S201–S207. https://doi.org/10.1111/1556-4029.12931

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nau A-M, Ditto P, Steadman DW, Mockus A. Identifying factors that help improve existing decomposition-based PMI estimation methods. J Forensic Sci. 2025;70:1249–60. https://doi.org/10.1111/1556-4029.70046