ORIGINAL ARTICLE

Variations in cuticular hydrocarbons of *Calliphora vicina* (Diptera: Calliphoridae) empty puparia: Insights for estimating late postmortem intervals

Swaima Sharif[1,2] · Cora Wunder[1,3] · Jens Amendt[1] · Ayesha Qamar[2]

Abstract

Necrophagous flies, particularly blowflies, serve as vital indicators in forensic entomology and ecological studies, contributing to minimum postmortem interval estimations and environmental monitoring. The study investigates variations in the predominant cuticular hydrocarbons (CHCs) viz. n-C25, n-C27, n-C28, and n-C29 of empty puparia of *Calliphora vicina* Robineau-Desvoidy, 1830, (Diptera: Calliphoridae) across diverse environmental conditions, including burial, above-ground and indoor settings, over 90 days. Notable trends include a significant decrease in n-C25 concentrations in buried and above-ground conditions over time, while n-C27 concentrations decline in buried and above-ground conditions but remain stable indoors. Burial conditions show significant declines in n-C27 and n-C29 concentrations over time, indicating environmental influences. Conversely, above-ground conditions exhibit uniform declines in all hydrocarbons. Indoor conditions remain relatively stable, with weak correlations between weathering time and CHC concentrations. Additionally, machine learning techniques, specifically Extreme Gradient Boosting (XGBoost), are employed for age estimation of empty puparia, yielding accurate predictions across different outdoor and indoor conditions. These findings highlight the subtle responses of CHC profiles to environmental stimuli, underscoring the importance of considering environmental factors in forensic entomology and ecological research. The study advances the understanding of insect remnant degradation processes and their forensic implications. Furthermore, integrating machine learning with entomological expertise offers standardized methodologies for age determination, enhancing the reliability of entomological evidence in legal contexts and paving the way for future research and development.

Keywords Cuticular hydrocarbons · GC-MS · Age estimation · Forensic entomology · XGBoost

✉ Ayesha Qamar
ayesha.zoology@gmail.com

Swaima Sharif
swaimasharif@gmail.com

Cora Wunder
wunder@uni-mainz.de

Jens Amendt
amendt@em.uni-frankfurt.de

1   Institute of Legal Medicine, Forensic Biology, University Hospital, Goethe University, Frankfurt am Main, Germany

2   Present Address: Section of Entomology, Department of Zoology, Aligarh Muslim University, Aligarh 202002, U.P, India

3   Present Address: Institute of Legal Medicine, Johannes Gutenberg University Medical Center, Am Pulverturm 3, 55131 Mainz, Germany

Introduction

Insect physiology encompasses a broad array of topics, ranging from the intricacies of metabolic pathways to the physiological adaptations that enable survival in diverse environments. Among the myriad physiological processes studied within this discipline, the composition and temporal variation of cuticular hydrocarbons (CHCs) in insects have garnered significant attention [1–5]. CHCs, comprising a complex mixture of long-chain hydrocarbons and their derivatives, are vital components of the insect cuticle, playing pivotal roles in various ecological interactions, such as communication, reproduction, and defense [5–7].

The Calliphoridae family, commonly known as blow flies, comprises a diverse group of insects with considerable ecological and forensic significance. Among these, *Calliphora vicina* holds particular interest due to its widespread distribution and close association with carrion decomposition [8–10]. The development stages of *C. vicina* serve as

a critical temporal marker in forensic entomology, aiding in estimating minimum postmortem intervals (PMI min) in forensic investigations. As blow flies progress through their life cycle, transitioning from larval stages to puparia and ultimately emerging as adult flies, the puparia retain residual hydrocarbons on their cuticles. Understanding the temporal dynamics of CHCs in *C. vicina* puparia holds excellent potential for forensic applications.

In the late stages of decomposition, when the body has reached the skeletal stage, and no soft tissues or dead insects are available, empty puparia remain significant forensic evidence [11, 12]. These puparia can be found in various locations, such as on the body, beneath the body, or buried in soil [13, 14]. This distribution occurs as insects migrate away from the original feeding source to find a suitable place for pupation. Therefore, empty puparia have a high potential to be discovered as evidence in forensic investigations [15]. Puparia play a vital role in forensic entomology and ecological research by providing insights into decomposition timelines, species identification, and environmental processes [12, 16–19]. Studies on insect stages other than empty puparia are abundant [18, 20–24]. Yet, research focusing specifically on empty puparia of blow flies, specifically in field conditions, as crucial evidence during advanced stages of decomposition is relatively limited in comparison [18, 25, 26]. Addressing the limitations and knowledge gaps associated with their use requires interdisciplinary collaboration, methodological advancement, and a detailed understanding of insect biology and ecosystem dynamics. Overcoming these challenges will enhance the reliability and applicability of puparial analysis in both forensic science and ecological studies, contributing to advancements in understanding environmental health and forensic investigation methodologies.

Given the interdisciplinary nature of CHCs in *C. vicina* puparia, this study seeks to integrate principles from insect physiology, ecology, and forensic entomology to shed light on the temporal variation of CHCs in a species of significant ecological and forensic importance. Prior research has emphasized that more than half of the lipids from blowfly puparia comprise alkanes falling within the n-C25 to n-C30 range [27]. In this context, our investigation specifically focused on the analysis of CHCs in empty blowfly puparia, with a particular emphasis on five predominant CHCs - n-Pentacosane (n-C25), n-Hexacosane (n-C26), n-Heptacosane (n-C27), n-Octacosane (n-C28), and n-Nonacosane (n-C29) [2, 4, 28]. These CHCs are chosen for their recognized stability and lower volatility when compared to smaller chain CHCs. This study employs the machine learning model XGBoost for age estimation of empty puparia based on hydrocarbon concentrations, connecting baseline research with the practical utilization of blowfly puparia in environmental forensics and ecological studies.

Materials and experiment design

The rearing process followed the protocol outlined in our previous study [4] 300 larvae were divided into three groups of 100 larvae each. These larvae were placed in plastic cups with 20 g of minced pork meat and transferred to plastic boxes filled with sawdust to aid pupation (refer to Fig S1 in Supplementary file). Empty puparia were collected from these three groups and placed in three conditions: buried outdoors, placed above-ground outdoors, and stored indoors (Fig S2). For the outdoor buried condition, ten empty puparia were buried approximately 5 inches deep in nylon stockings at the Institute of Legal Medicine in Frankfurt am Main, Germany. The study was conducted from late October to early February, encompassing the winter season. Sampling occurred on days 1, 28, 56, and 90, with two empty puparia collected per time point per replication. Similarly, above-ground outdoor and indoor conditions each involved the same sampling protocol. Samples were prepared by submerging two puparial cases in n-hexane and methadone-d9 (used as internal standard), followed by ultrasonication, drying with nitrogen air, and GC-MS analysis with hexane reconstitution and blank runs for carryover prevention as detailed in our previous study [28].

CHC profile analysis

Chemical analysis was conducted using an Agilent GC-MS system comprising a 7693 GC and a 7890 B MSD, equipped with a split/splitless injector in splitless mode. Samples, injected with 1 μl at 250 °C, were separated on an Agilent VG-1 ms capillary column (30 m × 250 μm I.D. × 0.25 μm film thickness). The temperature program started with a 2-minute hold at 100 °C, followed by ramps to 200 °C at 25 °C/min, 260 °C at 3 °C/min, and 320 °C at 20 °C/min, holding for 2 min. Helium flowed at 1.2 ml/min. The 5977 B MSD operated in positive ion mode (70 eV), using Selected Ion Monitoring (SIM) and scan modes (m/z 45–600) from 4 min into the run. Quantification of n-C25 to n-C29 compounds utilized SIM, with specific ions (352, 366, 380, 394, 408) and methadone-d9 (303) as internal standard. Calibration curves (0.4–50 ng/μl) were constructed with Agilent Chemstation, yielding regression coefficients > 0.994 (S3). Regular calibrations and hexane blanks ensured analysis accuracy and reliability.

Statistical analysis

Statistical analysis encompassed One-way ANOVA followed by Dunnett's test to identify significant variations in hydrocarbon means across different age groups of empty puparia. To track changes in CHC concentrations over time, normalization involved dividing initial day one concentrations by subsequent

time points and expressing results as percentages. A correlation heatmap in IBM SPSS version 25 investigated relationships between compound concentrations and weathering time alongside canonical discriminant analysis (CDA) and regression analysis. Following these analyses, for practical utility in age prediction of empty puparia, we employed machine learning technique, specifically eXtreme Gradient Boosting (XGBoost). Recognizing that simple regression analysis and CDA may not directly serve this purpose, we utilized XGBoost to develop a predictive model relevant to age determination.

Machine learning via XGBoost modeled the age prediction of puparia using n-C25 to n-C29 concentrations after log transformation (because of heteroscedasticity in data) with the 'xgboost' library in R. 70% of the data was allocated for model training and the remaining 30% solely for testing the model's generalizability to unseen data points as test data. A robust 5-fold cross-validation approach was applied, dividing the dataset into five subsets to facilitate accurate performance assessment. The training set was used for parameter tuning via 5-fold cross-validation, where we explored a grid of parameters including the number of boosting rounds ('nrounds'), maximum tree depth ('max_depth'), and learning rate ('eta'). During tuning, regularization parameters such as 'gamma', 'colsample_bytree,' 'min_child_weigh't, and 'subsample' were fixed [29]. The selection of hyperparameters aimed at minimizing the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Below is the R script for this method.

```
# Load the required libraries

library(xgboost)

library(caret)

library(ggplot2)

# Read  data (assuming it's in a data frame named 'data')

# Replace 'data.csv' with the actual path to the data file

data <- read.csv("data.csv")

# Separate the target variable (age) and the features (concentrations)

target <- data$Age

features <- data[, -1]


# Apply logp(1+x) transformation to the features

transformed_features <- log1p(features)


# Combine the transformed features and the target back into a data frame

transformed_data <- data.frame(Age = target, transformed_features)


# Convert data to DMatrix format (required for XGBoost)

data_matrix    <-    xgb.DMatrix(data    =    as.matrix(transformed_data[,    -1]),    label    =
transformed_data$Age)
```

```
# Split the data into training and testing sets (70% training, 30% testing)

set.seed(123)

splitIndex <- createDataPartition(transformed_data$Age, p = 0.7, list = FALSE)

train_data <- transformed_data[splitIndex, ]

test_data <- transformed_data[-splitIndex, ]

tuneGrid <- expand.grid(

  .nrounds = c(50, 100, 150),          # Number of boosting rounds

  .max_depth = c(3, 6, 9),             # Maximum depth of trees

  .eta = c(0.01, 0.1, 0.3),            # Learning rate

  .gamma = 0,                          # Regularization parameter

  .colsample_bytree = 1,               # Column subsampling ratio for each tree

  .min_child_weight = 1,               # Minimum sum of instance weight (Hessian) needed in a child

  .subsample = 1                       # Subsample ratio of the training instance

)

# Perform cross-validation

xgb_cv <- train(

  x = as.matrix(train_data[, -1]),

  y = train_data$Age,

  trControl = trainControl(method = "cv", number = 5),  # 5-fold cross-validation

  tuneGrid = tuneGrid,                          # Use the specified tuning grid

  method = "xgbTree",

  metric = "RMSE"

)
```

```
# Access the best parameters from cross-validation

best_params <- xgb_cv$bestTune

# Convert best_params to a list

best_params_list <- list(

  objective = "reg:squarederror",  # Regression task

  max_depth = best_params$max_depth,

  nrounds = best_params$nrounds,

  eta = best_params$eta,

  gamma = best_params$gamma,

  colsample_bytree = best_params$colsample_bytree,

  min_child_weight = best_params$min_child_weight,

  subsample = best_params$subsample

)

# Train the final XGBoost model using the best parameters

final_xgb_model  <-  xgb.train(params  =  best_params_list, data  =  data_matrix, nrounds  =

best_params$nrounds)

# Make predictions on test data

test_matrix <- xgb.DMatrix(data = as.matrix(test_data[, -1]), label = test_data$Age)

test_predictions <- predict(final_xgb_model, test_matrix)

# Calculate RMSE and MAE for the test set

test_rmse <- sqrt(mean((test_data$Age - test_predictions)^2))

test_mae <- mean(abs(test_data$Age - test_predictions))
```

```
# Calculate TSS and RSS for test data

test_actual_mean <- mean(test_data$Age)

test_tss <- sum((test_data$Age - test_actual_mean)^2)

test_rss <- sum((test_data$Age - test_predictions)^2)


# Calculate R-squared for test data

test_r_squared <- 1 - (test_rss / test_tss)


# Print the R-squared values

cat("R-squared value for training data:", round(train_r_squared, 4), "\n")

cat("R-squared value for test data:", round(test_r_squared, 4), "\n")


# Create a scatter plot for the training set

train_plot <- ggplot(data = train_data, aes(x = Age, y = train_predictions)) +

 geom_point() +

 geom_abline(intercept = 0, slope = 1, linetype = "dashed") +

 labs(x = "Actual Age", y = "Predicted Age", title = "Training Set") +

 annotate("text", x = min(train_data$Age), y = max(train_predictions),

      label = paste("RMSE:", round(train_rmse, 2), "\nMAE:", round(train_mae, 2)),

      hjust = 0, vjust = 1) +  # Adjust hjust to 0 for left alignment

 theme_minimal()
```

```
# Display the training plot

print(train_plot)

# Create a scatter plot for the test set

test_plot <- ggplot(data = test_data, aes(x = Age, y = test_predictions)) +

  geom_point() +

  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +

  labs(x = "Actual Age", y = "Predicted Age", title = "Test Set") +

  annotate("text", x = min(test_data$Age), y = max(test_predictions),

        label = paste("RMSE:", round(test_rmse, 2), "\nMAE:", round(test_mae, 2)),

        hjust = 0, vjust = 1) +  # Adjust hjust to 0 for left alignment

  theme_minimal()


# Display the test plot

print(test_plot)
```

Results

The study explored variations in the CHCs of empty puparia of *C. vicina* across different natural environmental conditions for 90 days. The chromatograms depicting these conditions are presented in Fig. 1. Examination of hydrocarbons at different time intervals revealed a consistent highest concentration of n-C27 hydrocarbon in all three conditions, as detailed in Table 1. N-C26 is omitted in the analysis due to concentrations below 2%. Subsequent to n-C27, the concentrations of hydrocarbons followed a descending order: n-C29, n-C28, and n-C25. From the data logger readings, the highest temperature observed outdoors in buried conditions reached 12.5 °C, while the lowest dropped to 2 °C.

Conversely, outdoor above-ground conditions experienced a maximum of 10 °C and a minimum of -3 °C. Indoors, temperatures ranged from a maximum of 21.5 °C to a minimum of 16 °C (Fig S4). Key trends include a significant decrease in n-C25 concentrations in buried and above-ground settings over time. n-C27 concentrations consistently decline in buried and above-ground conditions while remaining stable indoors. n-C28 concentrations decrease in above-ground and indoor settings. n-C29 concentrations decrease in buried and above-ground conditions, with a slight, statistically insignificant increase indoors.

In the buried condition, n-C27 and n-C29 exhibited a significant ($p < 0.0001$) declining trend over the observation period, indicating decreased concentrations from Day 1 to Month 3. However, no significant changes were observed in the concentrations of n-C25 ($p = 0.419$) and n-C28 ($p = 0.223$) during the same period. This suggests that the burial environment may selectively influence certain hydrocarbons in empty puparia of *C. vicina*, with n-C27 and n-C29 showing a more pronounced response (Fig. 2A).

Conversely, all hydrocarbons, including n-C27, n-C28, n-C25, and n-C29, displayed a significant decline in the above-ground condition. Concentrations decreased notably from Day 1 to Month 3. This uniform decline suggests a consistent environmental influence on the hydrocarbon composition of puparia when placed above-ground (Fig. 2B).

In the indoor condition, no significant ($p > 0.05$) changes were observed in the concentrations of the analyzed hydrocarbons over time. This relative stability suggests that the indoor environment has a less discernible impact on the hydrocarbon composition of empty puparia compared to the outdoor conditions (Fig. 2C). The distinct trends observed in different environmental settings highlight the nuanced responses of *C. vicina* puparia to varying conditions, providing valuable insights into the environmental factors shaping their hydrocarbon profiles.

Fig. 1 Chromatograms representing the cuticular hydrocarbons (CHCs) extracted from empty puparia of *Calliphora vicina* in three conditions i.e. (I) Buried, (II) Above-ground and (III) Indoor showing peaks 1 to 4 for n-C25, n-C27, n-C28, and n-C29, respectively
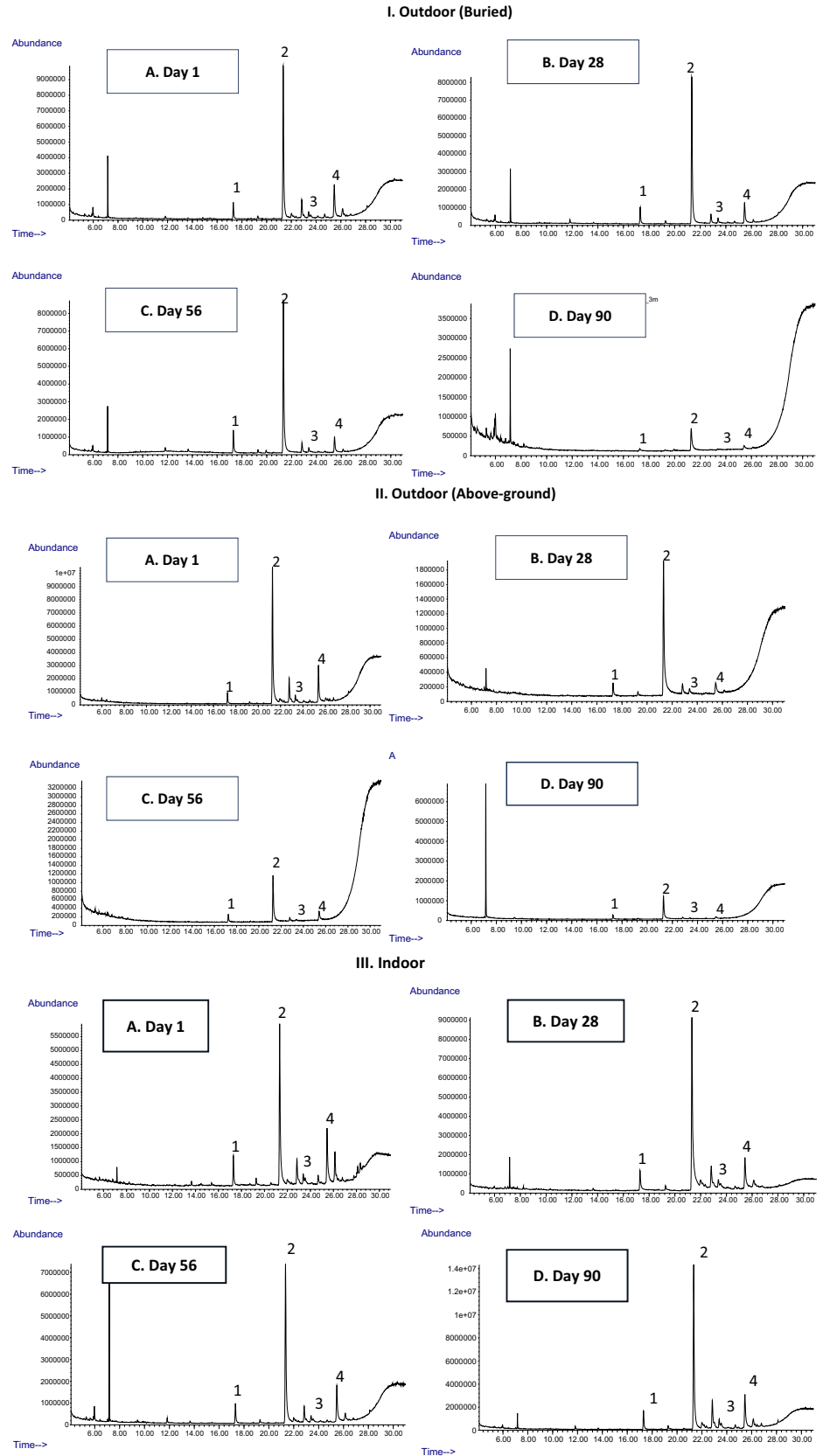
Table 1 Cuticular Hydrocarbons with their mean of concentrations (ng/μl) and standard deviation for each time point/Age (in days) for three environmental conditions

| CHC	Buried/Outdoor				Above ground/Outdoor				Indoor			
	Day 1 (n=15)	Week 4 (n=15)	Week 8 (n=15)	Month 3 (n=15)	Day 1 (n=15)	Week 4 (n=15)	Week 8 (n=15)	Month 3 (n=15)	Day 1 (n=15)	Week 4 (n=15)	Week 8 (n=15)	Month 3 (n=15)
n-C25	21.16±13.66	14.02±5.92	17.52±9.44	18.26±12.19	32.67±15.98	10.78±5.37	8.06±2.82	6.55±2.43	33.94±17.91	40.50±0.19.78	32.18±12.51	35.39±17.73
n-C27	309.86±119.62	169.08±57.74	23.83±7.30	15.36±6.25	279.66±123.67	268.11±175.27	204.20±77.27	78.11±25.71	351.93±215.16	383.29±138.29	360.28±114.09	292.16±126.12
n-C28	24.20±10.78	18.07±7.87	19.57±7.64	17.99±10.19	35.39±25.49	15.39±6.21	9.52±2.83	4.83±2.85	22.14±14.05	27.36±21.13	28.82±16.13	17.17±9.40
n-C29	148.97±34.99	156.08±41.16	70.72±24.96	43.62±18.65	137.12±28.52	127.72±28.06	78.42±14.90	29.65±12.95	81.84±52.81	89.67±40.60	99.84±46.79	80.48±55.91

The correlation analysis between weathering time and CHC concentrations in empty puparia buried in soil revealed intriguing relationships (Fig. 3A). Notably, a negligible correlation was observed between weathering time and n-C25 concentration ($r = -0.0471$), indicating minimal impact of time on this specific hydrocarbon. In contrast, a strong negative correlation was identified between weathering time and n-C27 concentration ($r = -0.8248$), suggesting a significant decrease in n-C27 levels over time. Similarly, a strong negative correlation ($r = -0.7778$) was found between weathering time and n-C29 concentration, indicating a pronounced reduction in n-C29 levels with increasing time. When exploring relationships between CHCs, weak to very weak correlations were identified. For instance, a weak positive correlation ($r = 0.1425$) was observed between n-C25 and n-C27 concentrations, while a moderate positive correlation ($r = 0.6451$) existed between n-C27 and n-C29 concentrations.

The correlation analysis between weathering time and CHC concentrations in empty puparia placed outdoors above ground yielded noteworthy insights into the dynamics of CHC changes over time in a natural environment. Moderate to strong negative correlation was observed between weathering time and each of the specific CHCs, namely n-C25 ($r = -0.6546$), n-C27 ($r = -0.8171$), n-C28 ($r = -0.6765$), and n-C29 ($r = -0.8733$). When examining the interplay between different CHCs, significant positive correlations were identified. For instance, a strong positive correlation ($r = 0.7606$) was observed between n-C25 and n-C27 concentrations, indicating a concurrent increase or decrease in both hydrocarbons over time. Similarly, moderate positive correlations were found between n-C25 and n-C28 ($r = 0.6164$), n-C27 and n-C28 ($r = 0.6301$), and n-C27 and n-C29 ($r = 0.6416$) (Fig. 3B). These correlations suggest potential synchronized variations in the concentrations of these CHCs during the weathering process.

In indoor conditions, weathering time shows a weak negative correlation with n-C25 ($r = -0.026$), indicating a slight tendency for the concentration of this hydrocarbon to decrease with increasing weathering time. Similarly, the correlations with n-C27 ($r = -0.157$) and n-C28 ($r = -0.051$) are negative but remain weak, suggesting only marginal decreases in their concentrations over time. The positive correlation between weathering time and n-C29 ($r = 0.123$) is also weak, signifying a slight increase in concentration with more extended weathering periods. However, the correlations between weathering time (days) and the concentrations of CHCs appear to be generally weak and insignificant. The correlation coefficients range from -0.026 to 0.366, suggesting subtle or limited associations between these variables (Fig. 3C).

In the CDA conducted to assess the discriminative potential of CHCs among four distinct age groups, the results revealed significant differentiation between the groups, with an overall accuracy

Fig. 2 Residual ratio (%) vs. Weathering time (days) plot. **A.** Outdoor (buried) Among the studied hydrocarbons, n-C7, and n-C29 displayed a decreasing trend after 28 days. **B** Outdoor (Above-ground) displaying decreasing trend over the time. **C** Indoor, where hydrocarbons showed stability
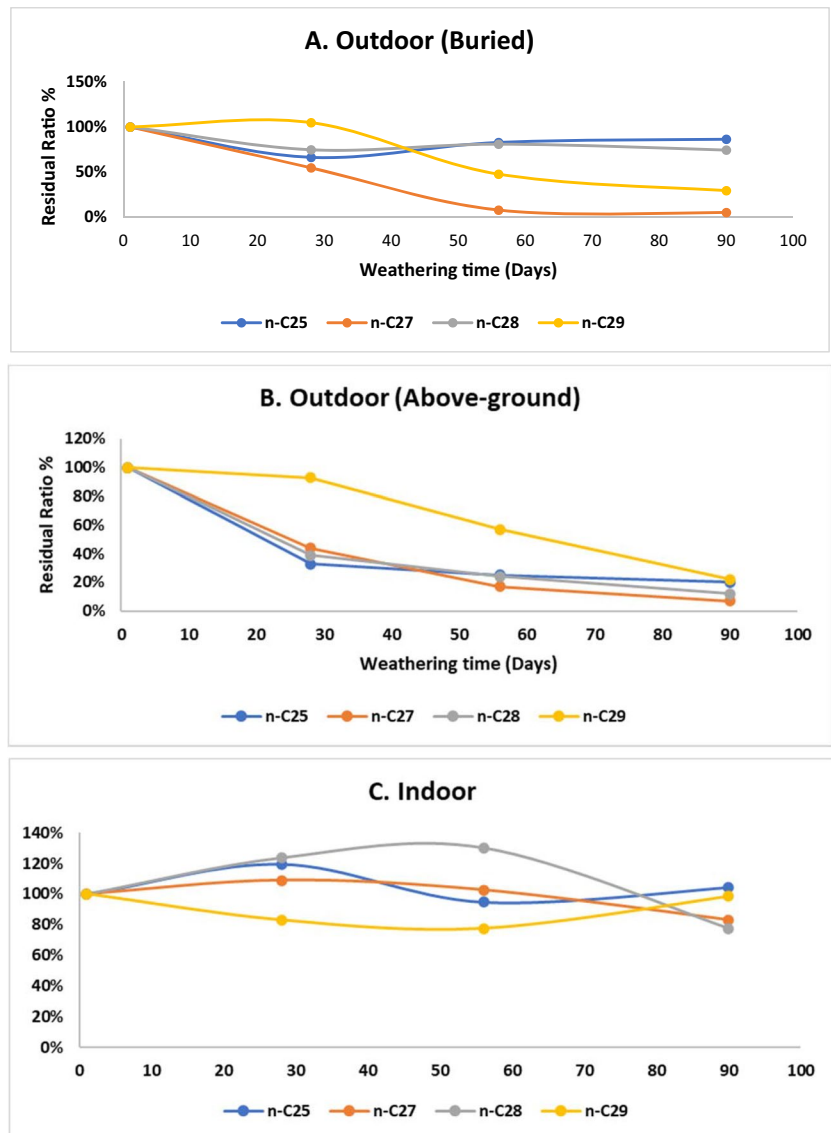
of 76.7%. Specifically, variables nC27 and nC29 emerged as crucial contributors to accurate group classification. The Eigenvalues provided further insight, with the first function explaining a substantial portion of the total variance (94.2%), indicating strong discriminatory power. However, in varying scenarios, such as in outdoor above-ground conditions, where the first function explained a remarkably high percentage (98.5%) of the total variance, and in indoor conditions, where it accounted for a relatively lower proportion (57.7%) of the cumulative variance, the discriminatory power varied, highlighting interpretations of the data across different contexts (for details refer to Fig. S5 in Supplementary information). Further, regression analyses conducted on three datasets—burial, above-ground, and indoor—revealed significant findings for CHCs in *Calliphora vicina* empty puparia. In the burial dataset, the model showed a strong correlation ($R = 0.888$, $R^2 = 0.789$) with nC27 and nC29 as significant predictors ($p < 0.001$). Similarly, the above-ground dataset exhibited a

robust model ($R = 0.947$, $R^2 = 0.896$) emphasizing the importance of nC29 ($p < 0.001$). Conversely, the indoor dataset demonstrated a weaker model ($R = 0.361$, $R^2 = 0.130$) with non-significant predictors overall ($p > 0.05$). Detailed results for each dataset can be found in the Supplementary file S6. To enhance precision in age estimation, the study transitioned to machine learning models, capitalizing on their effectiveness in predicting CHC dynamics.

Age estimation of empty puparia placed in outdoor buried using XGBoost model

The predictive model's performance in estimating the age of puparia based on hydrocarbon concentrations in buried conditions was evaluated and presented in Fig. 4. The model optimization was carried out using a robust 5-fold cross-validation methodology coupled with hyperparameter
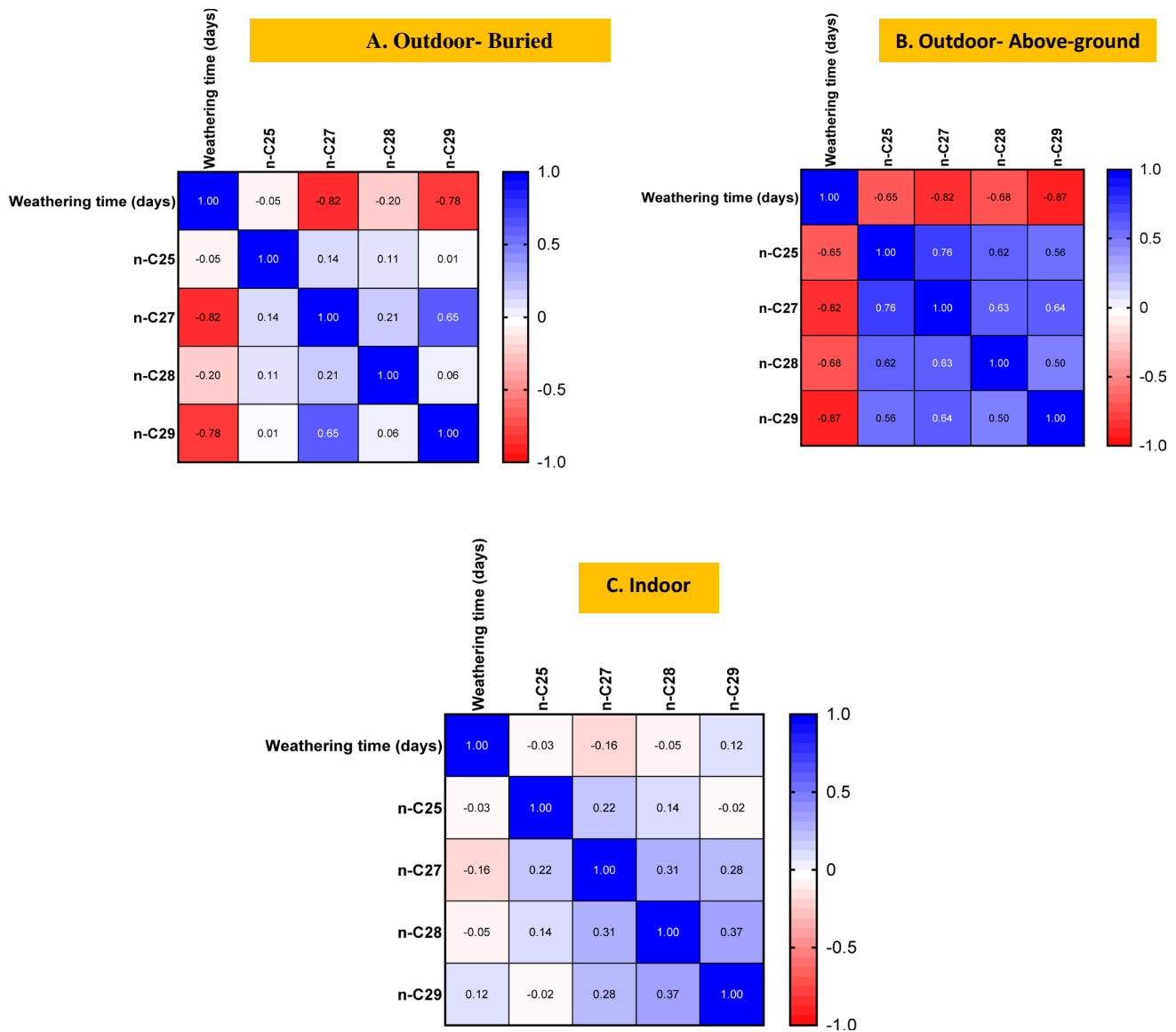
Fig. 3 Heatmap displaying the correlation matrix of CHCs and Weathering time (days). **A** Outdoor-buried. **B** Outdoor-Above-ground. **C** Indoor. Each cell in the heatmap represents the correlation coefficient between the corresponding CHCs and weathering time. Negative correlations are represented by red color, positive correlations by blue

tuning through a grid search approach. The optimized model parameters included Boosting Rounds set to 100, a Maximum Tree Depth of 6, and a Learning Rate of 0.1.

The cross-validation results demonstrated that the model achieved an average training RMSE of 5.36 and an MAE of 4.05. When applied to the test dataset, the model's performance was comparable, with an RMSE of 5.54 and an MAE of 3.57. These error metrics indicate the model's capability to predict the age of puparia with high precision.

Furthermore, the R-squared values underscore the model's accuracy and reliability. The training dataset yielded an R-squared value of 0.9737, while the test dataset exhibited a closely matching R-squared value of 0.9719. These high R-squared values reflect the model's strong predictive power and robustness in estimating the age of puparia under buried conditions.

Age estimation using XGBoost on outdoor (above-soil)

The optimized model parameters through grid search included Boosting Rounds set to 50, a Maximum Tree Depth of 3, and a Learning Rate of 0.3. The cross-validation results demonstrated that the model achieved an average training RMSE of 0.03. When applied to the test dataset, the model's performance was consistent, with an
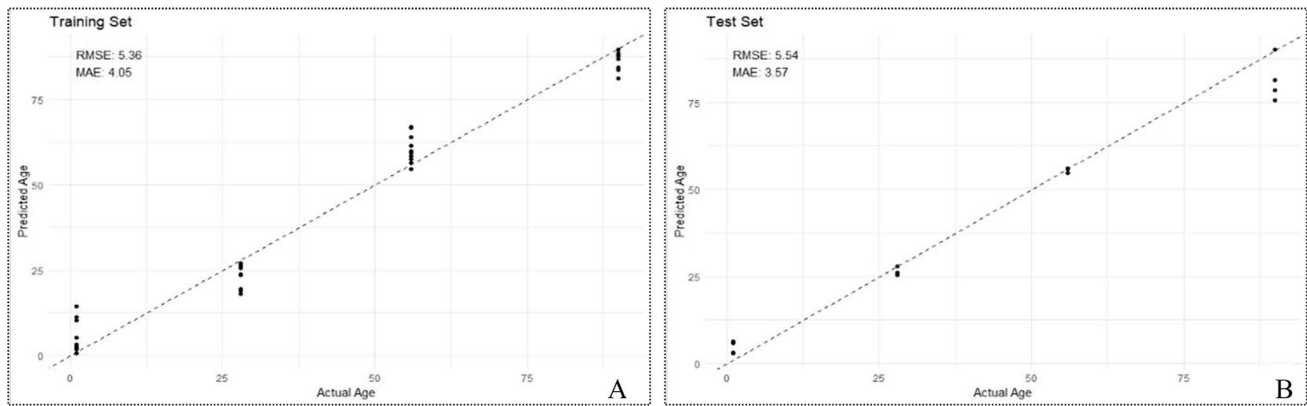
Fig. 4 Illustrating the predicted age versus the actual age of puparia based on hydrocarbon concentrations in buried conditions. The robust 5-fold cross-validation methodology and hyperparameter optimization through grid search led to optimized parameters (Boosting Rounds: 100, Maximum Tree Depth: 6, Learning Rate: 0.1). Cross-validation results showed an average training RMSE of 5.36 and MAE 4.05, while the model achieved an RMSE of 5.54 and an MAE of 3.57 on a test dataset. The R-squared values further emphasize the model's accuracy, with a value of 0.9737 for the training data (**A**) and 0.9719 for the test data (**B**)

RMSE of 0.05 and an MAE of 0.04. These error metrics indicate the model's capability to predict the age of puparia with very high precision.

Furthermore, the R-squared values underscore the model's accuracy and reliability. The training dataset yielded an R-squared value of 0.9999, while the test dataset exhibited an R-squared value of 0.999. These exceptionally high R-squared values reflect the model's strong predictive power and robustness in estimating the age of puparia under above-ground outdoor conditions (Fig. 5).

Age estimation of empty puparia placed indoor

The model was optimized with 50 Boosting Rounds, a Maximum Tree Depth of 6, and a Learning Rate of 0.1. The cross-validation results demonstrated an average training RMSE of 9.2. For the test dataset, the model achieved an RMSE of 10.32 and an MAE of 8.54.

The R-squared values further emphasize the model's accuracy, with the training data showing an R-squared value of 0.902 and the test data an R-squared value of 0.922. These high R-squared values also confirm the model's effectiveness and reliability in indoor conditions (Fig. 6).

Further results and details about the XGBoost model and interpretation are provided in Supplementary file S7.

Discussion

The investigation into the CHC profiles of empty puparia of *Calliphora vicina* across burial, above-ground, and indoor conditions over 90 days yielded valuable insights into the nuanced responses of these hydrocarbons to diverse environmental settings. Burial conditions revealed significant decreases in CHC concentrations, particularly notable for hydrocarbons such as n-C27 and n-C29. This decline suggests a pronounced response of CHCs to soil burial, likely driven by microbial degradation and chemical interactions with the soil matrix. Microbial activity in the soil, including the metabolism of organic compounds by bacteria and fungi, likely contributed to the degradation of longer-chain hydrocarbons [30–35]. Moreover, chemical processes such as oxidation and hydrolysis, facilitated by soil moisture and pH levels, may have further contributed to CHC degradation [36, 37].

In contrast, above-ground conditions exhibited uniform decreases in CHC concentrations across all hydrocarbons, indicating a consistent environmental influence on CHC degradation. Exposure to stressors like UV radiation, temperature fluctuations, and atmospheric pollutants likely accelerated hydrocarbon degradation processes. UV radiation, in particular, can induce photodegradation of CHCs, while temperature fluctuations and atmospheric pollutants can catalyze oxidative processes [3, 38–40]. Conversely, indoor conditions demonstrated relative stability in CHC concentrations over the observation period, suggesting a lesser impact of indoor environments on CHC degradation. Protection from environmental stressors like UV radiation and temperature fluctuations likely slowed down CHC degradation processes indoors. The weak correlations between weathering time and CHC concentrations in indoor conditions underscore the limited associations between environmental factors and CHC dynamics indoors. Understanding these influences is pivotal for accurate age estimation and forensic applications in entomology, and further research into the specific mechanisms driving CHC degradation in
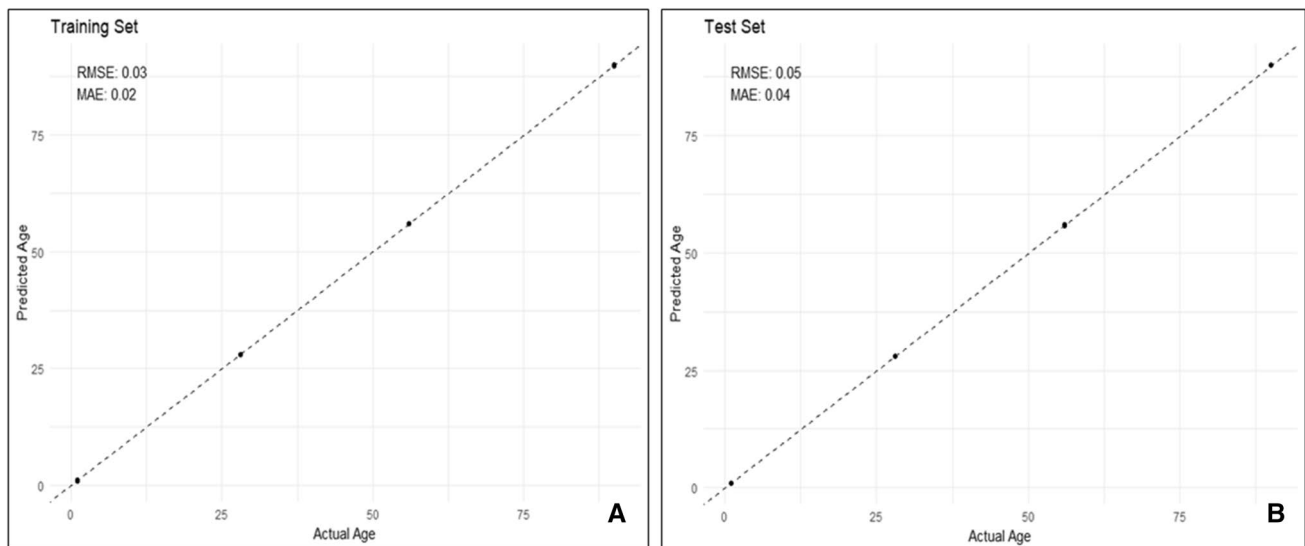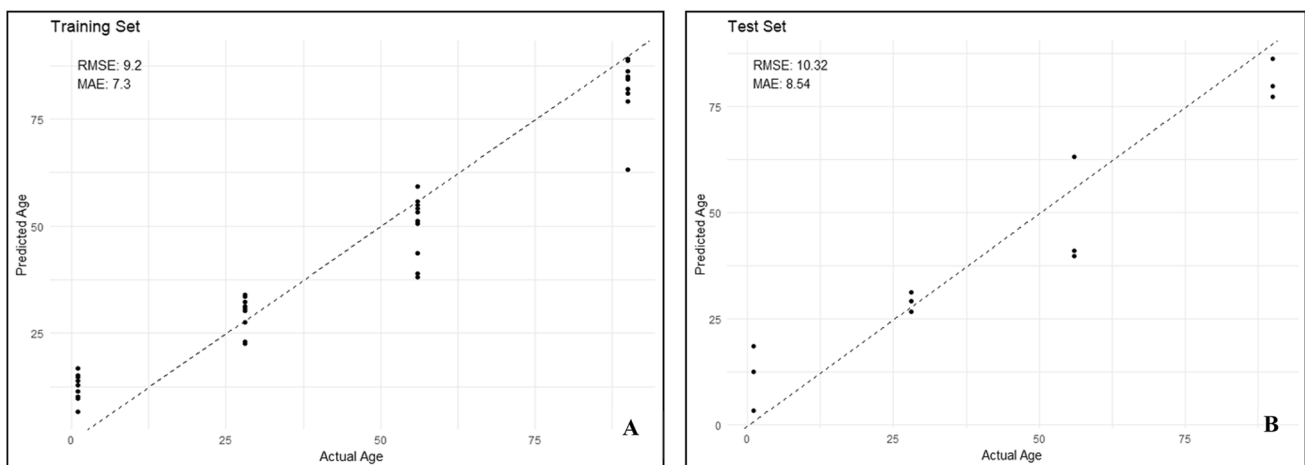
Fig. 5 The predicted age versus the actual age of puparia based on hydrocarbon concentrations in above-ground outdoor conditions with key insights into the model's performance. Cross-validation results (Boosting Rounds: 50, Maximum Tree Depth: 3, Learning Rate: 0.3) showed an average training RMSE of 0.03, while the model achieved an RMSE of 0.05 and a mean absolute error (MAE) of 0.04 on a test dataset. The R-squared values further emphasize the model's accuracy, with a value of 0.9999 for the training data (**A**) and 0.999 for the test data (**B**)

Fig. 6 Illustrating the predicted age versus the actual age of puparia based on hydrocarbon concentrations with key insights into the model's performance in Indoor conditions. Cross-validation results (Boosting Rounds: 50, Maximum Tree Depth: 6, Learning Rate: 0.1) showed an average training RMSE of 9.2, while the model achieved an RMSE of 10.32 and a mean absolute error (MAE) of 8.54 on an independent test dataset. The R-squared values further emphasize the model's accuracy, with a value of 0.902 for the training data (**A**) and 0.922 for the test data (**B**)

varied environments will enhance our comprehension of CHC dynamics and their forensic utility.

In our laboratory-based investigation, we assessed the changes in CHCs of empty puparia placed in two different storage media: soil (representing outdoor conditions) and paper towels (representing indoor conditions) [4]. After eight weeks, notable degradation was observed for all the CHCs studied in the soil medium. However, in the current study, only two of the four investigated CHCs showed degradation in buried conditions, while all CHCs exhibited a decline in concentration over time when puparia were placed above ground, resembling the effects of outdoor conditions. In contrast, the paper-towel medium, simulating indoor conditions, displayed a slower degradation pattern. Specifically, after 120 days, significant degradation was noted for n-C25 and n-C28, while concentrations of n-C27 and

n-C29 increased [4]. In the current indoor investigation, no significant variations were noted. Indeed, while laboratory conditions attempt to mimic real-world environments, there can still be variations in degradation patterns between laboratory and actual outdoor conditions. Factors such as temperature fluctuations, humidity levels, microbial activity, and exposure to sunlight may influence the degradation rates of organic materials, including insect remains, in natural settings compared to controlled laboratory environments. These environmental variables create dynamic conditions that can significantly impact the decomposition process.

In a notable study, the degradation of CHCs in empty puparia of *Chrysomya rufifacies* Macquart, 1842 (Diptera: Calliphoridae) was investigated in laboratory and field conditions. The findings revealed a stark contrast in the rate of hydrocarbon weathering between the two settings, with much more rapid degradation observed in the field [26, 41]. This disparity underscores the profound influence of environmental conditions on the degradation kinetics of puparial hydrocarbons. The accelerated weathering observed in the field suggests a complex interplay of factors such as temperature fluctuations, exposure to UV radiation, microbial activity, and moisture levels, all of which contribute to the breakdown of organic compounds [33, 38, 39, 42]. Another study compared the differences in hydrocarbon and fatty acid esters, as well as transesterified waxes, between lab-preserved old puparia and fresh ones of *Hydrotaea aenescens* (Diptera: Muscidae) over a 15-year period [43]. The comparison between recent (2012) and older (1997) puparia contents has revealed significant differences in composition. Specifically, there is an observed general decrease in the chain length of the n-alkane distribution pattern, coupled with an increase in the length of ester chains. Both extracts also contain traces of three hopane hydrocarbon congeners. Furthermore, differences in hydrocarbon and fatty acid esters from transesterified waxes are clearly discernible. Their distribution patterns exhibit similarities to those reported by Zhu et al. for puparia exposed to weathering over 90 days [26, 41]. A recent study utilized multivariate analysis to investigate the weathering and aging of empty puparial cases from two blow fly species over nine months [44]. Quantifying the weathering time of these insect traces could significantly aid in narrowing down longer PMI.

While laboratory studies offer controlled environments conducive to systematic experimentation and observation, they may not fully capture the intricacies of real-world degradation processes. Consequently, extrapolating findings from laboratory experiments to natural scenarios requires careful consideration of potential differences and limitations. Field experiments or observational studies play a crucial role in complementing laboratory research by providing insights into the complexities of decomposition dynamics in diverse environmental contexts. The current study is an attempt in this direction.

Also, a relevant study investigated temporal variations in the reflection spectrum of *Chrysomya megacephala* Fabricius, 1794 (Diptera: Calliphoridae) pupa using hyperspectral imaging (HIS), along with proposing the eXtreme Gradient Boosting Regression (XGBR) model as an optimal approach for estimating pupa development time based on HSI data [45]. Their findings shed light on the dynamic changes occurring during pupa development and the potential of HSI in monitoring such changes non-invasively. Moreover, the introduction of the XGBR model represents a promising advancement in accurately estimating pupa age, which is crucial in forensic entomology for determining PMImin [45]. Previous research has employed artificial neural networks (ANN) to identify fly maggots based on images of spiracles, predict the age of adult flies, and estimate the weathering time of empty puparia in other forensically significant fly species [46–48]. Our recent study focused on the weathering patterns of empty puparia of another forensically important blowfly *Lucilia sericata*, various machine learning models, including ANN, support vector machine (SVM), and XGBoost, was employed to predict the age of empty puparia. Our first approach was regression analyses (detailed in Supplementary file S6), which yielded moderate R^2 values of 0.78 for buried conditions and 0.89 for above-ground conditions. We sought to enhance accuracy and predictive performance, particularly in discerning age groups based on CHCs, so we turned to CDA. However, to achieve superior results, we implemented XGBoost. XGBoost is versatile and can be applied to various supervised learning tasks, including regression, classification, and ranking problems. The study reported that XGBoost demonstrated superior performance to the other models, achieving higher accuracy in age prediction. This highlights the effectiveness of advanced machine learning techniques, particularly XGBoost, in forensic entomology research for accurately estimating late PMI. A review study highlights the challenges in estimating PMI in highly decomposed bodies and skeletal remains, citing environmental factors and the scarcity of reliable time since death markers as key obstacles [49]. It emphasizes the need for a multidisciplinary approach, combining taphonomic, morphological, and entomological assessments and advanced techniques like fluorescence and proteomics [11, 49–55]. The study underscores the variability in decomposition rates and the importance of using multiple methods to enhance accuracy tailored to each case's specifics and available resources. Understanding the limitations and varying reliability of these methods is crucial for forensic practitioners aiming to achieve accurate PMI estimations in cases involving severe decomposition.

Limitations of the study include the fact that it did not investigate other influential environmental factors such as microbial activity, UV exposure, moisture levels, and geographical variations. While these factors are known to likely

impact the degradation of CHCs, their specific contributions require further empirical validation through subsequent research efforts. This study establishes a foundational baseline for future investigations in this direction.

Furthermore, enhancing the machine learning models used in this study by incorporating additional variables (as it was beyond the logistic and time frame of the current study), such as a broader range of CHCs and a more extensive dataset, is essential for advancing the accuracy and reliability of age estimation models based on CHC profiles. These enhancements represent critical steps towards refining forensic entomology methodologies to achieve more precise estimations of PMI. However, before these methods can be reliably applied in practical forensic settings, further research is needed to confirm and expand upon these findings.

Conclusion

The study provides valuable insights into the intricate dynamics of CHC profiles in empty puparia of *Calliphora vicina* across diverse environmental conditions. The findings reveal distinct degradation patterns, with burial and above-ground settings showing significant declines in CHC concentrations over time, while indoor conditions exhibit relative stability. These results highlight the complex interplay between environmental factors, microbial activity, and chemical processes influencing CHC degradation. The study emphasizes the importance of considering environmental variables in forensic entomology and ecological research, particularly for age estimation and environmental monitoring purposes. However, the study's limitations include its focus solely on selected environmental variables. Factors like microbial activity, UV exposure variations, moisture levels, and geographical influences, known to impact CHC degradation, were not comprehensively investigated. Future research should validate findings across broader environmental contexts, expand datasets, and incorporate additional variables to refine CHC dynamics understanding. Moreover, the integration of machine learning techniques with entomological expertise holds potential for standardized methodologies for age determination, enhancing the reliability of entomological evidence in legal contexts and opening avenues for future research and development.

Declarations

References

1. Otte T, Hilker M, Geiselhardt S (2018) Phenotypic plasticity of cuticular hydrocarbon profiles in insects. J Chem Ecol 2018 44(3):235–247. https://doi.org/10.1007/S10886-018-0934-4

2. Sharif S, Wunder C, Kaleem Khan M et al (2023) Degradation dynamics of cuticular hydrocarbons of empty puparia in forensically important blow fly Lucilia Sericata in soil and under room conditions: insights and machine learning applications. Forensic Chem 35:100519. https://doi.org/10.1016/J.FORC.2023.100519

3. Hatano E, Wada-Katsumata A, Schal C (2019) Environmental decomposition of cuticular hydrocarbons generates a volatile pheromone that guides insect social behavior. bioRxiv: 773937. https://doi.org/10.1101/773937

4. Sharif S, Wunder C, Khan MK et al (2023) Cuticular hydrocarbons as weathering biomarkers of empty puparia of the forensically important blowfly Calliphora vicina Robineau-Desvoidy, 1830 (Diptera: Calliphoridae) in soil v/s under room conditions. Forensic Sci Int 349:111748. https://doi.org/10.1016/J.FORSCIINT.2023.111748

5. Qadir I, Qamar A, Paul B, Mir AH (2021) Cuticular hydrocarbons C14-C36 are potential contact pheromonal elements modulating some behaviors in Zygogramma Bicolorata (Coleoptera: Chrysomelidae). Biol (Bratisl) 76:123–132. https://doi.org/10.2478/s11756-020-00515-w

6. Gołębiowski M, Stepnowski P (2022) Chemical composition of insect surface waxes: biological functions and analytics. Handbook of Bioanalytics 1–19:1. https://doi.org/10.1007/978-3-030-63957-0_29-1

7. Ma M, Luo J, Li C et al (2023) A life-and-death struggle: interaction of insects with entomopathogenic fungi across various infection stages. Front Immunol 14:1329843. https://doi.org/10.3389/FIMMU.2023.1329843/BIBTEX

8. Hodecek J, Jakubec P (2022) Spatio-temporal distribution and habitat preference of necrophagous Calliphoridae based on 160 real cases from Switzerland. Int J Legal Med 136:923–934. https://doi.org/10.1007/S00414-021-02769-8/TABLES/3

9. Singh CP, Sharif S, Qamar A (2024) Forensic entomology research in India: a critical analysis of current state and future directions. Munis Entomol Zool 19(1):278–304

10. Sharma R, Kumar Garg R, Gaur JR (2015) Various methods for the estimation of the post mortem interval from Calliphoridae: a review. Egypt J Forensic Sci 5:1–12. https://doi.org/10.1016/J.EJFS.2013.04.002

11. Sukontason KL, Kanchai C, Piangjai S et al (2006) Morphological observation of puparia of Chrysomya nigripes (Diptera:

Calliphoridae) from human corpse. Forensic Sci Int 161:15–19. https://doi.org/10.1016/J.FORSCIINT.2005.10.013

12. Wydra J, Matuszewski S (2021) The optimal post-eclosion interval while estimating the postmortem interval based on an empty puparium. Forensic Sci Med Pathol 17:192–198. https://doi.org/10.1007/S12024-020-00328-Y

13. Sharif S, Qamar A (2021) Insect faunal succession on buried goat carcass in Aligarh Region of Uttar Pradesh, India, with implications in forensic entomology. Egypt J Forensic Sci 11:1–8. https://doi.org/10.1186/S41935-021-00235-5/FIGURES/3

14. Sharif S, Qamar A (2022) Diptera and Coleoptera colonizing abandoned above-ground carcasses at various stages of decay in Uttar Pradesh, India. 148:273–282. https://doi.org/10.3157/061.148.0208

15. Marchetti D, Arena E, Boschi I, Vanin S (2013) Human DNA extraction from empty puparia. Forensic Sci Int: 229–e29. https://doi.org/10.1016/J.FORSCIINT.2013.03.043

16. Malejko J, Deoniziak K, Tomczuk M, Długokencka J (2020) Puparial cases as toxicological indicators: Bioaccumulation of Cadmium and Thallium in the forensically important Blowfly Lucilia Sericata. Front Chem 8:586067. https://doi.org/10.3389/fchem.2020.586067

17. Chophi R, Sharma S, Sharma S, Singh R (2019) Forensic entomotoxicology: current concepts, trends and challenges. J Forensic Legal Med 67:28–36. https://doi.org/10.1016/j.jflm.2019.07.010

18. Kula C, Amendt J, Drijfhout FP, Moore HE (2022) Geographical variation of cuticular hydrocarbon profiles of adult flies and empty Puparia Amongst three populations of Calliphora vicina (Diptera: Calliphoridae). J Med Entomol. https://doi.org/10.1093/JME/TJAC167

19. Moore H, Lutz L, Bernhardt V et al (2022) Cuticular hydrocarbons for the identification and geographic assignment of empty puparia of forensically important flies. Int J Legal Med 136:1791–1800. https://doi.org/10.1007/s00414-022-02786-1

20. Amendt J, Krettek R, Zehner R (2004) Forensic entomology. Naturwissenschaften 91:51–65. https://doi.org/10.1007/S00114-003-0493-5/TABLES/4

21. Lutz L, Zehner R, Verhoff MA et al (2021) It is all about the insects: a retrospective on 20 years of forensic entomology highlights the importance of insects in legal investigations. Int J Legal Med 135:2637–2651. https://doi.org/10.1007/S00414-021-02628-6/

22. León-Morán LO, Pastor-Belda M, Viñas P et al (2024) Discrimination of Diptera order insects based on their saturated cuticular hydrocarbon content using a new microextraction procedure and chromatographic analysis. Anal Methods. https://doi.org/10.1039/D4AY00214H

23. Moore HE, Butcher JB, Adam CD et al (2016) Age estimation of Calliphora (Diptera: Calliphoridae) larvae using cuticular hydrocarbon analysis and Artificial neural networks. Forensic Sci Int 268:81–91. https://doi.org/10.1016/j.forsciint.2016.09.012

24. Bernhardt V, Pogoda W, Verhoff MA et al (2017) Estimating the age of the adult stages of the blow flies Lucilia Sericata and Calliphora vicina (Diptera: Calliphoridae) by means of the cuticular hydrocarbon n-pentacosane. Sci Justice 57:361–365. https://doi.org/10.1016/J.SCIJUS.2017.04.007

25. Paula MC, Michelutti KB, Eulalio ADMM et al (2018) New method for estimating the postmortem interval using the chemical composition of different generations of empty puparia: indoor cases. PLoS ONE 13:e0209776. https://doi.org/10.1371/JOURNAL.PONE.0209776

26. Zhu GH, Jia ZJ, Yu XJ et al (2017) Predictable weathering of puparial hydrocarbons of necrophagous flies for determining the postmortem interval: a field experiment using Chrysomya rufifacies. Int J Legal Med 131:885–894. https://doi.org/10.1007/S00414-016-1507-0

27. Gilby AR, McKellar JW (1970) The composition of the empty puparia of a blowfly. J Insect Physiol 16:1517–1529. https://doi.org/10.1016/0022-1910(70)90250-7

28. Sharif S, Wunder C, Amendt J, Qamar A (2024) Deciphering the impact of microenvironmental factors on cuticular hydrocarbon degradation in Lucilia sericata empty puparia: bridging ecological and forensic entomological perspectives using machine learning models. Sci Total Environ 913:169719. https://doi.org/10.1016/J.SCITOTENV.2023.169719

29. Shams MY, Elshewey AM, El-kenawy ESM et al (2023) Water quality prediction using machine learning models based on grid search method. Multimed Tools Appl 83:35307–35334. https://doi.org/10.1007/S11042-023-16737-4/FIGURES/15

30. Austin RN, Callaghan AV (2013) Microbial enzymes that oxidize hydrocarbons. Front Microbiol 4. https://doi.org/10.3389/FMICB.2013.00338/FULL

31. Hollaway S, Faw G, Sizemore RK (1980) The bacterial community composition of an active oil field in the Northwestern Gulf of Mexico. Mar Pollut Bull 11:153–156. https://doi.org/10.1016/0025-326X(80)90141-1

32. Jones J, Knight M, Byrom J (1967) Effect of gross pollution by kerosine hydrocarbons on the microflora of a moorland soil. Nature. https://doi.org/10.1038/2271166a0

33. Leahy JG, Colwell RR (1990) Microbial degradation of hydrocarbons in the environment. Microbiol Rev 54:305. https://doi.org/10.1128/MR.54.3.305-315.1990

34. Mulkins Phillips GJ, Stewart JE (1974) Distribution of hydrocarbon utilizing bacteria in Northwestern Atlantic waters and coastal sediments. Can J Microbiol 20:955–962. https://doi.org/10.1139/M74-147

35. Van Beilen JB, Funhoff EG (2007) Alkane hydroxylases involved in microbial alkane degradation. Appl Microbiol Biotechnol 74:13–21. https://doi.org/10.1007/S00253-006-0748-0

36. Maeng JHO, Sakai Y, Tani Y, Kato N (1996) Isolation and characterization of a novel oxygenase that catalyzes the first step of n-alkane oxidation in Acinetobacter sp. strain M-1. J Bacteriol 178:3695. https://doi.org/10.1128/JB.178.13.3695-3700.1996

37. Moucawi J, Fustec E, Jambu P et al (1981) Biooxidation of added and natural hydrocarbons in soils: Effect of iron. Soil Biol Biochem 13:335–342. https://doi.org/10.1016/0038-0717(81)90073-0

38. Eriksson M, Jong-Ok KA, Mohn WW (2001) Effects of low temperature and freeze-thaw cycles on Hydrocarbon Biodegradation in Arctic Tundra Soil. Appl Environ Microbiol 67:5107. https://doi.org/10.1128/AEM.67.11.5107-5112.2001

39. Janus R, Kołomański K, Wądrzyk M, Lewandowski M Degradation of petroleum diesel fuel accelerated by UV irradiation: the impact of ageing on chemical composition and selected physicochemical properties. https://doi.org/10.1051/e3sconf/201910802003

40. Kebede G, Tafese T, Abda EM et al (2021) Factors Influencing the Bacterial Bioremediation of Hydrocarbon Contaminants in the Soil: Mechanisms and Impacts. J Chem 2021:1. https://doi.org/10.1155/2021/9823362

41. Zhu G, Xu X, Yu X et al (2007) Puparial case hydrocarbons of Chrysomya megacephala as an indicator of the postmortem interval. Forensic Sci Int 169:1–5

42. Woodrow RJ, Grace JK, Nelson LJ, Haverty MI (2000) Modification of Cuticular hydrocarbons of Cryptotermes brevis (Isoptera: Kalotermitidae) in response to temperature and relative humidity. Environ Entomol 29:1100–1107. https://doi.org/10.1603/0046-225X-29.6.1100

43. Frere B, Suchaud F, Bernier G, Cottin F, Vincent B, Dourel L, Lelong A, Arpino P (2014) GC-MS analysis of cuticular lipids in recent and older scavenger insect puparia. An approach to estimate the postmortem interval (PMI). Anal Bioanal Chem 406(4):1081–1088

44. Moore HE, Pechal JL, Benbow ME, Drijfhout FP (2017) The potential use of cuticular hydrocarbons and multivariate analysis

to age empty puparial cases of Calliphora vicina and Lucilia Sericata. Sci Rep 7:1933. https://doi.org/10.1038/s41598-017-01667-7

45. Zhang X, Qu H, Zhou Z et al (2024) Age determination of chrysomya megacephala pupae through reflectance and machine learning analysis. Insects 15:184. https://doi.org/10.3390/INSECTS15030184

46. Apasrawirote D, Boonchai P, Muneesawang P et al (2022) Assessment of deep convolutional neural network models for species identification of forensically-important fly maggots based on images of posterior spiracles. Sci Rep 12:1–9. https://doi.org/10.1038/s41598-022-08823-8

47. Zhang X, Bai Y, Ngando FJ et al (2022) Predicting the Weathering Time by the empty puparium of Sarcophaga Peregrina (Diptera: Sarcophagidae) with the ANN models. Insects 13:808. https://doi.org/10.3390/INSECTS13090808/S1

48. Moore HE, Butcher JB, Day CR, Drijfhout FP (2017) Adult fly age estimations using cuticular hydrocarbons and Artificial neural networks in forensically important Calliphoridae species. Forensic Sci Int 280:233–244. https://doi.org/10.1016/J.FORSCIINT.2017.10.001

49. Franceschetti L, Amadasi A, Bugelli V, Bolsi G, Tsokos M (2023) Estimation of late postmortem interval: where do we stand? A literature review. Biology 12(6):783. https://doi.org/10.3390/biology12060783

50. Schmidt VM, Zelger P, Wöss C et al (2022) Postmortem interval of human skeletal remains estimated with handheld NIR spectrometry. Biology (Basel) 11. https://doi.org/10.3390/BIOLOGY11071020

51. Nagy G, Lorand T, Patonai Z et al (2008) Analysis of pathological and non-pathological human skeletal remains by FT-IR spectroscopy. Forensic Sci Int 175:55–60. https://doi.org/10.1016/J.FORSCIINT.2007.05.008

52. Ortiz-Herrero L, Uribe B, Armas LH et al (2021) Estimation of the postmortem interval of human skeletal remains using Raman spectroscopy and chemometrics. Forensic Sci Int 329:111087. https://doi.org/10.1016/J.FORSCIINT.2021.111087

53. Vanin S, Huchet J-B (2017) Forensic entomology and funerary archaeoentomology. Taphonomy of human remains. In: Schotsmans EMJ, Márquez-Grant N, Forbes SL (eds) Taphonomy of human remains: forensic analysis of the Dead and the depositional environment. Wiley, New York, pp 167–186. https://doi.org/10.1002/9781118953358.CH13

54. Gilbert BM, Bass WM (1967) Seasonal Dating of burials from the Presence of fly Pupae. Am Antiq 32:534–535. https://doi.org/10.2307/2694081

55. Morrow JJ, Baldwin DA, Higley L et al (2015) Curatorial implications of Ophyra capensis (Order Diptera, Family Muscidae) puparia recovered from the body of the blessed Antonio Patrizi, Monticiano, Italy (Middle Ages). J Forensic Leg Med 36:81–83. https://doi.org/10.1016/J.JFLM.2015.09.005