



---

*Research article*

## Modeling discontinuous growth in reared *Panulirus ornatus*: A generalized additive model and Cox proportional hazard model approach

Chuan Hui Foo\*

Department of Mathematics, Faculty of Science and Mathematics, Sultan Idris Education University, Tanjung Malim 35900, Malaysia

\* **Correspondence:** Email: [foo.ch@fsmt.upsi.edu.my](mailto:foo.ch@fsmt.upsi.edu.my).

**Abstract:** Crustaceans exhibit discontinuous growth as they shed hard shells periodically. Fundamentally, the growth of crustaceans is typically assessed through two key components, length increase after molting (LI) and time intervals between consecutive molts (TI). In this article, we propose a unified likelihood approach that combines a generalized additive model and a Cox proportional hazard model to estimate the parameters of LI and TI separately in crustaceans. This approach captures the observed discontinuity in individuals, providing a comprehensive understanding of crustacean growth patterns. Our study focuses on 75 ornate rock lobsters (*Panulirus ornatus*) off the Torres Strait in northeastern Australia. Through a simulation study, we demonstrate the effectiveness of the proposed models in characterizing the discontinuity with a continuous growth curve at the population level.

**Keywords:** Markov property; discontinuous; unified likelihood approach; crustaceans; molt

---

### 1. Introduction

*Panulirus ornatus* is commonly found in the shallow coastal waters of the Indo-Pacific region including the coasts of Australia, Indonesia, Malaysia and Papua New Guinea [1, 2]. It typically inhabits coral reefs, rocky reefs and coastal waters with abundant hiding places such as crevices and caves where it can find shelter during the day [3]. Meanwhile, *Panulirus ornatus* is known for its sociable nature, often forming groups or colonies. They can be spotted in large groups of up to 50 individuals, especially during their juvenile stages. However, as they mature into adulthood, they tend to prefer solitude and may display territorial behavior, occasionally resulting in intraspecific aggression [4]. The optimal water temperature range for *Panulirus ornatus* is between 25 and 29 degrees Celsius, although they possess moderate tolerance for temperatures beyond this range [5]. As juvenile lobsters grow, they periodically molt their exoskeletons to accommodate their increasing

size [6]. To assess the growth and alleviate stress on the lobsters, measurements were taken of the carapace lengths of the shed exoskeletons. The carapace length of lobsters can vary based on factors such as age and gender. Although, adult lobsters generally have a carapace length ranging from 15 to 20 centimeters.

In fisheries research, modeling animal growth is essential for stock assessment and effective management [7]. Growth models are a fundamental aspect of modeling growth rates, age-at-recruitment or length-at-age. While growth in finfish and some invertebrates can generally be modeled as a continuous process, growth in crustaceans is discontinuous due to the molting process. Therefore, the validity of continuous growth models has been questioned especially for crustacean species [8]. As all individuals have exact molting times ( $TI$ ) and increments ( $LI$ ) from each molting throughout the study [9], it is relatively straightforward to estimate the size-at-age of animals as we observe their growth during the entire molting process. However, traditional models do not account for the stepwise growth trajectories of individual crustaceans which can result in difficulties in estimating the molt process such as missing ageing information when the shell sheds off during molting [10].

### 1.1. Modeling length increments $LI$ and time intermolts $TI$

The distinct growth trajectories resulting from molting make it challenging to determine the growth rate of these organisms accurately. The objective of this paper is to address this issue by characterizing both stochastic components  $LI$  and  $TI$  through a unified likelihood approach. We demonstrate that both components are conditionally independent given premolt length and the length of time between molt intervals which allows us to quantify the growth parameters of  $LI$  and  $TI$  separately.

Each individual has its own unique genes that can affect its metabolism and may experience different environmental conditions such as food availability [11] which can influence their actual growth rate. Therefore, a suitable growth model should consider both individual and environmental variability [12]. We must consider the nature of the data and the goals of the analysis before choosing a specific modeling approach. Hence, we propose using a generalized additive model (GAM) in the analysis of  $LI$ . It allows for the modeling of a nonlinear relationship and captures complex growth patterns, making it suitable for a better understanding of the growth dynamics of *Panulirus ornatus*.

To model the growth of individuals'  $TI$ , we apply a semiparametric method, specifically the Cox proportional hazard model (CPHM). The misspecification of distribution assumptions may result in biased estimation and misleading conclusions. For this reason, a semiparametric approach would be more robust as there is no assumption made about the distribution of survival times [13]. For instance, [14] provided a comprehensive review of survival data analysis for biological studies which supports the use of a semiparametric approach. By using both the GAM and CPHM, the study provides a comprehensive analysis of the growth patterns in *Panulirus ornatus*. The GAM allows for the modeling of continuous growth curves, accommodates nonlinear growth patterns and captures the overall growth trajectory of the population.

On the other hand, the CPHM is particularly useful when dealing with censored data where the exact event time is unknown or only partially observed. These censored observations can be incorporated into the model, providing more accurate estimates of growth rates and survival probabilities. The study employs simulation techniques to evaluate the models' performance and assess the impact of various factors on growth patterns. Simulation allows for different growth scenarios and investigates the effects

of specific variables on the growth trajectory and discontinuities in the population. Therefore, it can provide valuable insights into the underlying mechanisms and factors influencing growth in *Panulirus ornatus*.

The contribution of this paper is to characterize the discontinuous trajectories of lobster data through two stochastic processes, moult increments and time intermoult. However, both stochastic processes need to be modeled separately to determine their respective growth parameters. For modeling moult increments, we consider a Generalized Additive Model, while we use a semiparametric approach namely Cox Proportional Hazard Model to estimate the time intermoult. Once the estimated parameters are obtained, we combine the two models through simulation to explicitly infer the growth rate. This allows us to estimate the individual discontinuous growth patterns as a continuous smooth curve at the population level for lobsters. Understanding the growth patterns and discontinuities in crustaceans like *Panulirus ornatus* has important implications for their management and conservation. By using a combined GAM and CPHM approach, the study provides a robust framework for predicting and analyzing growth patterns, allowing for better-informed decisions in aquaculture, fisheries and conservation efforts. This knowledge can aid in optimizing rearing conditions, implementing appropriate harvesting strategies and monitoring population dynamics more effectively.

## 2. Materials and methods

### 2.1. Data

#### 2.1.1. Rearing ornate rock lobsters (*Panulirus ornatus*)

There are two fundamental approaches to estimating the growth data of lobsters, the tag-recapture method [15] and the tank data setting. For this study, we focus on tank data where all the samples are reared in the aquarium and their growth processes are observed directly. Implementing a new experimental research design in laboratory settings can be expensive due to limitations in data collection or ongoing monitoring efforts such as ensuring food supply, close monitoring and cleansing the bottom of aquariums to investigate long-term trends or patterns in lobster populations. In such cases, utilizing reared data provided by the Commonwealth Scientific and Industrial Research Organization (CSIRO) can offer a cost-effective alternative. A dataset of 75 ornate rock lobsters including 39 females and 36 males was previously collected from Cairns, Australia between 1995 and 1999. The time intervals for molting ranged from 45 to 260 days for lobsters reared in tanks with water temperatures ranging from 25 to 30 degrees Celsius. All lobsters were measured for their carapace length which ranged from 7.9 to 158.3 mm.

The samples will be housed in a controlled environment in the laboratory with appropriate water quality parameters such as temperature, salinity and pH maintained at optimal levels for lobster growth. On a daily basis, the lobsters will be provided with a suitable diet that mimics their natural diet typically consisting of frozen penaeid prawns or frozen bivalve molluscs. The juveniles were randomly assigned to twelve separate aquaria each with a varying density ranging from 1–10, a capacity of 108 liters and dimensions of 60 cm x 60 cm x 30 cm (length x width x depth). The aquaria were supplied with filtered sea water at a rate of 4.5 liters per hour. Six groups of juveniles were reared at a temperature range of 25–30 °C and each tank was oxygenated and equipped with two concrete-block shelters placed on a

bed of loose sand.

The juvenile lobsters will be regularly monitored and each molting event will be recorded. This will involve visually observing the lobsters during the molting process which is a vulnerable time for lobsters as they shed their old exoskeleton and harden their new exoskeleton. The molt increment which is the difference in size between the old and new exoskeleton will be measured and recorded for each lobster. This will provide information about the growth rate of the lobsters during each molt cycle. From February 1995 to June 1999, juvenile *Panulirus ornatus* were collected at Cairns, Australia and their carapace lengths were measured (initial size range: 7.9–158.3 mm CL).

In addition, the intermolt time which is the duration between two consecutive molts will be recorded for each lobster. This can be determined by tracking the time elapsed between successive molt events. The intermolt time can vary depending on various factors such as lobster size, temperature and nutrition. Studying it can provide insights into the growth and development of lobsters in a laboratory setting.

## 2.2. Analytical approach

### 2.2.1. Modeling *LI* through generalized additive model GAM

The reason for using GAM is that the growth patterns are more intricate and are better captured by flexible models like GAM. GAM can capture complex patterns and interactions in the data without relying on strict assumptions about the functional form of the relationship. GAMs are particularly effective when the relationship between the dependent and independent variables is nonlinear (as growth paths generally are) which can be challenging for other regression techniques.

We consider a nonlinear model GAM to model length increment at molt, denoted as *LI*. In general, the growth trajectory of an individual is often non-linear and may follow a more complex pattern such as a sigmoidal curve. In such cases, a linear model may not be appropriate as it assumes a linear relationship between the response variable and the covariates. Furthermore, the assumption of normality of residuals may be violated when the growth trajectory is non-linear, leading to biased estimates and incorrect statistical inferences.

Therefore, the GAM is more feasible for capturing complex nonlinear relationships between the response variable and the covariates. The smooth functions used in GAMs can take many different forms, including splines, polynomials and other functions. GAM is proposed for estimating parameters of lobsters by modeling the relationship between carapace length and age using a non-linear function such as a cubic spline. The GAM can be written in the following form:

$$g(E(Y)) = \alpha + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) + \varepsilon, \quad (2.1)$$

where  $g(\cdot)$  is a link function,  $\alpha$  is a y-intercept,  $E(Y)$  is the mean of the outcome  $Y$ ,  $f_1(X_1) + f_2(X_2) + \dots + f_p(X_p)$  are smooth functions of the predictors  $X_1, X_2, \dots, X_p$  and  $\varepsilon$  is the error term.

To apply this equation to model lobster growth, we denote  $Y$  as the carapace length ( $L$ ),  $X_1$  as the age-at-molt ( $age$ ),  $X_2$  as the premolt length ( $L^-$ ),  $X_3$  as the second order of the premolt length [ $(L^-)^2$ ] and  $X_4$  as the molt increments ( $I$ ). The statistical approach of utilizing the GAM enables us to explore the relationship between the carapace length of lobsters and different explanatory variables. By capturing the effects of the independent variables on the dependent variable, we can infer the nature

of their association in a more comprehensive manner. Mathematically,  $L = I + L^-$ , the carapace lengths  $L$  of an individual can be described explicitly given that the molt increment  $I$  and premolt length  $L^-$  are known.

### 2.2.2. Modeling $TI$ through Cox proportional hazard model, CPHM

For the Cox proportional hazard model, it can be used to estimate time-to-moult data particularly in the presence of censored observations. Regarding distribution assumptions, the Cox model does not require explicit assumptions about the distribution of intermoult times. This flexibility allows it to be used with data that may not follow a specific distribution.

For the growth of crustaceans, a molt is considered an event. Thus, the time intervals between two consecutive molts denoted as  $TI$  can be deemed as “time-to-event” in survival analysis. Generally, the nonparametric method is preferable to its counterpart since the subsequent molting times remain unknown at the end of the experiments (due to right-censoring). The hazard function is

$$h(t|x) = h_0(t) \exp(x^T \beta),$$

where  $h_0(t)$  is baseline hazard function with independent covariates,  $x$  while the exponential part ensures the hazard function to be non-negative and time-independent. Given that the baseline hazard function has no distributional assumption, we can opt out the parameter estimates for  $h_0(t)$ .

To investigate the correlation between molting time and covariates  $x$  in lobsters, we utilize a hazard function to analyze molt time intervals in the form of

$$h(t|x) = h_0(t) \exp[\beta_1 L^- + \beta_2 (L^-)^2 + \beta_3 T + \beta_4 D], \quad (2.2)$$

where  $L^-$ ,  $(L^-)^2$  are the first and second order of premolt length,  $T$  is the tank water temperature and  $D$  is the density of tank lobsters. Parameter estimates in the CPHM are obtained by optimizing the log-partial likelihood function

$$l(\beta) = \sum_{\delta_i=1} [x_i \beta] - \log \left\{ \sum_{j \geq i} \exp(x_j \beta) \right\}$$

of which  $\delta_i$  is an indicator function implying 1 for uncensored and 0 for censored data.

### 2.3. Simulation through unified likelihood approach

After obtaining all growth parameters for both stochastic components  $LI$  and  $TI$ , we derive a joint likelihood function to combine these two processes, resulting in a population mean curve for all lobsters.

Let lobster  $i$  has  $n_i$  repeated measurements  $\{L_{ij-1}, \Delta_{ij}, I_{ij}\}_{j=1}^{n_i}$  where  $L_{ij-1}$  is the  $j$ th premolt length,  $\Delta_{ij}$  is the  $j$ th molt time intervals and  $I_{ij}$  is the  $j$ th molt increment. For instance, lobster  $i$  molted 3 times with joint density function of

$$f\{(L_{i2}, \Delta_{i3}, I_{i3}), (L_{i1}, \Delta_{i2}, I_{i2}), (L_{i0}, \Delta_{i1}, I_{i1})\}.$$

In general, the joint function for  $i$ -th individual can be written as

$$\prod_{j=1}^{n_i} f(\Delta_{ij}|L_{ij-1}, \Delta_{ij-1}) \prod_{j=1}^{n_i} g(I_{ij}|L_{ij-1}, \Delta_{ij})h(L_{i0}).$$

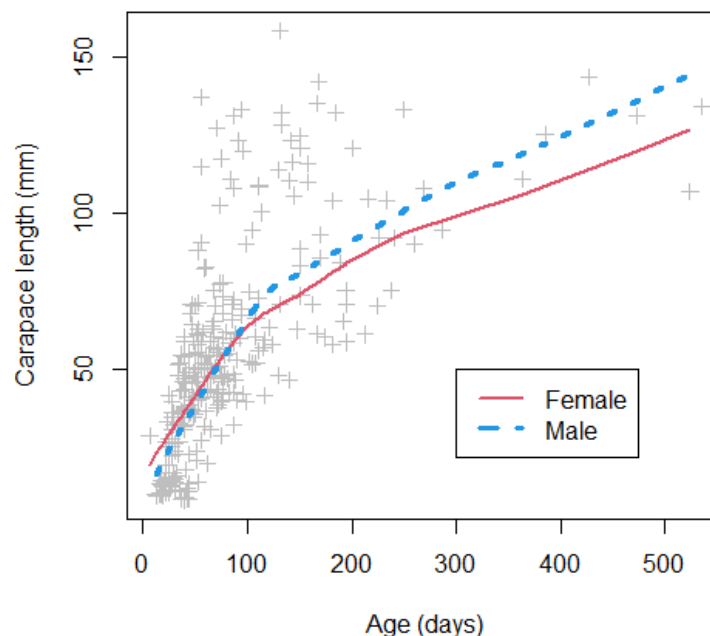
Suppose that we have a total of  $N$  lobsters. The likelihood function for individuals is

$$\prod_{i=1}^N \prod_{j=1}^{n_i} f(\Delta_{ij}|L_{ij-1}, \Delta_{ij-1}) \prod_{i=1}^N \prod_{j=1}^{n_i} g(I_{ij}|L_{ij-1}, \Delta_{ij})h(L_{i0}). \quad (2.3)$$

For the first function,  $f(\cdot)$  which refers to the CPHM as a time-to-event model Eq (2.1) while the second function  $g(\cdot)$  represents the GAM as a function of time (refer Eq (2.2)). Both functions condition on premolt length and molt time intervals. Finally, a population mean can be obtained through a simulation study.

### 3. Results

We propose the flexible model, GAM (refer to Eq (2.1)) for modeling molt increments. This model allows for nonlinear relationships between predictors and the response variable, enabling us to estimate the mean growth rate of lobsters as a function of time. Figure 1 illustrates the nonlinear growth trajectories of both male and female lobsters. In general, male lobsters grow bigger than females because females store more energy for reproductive purposes. As a result, the growth rate of female lobsters slows down after the maturity stage which typically occurs around day 150.



**Figure 1.** Observed (+ in grey) and estimated size of lobsters over time using GAM.

The adjusted R-squared indicates that 99.8% and 99.7% of the variance in the dependent variable is explained by the independent variables in the model for females and males, respectively. Additionally, both smoothing terms appear to be highly significant as their p-values are less than 0.05. This finding implies that the association between the predictors and the response is not linear and a smoothing function is needed to adequately capture these relationships (Table 1).

**Table 1.** Estimate of parameters using the generalized additive model.

Sex	Significance of smooth term	Edf	p-value
Female	<i>age</i>	3.274	$< 2e^{-16}$
	<i>L</i>	3.054	$< 2e^{-16}$
	$L^2$	4.000	$< 2e^{-16}$
	<i>I</i>	3.833	$< 2e^{-16}$
Male	<i>age</i>	3.550	$4.9e^{-6}$
	<i>L</i>	3.454	$< 2e^{-16}$
	$L^2$	3.668	$< 2e^{-16}$
	<i>I</i>	4.000	$< 2e^{-16}$

As discussed earlier, to model molting time intervals, the baseline hazard function  $h_0(t)$  is categorized into male and female strata. Two survival curves can be described based on sex-specific traits, as shown in Figure 2. The plot depicts the probability of molting times over time (in days). In the juvenile stages of development, crustaceans typically molt several times a year while adults only molt once a year. This is because they are growing rapidly and need to replace their exoskeleton to accommodate their increasing size. As they mature, the rate of molting tends to decrease as they near their terminal size and their exoskeleton becomes more resistant to damage. The figure shows that after 100 days, female lobsters were more likely to molt than males.

Referring to CPHM as in Eq (2.2), the test is not statistically significant for each of the covariates and the global test is also not statistically significant. Therefore, we claim that the proportional hazards assumption is not violated through statistical test (Table 2). In [16], the assumption of hazard function can be examined by scaled Schoenfeld residuals (see Appendix A1). Figure 3 displays a plot of scaled Schoenfeld residuals over time, where the smooth line indicates the fitted values while the dashed lines are  $\pm 2$  standard error around the fit. Any systematic departures from horizontal lines symbolize non-proportional hazards [17]. From the graphical display, there is no pattern with time for all of the covariates. Therefore, the assumption of the proportional hazard model in (2.2) is supported for the covariates premolt length, temperature and density. We can then conclude that the premolt length  $L$ , water temperature  $T$  and density  $D$  are the key factors of the growth rate.

By conducting simulations, we can characterize the growth trajectories of lobsters following the steps outlined below:

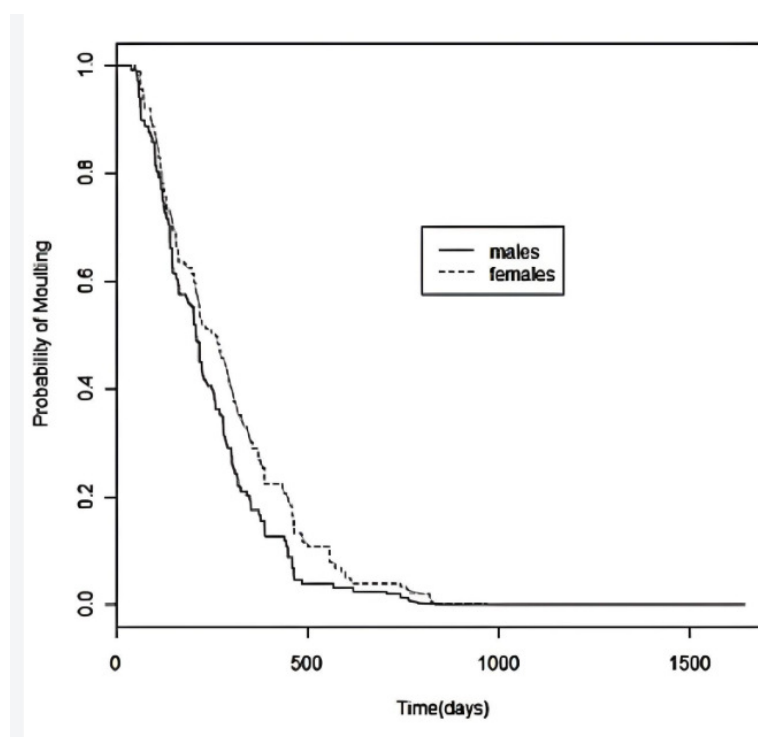
First, we collect data on lobsters, including their growth patterns, survival rates and environmental factors such as tank water temperature and premolt length that might influence their growth and survival.

Second, we develop a GAM to model the relationship between lobster growth length and age-at-molt. The statistical analysis is performed using the  $gam(\cdot)$  function with the logarithm of carapace length as a smooth function of time-to-molt. This allows us to calculate and predict carapace length

values over time.

Third, we develop a CPHM (refer Eq (2.2)) by utilizing the *Surv* function to specify survival times, events, covariates and gender stratification. The CPHM is then fitted through the constructed formula using the *coxph* function to calculate the expected survival probability of lobsters with respect to time (see Appendix A2).

Finally, we combine the GAM and CPH models to create a joint model as shown in Eq (2.3) that can capture the complex relationship between lobster growth and survival. This can be achieved using a unified likelihood approach where the parameters of both functions are estimated simultaneously. The predicted carapace length values from the GAM model are then updated into the CPHM. The mean population of males and females is displayed in a nonlinear smooth curve as shown in Figure 4.

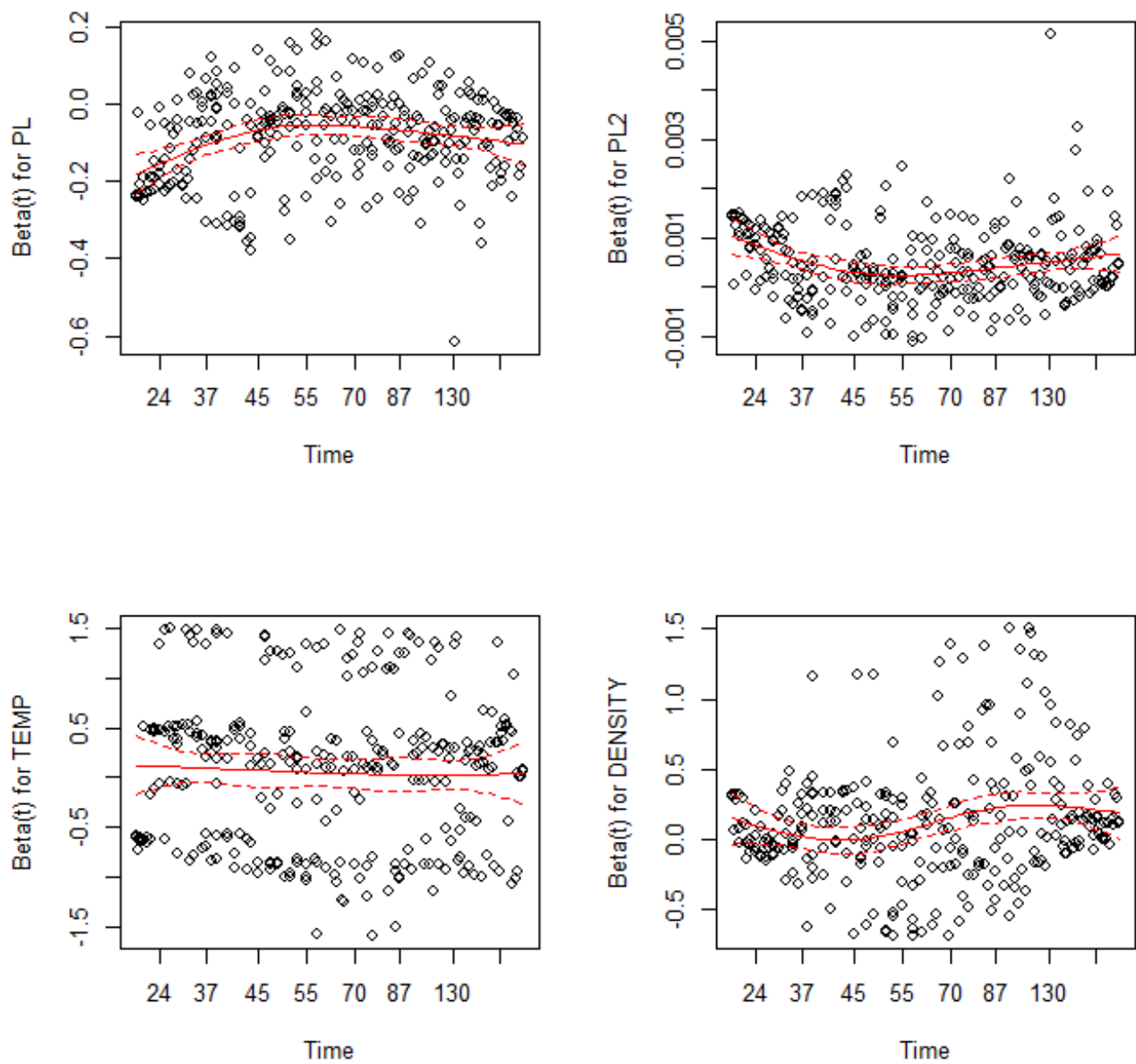


**Figure 2.** Survival curves for male and female lobsters.

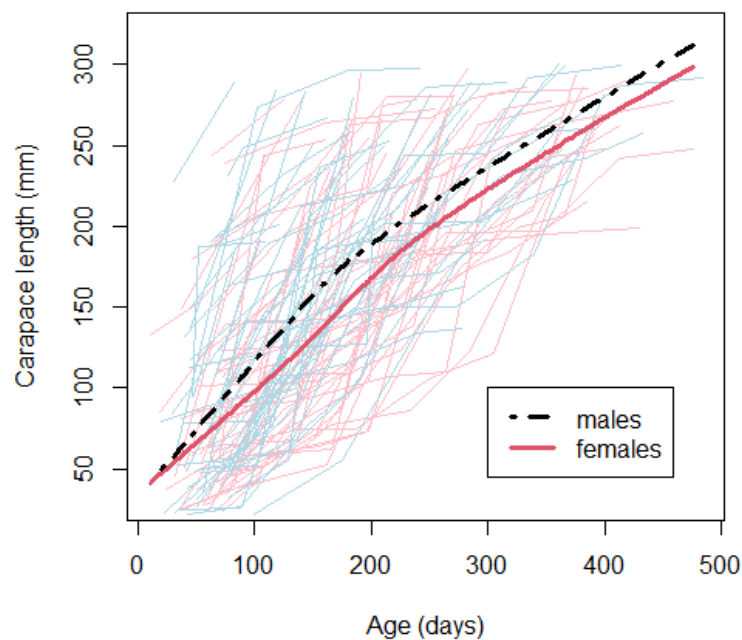
**Table 2.** Parameter estimates of Cox proportional hazard model.

Parameters	rho	chisq	p-value
$L$	0.057	0.949	0.330
$L^2$	-0.034	0.321	0.571
$T$	0.012	0.049	0.824
$D$	-0.026	0.174	0.677
GLOBAL	NA	1.844	0.764





**Figure 3.** Scaled Schoenfeld residuals over time (days) for different covariates.



**Figure 4.** Simulated data and predicted growth curves (dashed and solid lines) over time by unified likelihood approach.

#### 4. Conclusions

In practice, molt increments, as well as molt time intervals, occur concurrently throughout the molting process. Therefore, the discontinuous growth of crustaceans during the molting process complicates growth estimations. We applied a GAM for molt increment analysis and a CPHM to estimate the molt time intervals in lobsters. GAM is known for being a non-parametric smooth function that is robust to outliers in the data, making it a flexible tool for data analysis. It is a type of statistical model that can capture complex data patterns that may not be captured by a simple parametric model. Both GAM and CPHM account for individual and environmental variability to characterize individual stochastic growth trajectories in crustaceans.

In reality, we cannot guarantee which type of distribution fits the data well. Additionally, the molting time of lobsters is often considered as censored data because the exact time of molting is not always observed. CPHM can handle censored data by using maximum likelihood estimation to estimate the hazard function which describes the instantaneous probability of an event occurring at any given time, given that it has not occurred yet. A semi-parametric approach is treated as an alternative assumption since it is less restricted by any violated assumptions. As a result, this statistical framework can produce more robust outcomes compared to a parametric approach.

We have proposed a joint likelihood function to quantify the growth parameters for both stochastic components in crustaceans. This allows us not only to describe the discontinuous growth paths but also to mathematically characterize the population mean growth curve for lobster data in this case. We claim that both  $LI$  and  $TI$  are conditionally independent which means that the parameters in each function can be estimated separately. Through a simulation study, we can integrate both discontinuous functions to yield a smooth, continuous growth function for crustaceans.

Characterizing growth trajectories of lobsters can contribute to our understanding of crustacean biology and ecology. Not only it can provide valuable data for studying various aspects of lobster physiology, behavior and life history, but it can also serve as a foundation for further research on population dynamics, species interactions and ecosystem dynamics. By accurately estimating the lobsters' growth parameters can aid in the effective management of lobster fisheries. Knowledge of molt increments  $LI$  and molt intervals  $TI$  can provide insights into the growth rates of lobsters and help determine optimal harvesting strategies such as setting appropriate size limits or determining seasonal closures to ensure sustainable harvest practices.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

The author is grateful to the Referees for their valuable hints. We would also like to thank the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for providing the data on the ornate spiny lobster tank.

### Conflict of interest

The author declares there is no conflict of interest.

### References

1. R. Wahyudin, A. Hakim, Y. Qonita, M. Boer, A. Farajallah, A. Mashar, et al., Lobster diversity of Palabuhanratu Bay, South Java, Indonesia with new distribution record of *Panulirus ornatus*, *P. polyphagus* and *Parribacus antarcticus*, *Aquacult. Aquarium, Conserv. Legis.*, **10** (2017), 308–327.
2. A. Indarjo, G. Salim, T. I. Maryanto, L. A. Linting, M. Firdaus, Growth patterns and mortality of lobster *Panulirus ornatus* from the catch of bottom gill net fishers in the western waters of Tarakan Island, *HAYATI J. Biosci.*, **30** (2023), 532–542. <http://dx.doi.org/10.4308/hjb.30.3.532-542>
3. P. Briones-Fourzán, E. Lozano-Álvarez, Lobsters: Ocean icons in changing times, *ICES J. Mar. Sci.*, **72** (2015), i1–i6. <https://doi.org/10.1093/icesjms/fsv111>
4. M. Pandolfi, M. F. Scaia, M. P. Fernandez, Sexual dimorphism in aggression: Sex-specific fighting strategies across species, *Front. Behav. Neurosci.*, **15** (2021), 1–12. <https://doi.org/10.3389/fnbeh.2021.659615>
5. M. C. Ooi, E. F. Goulden, G. G. Smith, A. R. Bridle, Haemolymph microbiome of the cultured spiny lobster *Panulirus ornatus* at different temperatures, *Sci. Rep.*, **9**, (2019), 1–13. <https://doi.org/10.1038/s41598-019-39149-7>
6. N. M. Noordin, C. Zeng, P. C. Southgate, Survival, molting pattern, and growth of early blue swimmer crab, *Portunus pelagicus*, juveniles fed diets containing varying levels of cholesterol, *J. World Aquacult. Soc.*, **51**, (2020), 255–265. <https://doi.org/10.1111/jwas.12623>

7. I. Somers, Y. G. Wang, A simulation model for evaluating seasonal closures in Australia's multispecies Northern Prawn Fishery, North American, *North Am. J. Fish. Manage.*, **17** (1997), 114–130.
8. T. Miller, S. G. Smith, Modeling crab growth and population dynamics: Insights from the Blue Crab Conference, *Bull. Mar. Sci.*, **72** (2003), 537–541.
9. C. H. Foo, Parameter estimation of laboratory-reared *Panulirus ornatus*, *SN Appl. Sci.*, **2** (2020), 1–6. <https://doi.org/10.1007/s42452-020-2612-8>
10. Y. J. Chang, C. L. Sun, Y. Chen, S. Z. Yeh, Modelling the growth of crustacean species, *Rev. Fish Biol. Fish.*, **22** (2012), 157–187. <https://doi.org/10.1007/s11160-011-9228-4>
11. G. C. Rilling, E. D. Houde, Regional and temporal variability in growth and mortality of bay anchovy, *Anchoa mitchilli*, larvae in Chesapeake Bay, *Fish. Bull.*, **97** (1999), 555–569. <https://doi.org/10.2307/1353087>
12. T. Russo, P. Baldi, A. Parisi, G. Magnifico, S. Mariani, S. Cataudella, Levy processes and stochastic von Bertalanffy models of growth, with application to fish population analysis, *J. Theor. Biol.*, **258** (2009), 521–529. <https://doi.org/10.1016/j.jtbi.2009.01.033>
13. D. G. Kleinbaum, M. Klein, *Survival Analysis A Self-Learning Text*, third edition, Springer, 2012. <https://doi.org/10.1007/978-1-4419-6646-9>
14. P. Hougaard, Multivariate non-parametric estimates, in *Analysis of Multivariate Survival Data*, Springer, (2000), 442–482. <https://doi.org/10.1007/978-1-4612-1304-8>
15. C. H. Foo, Crustacean growth estimation from tag-recapture data, *J. Adv. Res. Dyn. Control Syst.*, **11** (2019), 1859–1867.
16. M. Tableman, J. S. Kim, *Survival analysis using S: Analysis of time-to-event data*, CRC Press, (2003).
17. D. Schoenfeld, Partial residuals for the proportional hazards regression model, *Biometrika*, **69** (1982), 239–241. <https://doi.org/10.1093/biomet/69.1.239>

## Appendix

### A1. Source code for schoenfeld residuals test

```
library (survival)
attach (lobster)
intcox <- coxph (Surv (START, STOP, STATUS) PL + PL2 + DENSITY + TEMP)
summary (intcox)
cox.zph (intcox)
plot (cox.zph (intcox))
```

### A2. Simulation source code

```
# Create empty data frame to store simulated data
simdata <- data.frame ()
```

---

```
# Generate 100 unique lobsters
uniquelobsters <- paste 0 ("Lobster", 1:100)

# Loop over each unique lobster
for (lobster in uniquelobsters)
  num.measures <- sample (2:7, 1)

# Randomly assign a sex to the lobster (1 for female, 2 for male)
sex <- sample (c (1, 2), 1)

# Generate repeated measurements for each variable
density <- rnorm (num.measures, mean = 2, sd = 1)
lobster_sex <- rep (sex, num.measures)
cl <- sort (runif (num.measures, 20, 300))
inc <- sort (runif (num.measures, 10, 100))
pl <- cl - inc
int <- rnorm (num.measures, mean = 60, sd = 20)
nint <- cumsum (int)
pl2 <- pl*pl

# Create a column with the lobster label
lobster_label <- rep (lobster, num.measures)

# Combine the generated data into a data frame
data <- data.frame (LOBSTER = lobster_label, INC = inc, DENSITY = density, SEX = lobster_sex,
  CL = cl, PL = pl, PL2 = pl2, INT = nint, STATUS = status, START = int)

# Append the lobster data to the simulated data frame
simdata <- rbind (simdata, data)

# Reset row names of the simulated data frame
row.names(simdata) <- NULL

# Load required libraries
library (plyr)
library (mgcv)

# Separate the data into female and male lobsters
female <- subset (simdata, SEX == 1)
male <- subset (simdata, SEX == 2)
```

---

```

# Plot the relationship between CL and INT for female lobsters
plot (simdata $ INT), simdata $ CL, type = "n", xlab = "Age (days)", ylab = "Carapace length (mm)")
CLcount 1 <- count (female $ CL, "LOBSTER")
data 1 <- unique (female $ LOBSTER)

for (i in 1:length (data 1))
{
sub <- t (matrix (NA, nr = CLcount 1 $ freq [i], nc = 2))
sub <- subset (female, LOBSTER == data 1 [i], select = c ("CL", "INT"))
lines (sort (sub [, 2]), sub [, 1], col = "pink")
}

CLcount 2 <- count (male $ CL, "LOBSTER")
data 2 <- unique (male $ LOBSTER)

for (i in 1:length (data 2))
{
sub <- t (matrix (NA, nr = CLcount2 $ freq [i], nc = 2))
sub <- subset (male, LOBSTER == data 2 [i], select = c ("CL", "INT"))
lines (sort (sub [, 2]), sub [, 1], col = "light blue")
}

# Fit the survival Cox model
cox_model <- coxph (Surv (START, STATUS) ~ PL + PL2 + INT + DENSITY, data = female)

# Obtain the survival probabilities
newdata <- data.frame (PL = female $ PL, PL2 = female $ PL2, INT = female $ INT, DENSITY =
female $ DENSITY)
fit <- survfit (cox_model, newdata = newdata)

# Calculate estimated survival time for females (SEX == 1)
subset <- female $ SEX == 1
time <- fit $ time [subset]

# Fit the survival time into the gamma function
gam_model <- gam (log (female $ CL) ~ s (time) + s (female $ PL) + s (female $ INC) + s (female $
PL2), data = female)
smooth <- exp (predict (gam_model))

# Prediction line
lines (lowess (female $ INT, smooth, f = 0.85), lwd = 3, lty = 1, col = 2)

```

---

*# Fit the survival Cox model*

```
cox_model <- coxph (Surv (START, STATUS) ~ PL + PL2 + INT+ DENSITY, data = male)
```

*# Obtain the survival probabilities*

```
newdata <- data.frame (PL = male $ PL, PL2 = male $ PL2, INT = male $ INT, DENSITY = male $ DENSITY)
```

```
fit <- survfit (cox_model, newdata = newdata)
```

*# Calculate estimated survival time for females (SEX == 1)*

```
subset <- male $ SEX == 2
```

```
time <- fit $ time [subset]
```

*# Fit the survival time into the gamma function*

```
gam_model <- gam (log (male $ CL) ~ s (time) + s (male $ PL) + s (male $ INC) + s (male $ PL2), data = male)
```

```
smooth <- exp (predict (gam_model))
```

*# Prediction line*

```
lines (lowess (male $ INT, smooth, f = 0.8), lwd = 3, lty = 4, col = 1)
```

```
legend (300, 100, c ("males", "females"), col = c (1, 2), lty = c (4, 1), lwd = 3:3)
```



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)