

# Longitudinal activity monitoring and lifespan: quantifying the interface

Su I Iao<sup>1,\*</sup>, Poorbita Kundu<sup>1,\*</sup>, Han Chen<sup>1</sup>, James R. Carey<sup>2</sup>, Hans-Georg Müller<sup>1</sup>

<sup>1</sup>Department of Statistics, University of California, Davis, CA 95616, USA

<sup>2</sup>Department of Entomology, University of California, Davis, CA 95616, USA

\*Equal contribution

**Correspondence to:** Hans-Georg Müller; email: [hgmueler@ucdavis.edu](mailto:hgmueler@ucdavis.edu)

**Keywords:** age-at-death, force of mortality, functional data analysis, longitudinal data, mediterranean fruit fly

**Received:** July 23, 2024

**Accepted:** August 13, 2024

**Published:** September 9, 2024

**Copyright:** © 2024 Iao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Understanding the relationship between activity over the entire lifespan and longevity is an important facet of aging research. We present a comprehensive framework for the statistical analysis of longitudinal activity and behavioral monitoring and their relationship with age-at-death at the individual level, highlighting the importance of advanced methodological approaches in aging research. The focus is on animal models, where continuous monitoring activity in terms of movement, reproduction and behaviors over the entire lifespan is feasible at the individual level. We specifically demonstrate the methodology with data on activity monitoring for Mediterranean fruit flies. Advanced statistical methodologies to explore the interface between activity and age-at-death include functional principal component analysis, concurrent regression, Fréchet regression and point processes. While the focus of this perspective is on relating age-at-death with data on movement, reproduction, behavior and nutrition of Mediterranean fruit flies, the methodology equally pertains to data from other species, including human data.

## INTRODUCTION

In recent years there has been a surge in the availability of complex longitudinal data capturing various aspects of organismal life history. Relating such data to individual lifespan and age-at-death is of paramount interest for the study of aging and longevity. We provide an overview of advanced statistical methodologies that are particularly well-suited for analyzing such data, with a focus on understanding the complex relationships between age-at-death and activity, reproduction and diet at the individual level. While we illustrate the methodology with data for lifetime monitoring of Mediterranean fruit flies, it is generic and easily adapts to other species. The methods are interpretable and enjoy broad applicability, including for human longitudinal aging and activity monitoring data [1–5].

The longitudinal monitoring data we consider here encompass three key areas of life history: Activity (quantified as counts of movements per time unit) [6], reproduction (quantified as egg-laying per day for female flies) [7] and longitudinal behaviors (quantified as frequency of each behavior per time unit) [8]. Of central interest is the prediction of remaining lifespan for an individual who has survived to a current age  $a$ , based on the available life history (activity, diet etc.) data from birth to  $a$  for the individual, i.e., the longitudinal monitoring data available in the time interval  $[0, a]$ . Predicting the exact remaining lifespan for an individual at current age  $a$  is futile, but it is possible to predict the distribution of remaining lifetime for an individual [9] by utilizing Fréchet regression [10], as we demonstrate below.

Of interest is also the modeling of the longitudinally observed activity or egg-laying data to elucidate patterns and modes of variation of activity, using covariance and functional principal component modeling [11–13], where we specifically demonstrate the applicability of product functional principal component analysis [14] to jointly model two components of the activity data, intra-day activity measured in hours within each day and age-dependent activity, where age is measured in days, giving rise to a function-valued stochastic process [6].

Providing another perspective, we demonstrate the modeling of daily egg-laying data for female medflies with a global Cox point process model [15] that can be coupled with a functional linear model [16] to quantify the relationship between longevity and reproductive activity [7, 17]. To study the effect of the current level of reproduction on the immediate reproductive rate, we demonstrate below the application of a concurrent regression [18] approach.

Another type of important life history data in aging research are age-varying behavioral data, where several categories of behaviors are observed simultaneously and one aims to establish patterns of behavior change as individuals age [8, 19–21]. Riemannian functional principal component analysis [22] is shown to be uniquely suited for the simultaneous longitudinal study of several behavioral components, accounting for the complex compositional and longitudinal nature of such data.

We note that these approaches can all be adapted to human longitudinal aging data when available, offering a statistical toolkit for researchers investigating human aging and longevity. The statistical methods for the most part are grounded in functional data analysis, which aims at flexible modeling of trajectories and longitudinal data [23–26], allowing for biological interpretations and facilitating the detection of patterns of aging and their relation with longevity.

## RESULTS

### Activity monitoring, diet and longevity

#### *Movement activity profiles of mediterranean fruit flies*

We illustrate the analysis of activity patterns and their relationship with diet and longevity with the movement activity profiles of 96 female adult Mediterranean fruit flies (medflies). The experiment utilized the Monitor-LAM25 system to obtain repeated observations of the 24-hour locomotory activity for the 96 medflies, where each fly was fed with one of three agar-based gel diets that differed in their sugar and yeast hydrolysate content

(50%, 20% and 10% and labeled C50, C20 and C10). Each diet group consisted of 32 female medflies. Each fly was placed in its own glass tube (25 mm diameter, 125 mm length) and the activity count, defined as the number of times a fly passed through an infrared beam placed in the middle of the tube, was recorded for each minute until the death of the fly; a full description with further details can be found in Chen et al. [6].

#### **Predicting remaining lifetime**

Predicting the exact remaining lifetime for an individual still alive at a specific age  $a$  is futile due to the high random variation of individual age-at-death. A more feasible target is to predict the distribution of remaining lifetime for an individual alive at age  $a$ . We demonstrate this by demonstrating the effect of diet by predicting the distribution of remaining lifetime with Fréchet regression (see Appendix) for individuals in each of the diet groups C50, C20 and C10, where the predicted distribution of remaining lifetime is conditional on an individual still being alive at age  $a$ , where  $a$  varies. Of course, the larger the value of  $a$ , the more compressed the remaining lifetime distribution is, as remaining lifetime is becoming shorter with increasing age  $a$ .

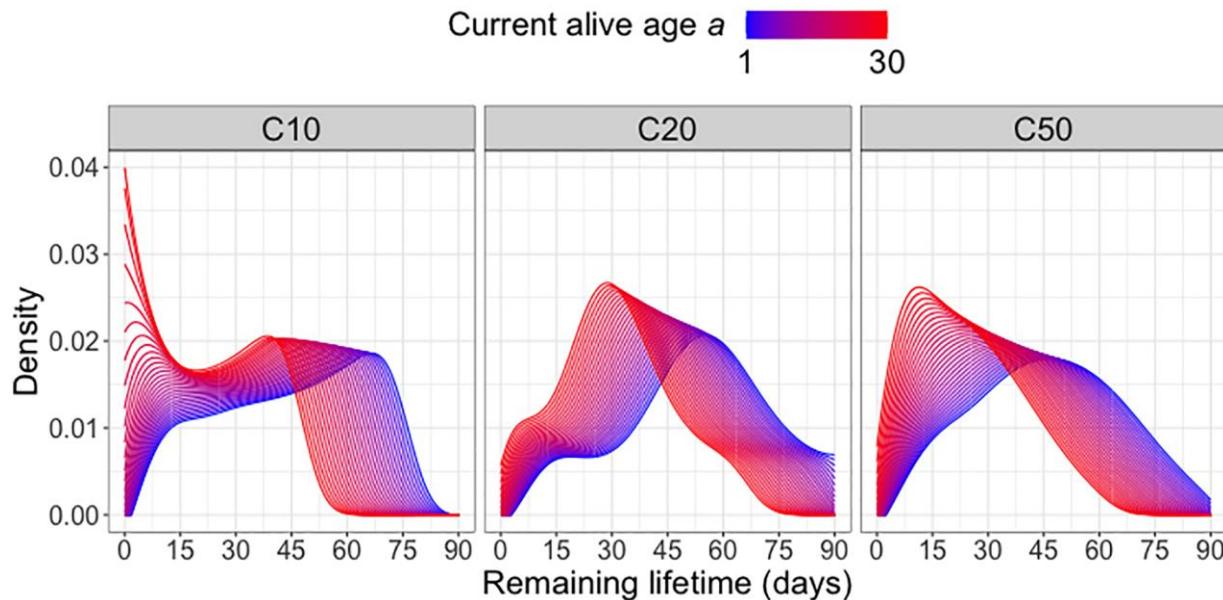
The target to be predicted is the remaining lifetime distribution at current age alive  $a$ , while the predictors are  $(a, C20, C50)$ , in addition to current age alive  $a$  these are indicators for diet groups C20 and C50, where  $C20 = 1$  if a fly is in diet group C20 and otherwise  $C20 = 0$  and analogously for C50; if both  $C20 = 0$  and  $C50 = 0$ , the fly is in diet group C10. The results for the predicted remaining lifetime distributions are shown in terms of the densities of these distributions in Figure 1. As expected, the densities of the predicted remaining lifetime distributions shift to the left as the current age alive  $a$  increases, consistent with the expectation of shorter remaining lifetimes for older individuals. This pattern is observed for all diet groups, however there are variations in the magnitude and shape of the shift. Specifically, in the C10 diet group, the density peak shifts from approximately 70 days at  $a = 1$  (i.e., right after eclosion) to around 40 days as current age alive  $a$  increases to 30 days. At  $a = 30$  days, the density values for the next 10 days shoot up rapidly, indicating elevated risk of death within the first 10 days after surviving past 30 days.

Predicted remaining lifetime distributions can be equivalently visualized as mortality rate (force of mortality, hazard function). Hazard functions can be derived from the density functions of the remaining lifetime distributions or alternatively directly estimated from lifetable data, with proper adjustments for right tail estimation [27, 28]. Figure 2 illustrates that, as

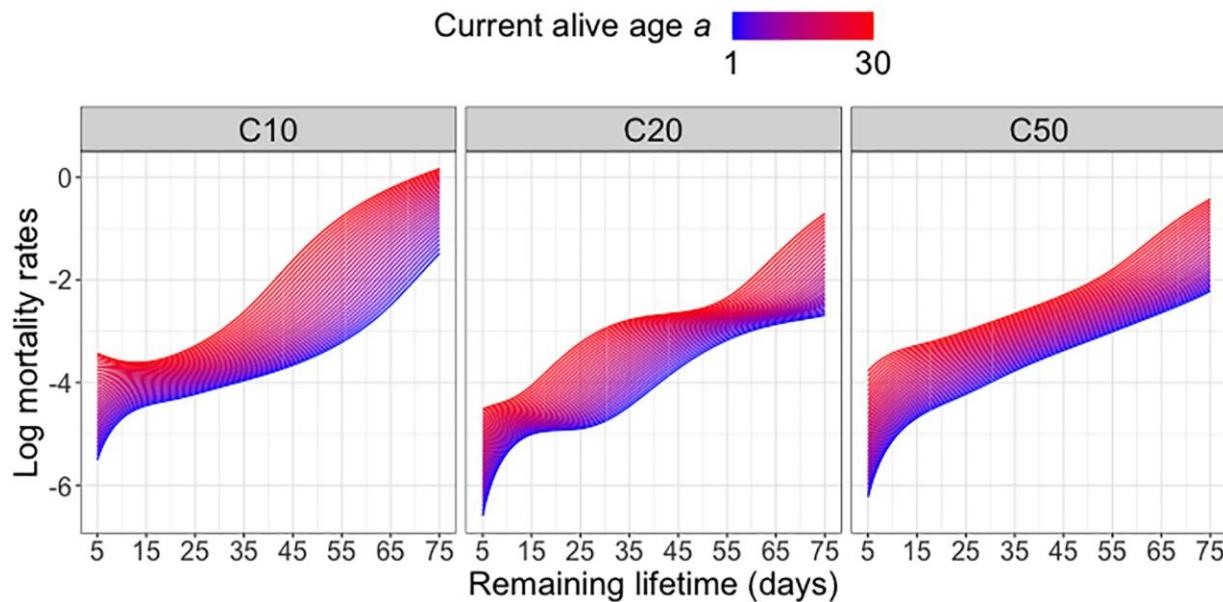
expected, the mortality risk (represented by higher values of the log hazard function) increases with increasing current age alive  $a$  for all diet groups. Across all diet groups, higher ages alive  $a$  are associated with elevated mortality risk at all subsequent ages, in line with expected increased mortality risk in older populations. Among the three diet groups, the C10-fed medflies are subject to the highest mortality risk at older ages  $a$ , while the C20-fed medflies are seen to have a relatively lower mortality risk. However, the overall

patterns of mortality risk across diet groups do not show substantial differences.

Focusing on the diet group C20, which according to Figure 2 is associated with the lowest mortality risk, we further examine the effect of current alive age  $a$  on the remaining lifetime distribution when also adjusting for the overall movement activity from time 0 until age  $a$ . Within each cohort surviving beyond age  $a$ , we classify the subjects into “high” (above median) or “low”



**Figure 1. Remaining lifetime distribution for different diet groups.** Predicted remaining lifetime distributions (visualized as densities) at different values of current alive age  $a$ , where  $a$  varies from  $a = 1$  (blue) to  $a = 30$  (red), for three different diet groups C10, C20, C50 (yeast hydrolysate content 10%, 20% and 50%), using global Fréchet regression (4) (see Appendix).



**Figure 2. Log mortality rates for different diet groups.** Log mortality rates (log hazard functions) of predicted remaining lifetime distributions across varying current alive ages  $a$  at which a subject is still alive for different diet groups; see equation (4) in the Appendix.

(below median) activity groups and implement global Fréchet regression (see (2) in the Appendix) with the remaining lifetime distribution as response and predictors  $a$  and activity group (coded as an indicator variable). Figure 3 indicates a movement of the density peak towards the left and upward with increasing currently alive age  $a$  that is more pronounced for the “high” activity group, which means there is a higher immediate death rate for this group. With increasing age  $a$ , the densities tend to have higher and sharper peaks, followed by a steadier decline. The global Fréchet regression models are fitted using the R function *GloDenReg* in the *frechet* package [29].

#### Analysis of continuous activity monitoring data

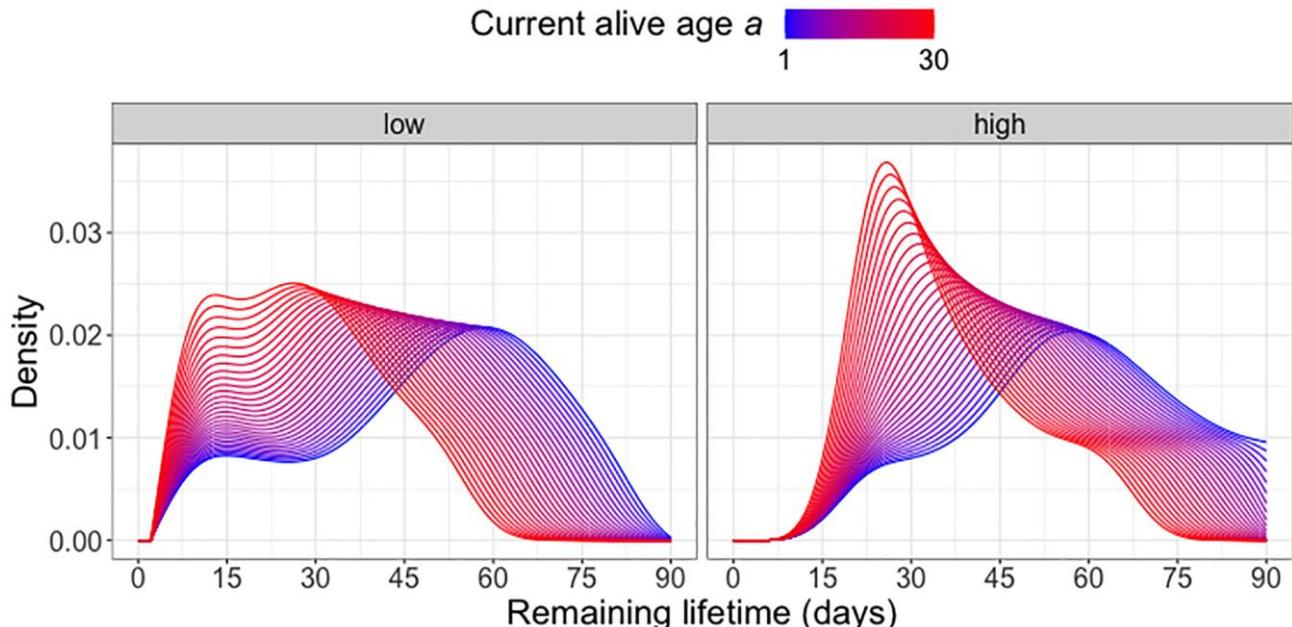
Continuous activity monitoring data can be viewed as a function-valued stochastic process  $X(s, t)$ , representing the activity count at age  $t$  (in days), which is the time index of the process and within-day hour  $s$ , which is the argument of the function observed at each age  $t$ . A key methodology is functional principal component analysis [30–32], which can be implemented with two-dimensional eigenfunctions [6] or, as we demonstrate here, with product eigenfunctions; for background see Chen et al. [14] and the Appendix). Figure 4 shows the first four eigenfunction surfaces for the product approach (see equation (5) in the Appendix), which is based on separate modeling of the hour and day dimensions, capturing potentially distinct patterns of variation for each of the day (age) and hour (within day) time scales.

The eigenfunction products delineate the main modes of variation for these processes, decomposing the total variation into interpretable components. Here the first eigenfunction surface  $\phi_1(t)\psi_1(s)$  (upper left) illustrates a contrast between movement activity early (0–8 am) and late in the day (after 8 am), with diminishing contrast at older ages. The second eigenfunction surface  $\phi_2(t)\psi_1(s)$  (upper right) highlights a contrasting pattern between activity at early and late ages (before and after 12 days), while the third eigenfunction surface  $\phi_1(t)\psi_2(s)$  (bottom left) shows another contrast that divides the intra-day functions into periods before and after 4 pm. The fourth eigenfunction surface  $\phi_2(t)\psi_2(s)$  (bottom right) emphasizes a reversal of the contrast between early and late activity at younger ages (before 4 pm until age 15 days) and late activity at older ages (after 4 pm beyond age 15), but overall does not explain much of the variation. Using the functional principal component scores for this product approach, one can further investigate the relationship between remaining lifetime and activity patterns; see [6] for more details. Implementations of these methods are available in the R package *fdarep* [33] via functions *Dense2dFPCA* and *DenseProductFPCA*.

#### Reproduction and longevity

##### Cost of reproduction

The well-known cost of reproduction hypothesis in biodemography and life history analysis is still somewhat controversial due to a lack of clearly



**Figure 3. Remaining lifetime distribution for different activity levels.** Predicted remaining lifetime distributions (visualized as densities) at varying alive ages  $a$  at which a subject is still alive and low (below median)/high (above median) activity levels, using global Fréchet regression (see (4) in the Appendix). Only C20-fed (yeast hydrolysate content 20%) flies are included in the analysis.

identifiable proximal factors. The general idea is that finite resources must be split between maintenance to extend lifespan and reproduction, both of which require protein sources [7, 34, 35], implying that high reproductive activity could be associated with reduced remaining lifespan. There is some evidence for a regulatory mechanism whereby lifespan is determined by remaining egg-laying potential; if this potential is low because eggs have been depleted then the remaining lifespan tends to be short and vice versa [17].

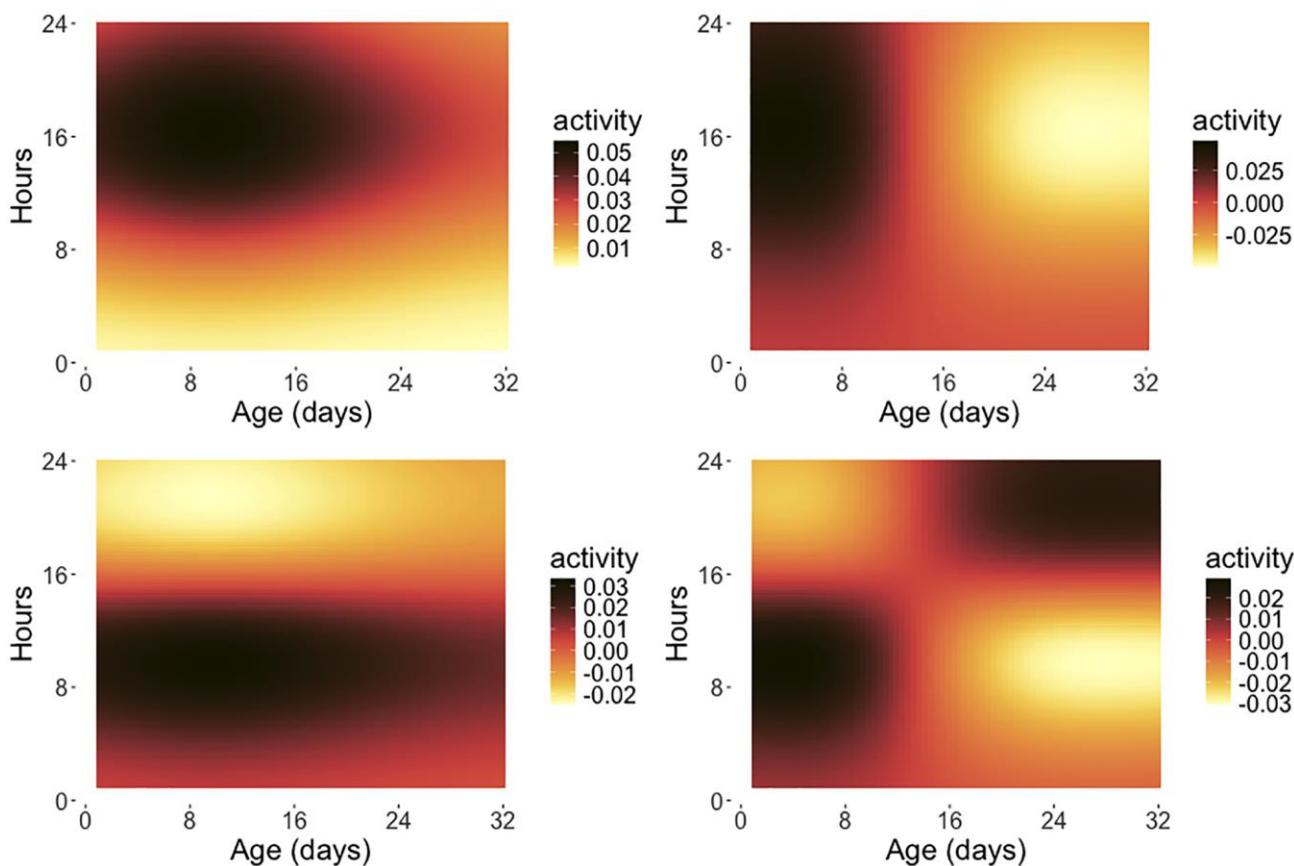
To study cost of reproduction related questions for cohorts of flies where reproduction can be measured as longitudinally measured daily egg-laying requires dedicated statistical methodology. We demonstrate this for a study of 473 female adult Mediterranean fruit flies who survived past age 35 days. The experiment was conducted at temperature  $26 \pm 2^\circ\text{C}$ , relative humidity  $80 \pm 10\%$ , and a 12:12 light:dark cycle. The medflies were placed in their own cages ( $6.5 \times 6.5 \times 12$  cm plastic bottles kept horizontally) with a lid replaced by a fine

mesh through which flies would lay their eggs that, in turn, would fall to a dish lined with a damp, black cloth. These oviposition dishes were collected daily for egg counting. On average, the daily number of eggs laid by female Mediterranean fruit flies tends to peak between 10 and 15 days after emergence and subsequently declines with age, as shown in previous analyses of these data [7, 17, 34].

The observed data are longitudinal daily egg counts, visualized in Figure 5. When analyzing such data, it is important to consider cohorts where all flies survive through a specified age  $a$ , in Figure 5 chosen as  $a = 35$  days, to avoid censoring issues that will lead to bias. The figure shows that these data are afflicted by high variance, with high day-to-day variation.

#### Associations of daily egg count data and longevity

Elucidating the relationship between reproduction and longevity has been of long-standing interest in biodemography [7, 17, 36, 37]. Here we demonstrate the application of a global Cox point process regression

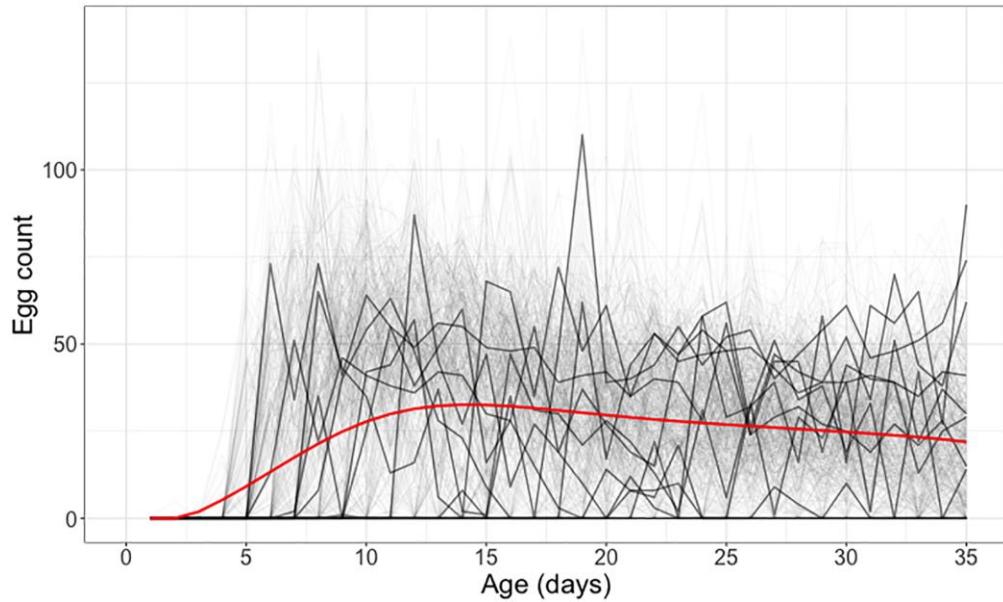


**Figure 4. Eigensurfaces for early age medfly activity data.** The first four eigensurfaces for product functional principal component analysis are depicted, with  $\phi_1(t)\psi_1(s)$  (top left, 13.62%),  $\phi_2(t)\psi_1(s)$  (top right, 8.62%),  $\phi_1(t)\psi_2(s)$  (bottom left, 3.65%) and  $\phi_2(t)\psi_2(s)$  (bottom right, 1.50%) as per (5), where the percentages represent the fraction of variance explained by the respective eigensurface and the  $\phi_j$  are the eigenfunctions reflecting the variation over the age span in days while the  $\psi_k$  are the eigenfunctions for the intra-day variation. The x-axis indicates the age coordinate  $t$  (measured in days) and the y-axis represents the hour coordinate  $s$  within a given day (0–24 hours). Only flies surviving 32 days are included in this analysis.

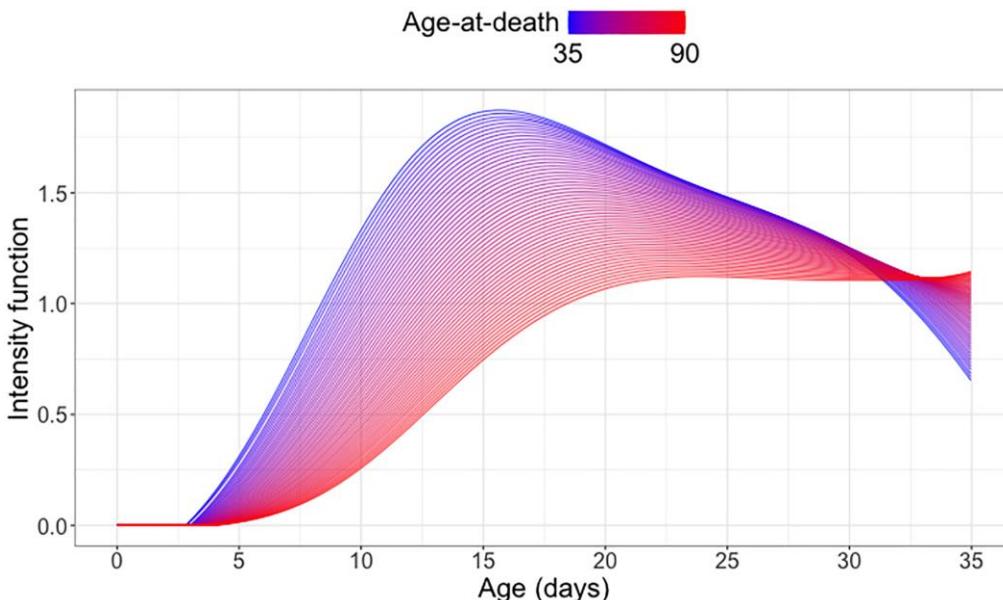
model [15] for such data, where we use a reverse regression approach with age-at-death as predictor and the intensity function of the egg-laying process as response. While reversing the time order, this approach can serve to quantify and visualize associations between longevity and the egg-laying process and we implemented it with the function *GloPointPrReg* in the R package *frechet* [29]; technical details are in the Appendix. Additional tools include a graphical method

to illustrate the connection between reproduction and longevity [38] and a forward prediction approach for remaining lifetime as response with the functional principal component scores obtained from longitudinally observed activity or reproductive trajectories up to current age alive  $a$  as predictors [6].

Figure 6 shows a distinct association between egg count trajectories and age at death. For medflies with lower



**Figure 5. Individual daily egg-laying counts for medflies surviving through 35 days.** Egg-laying counts for 10 randomly selected medflies are highlighted in black, while the other elements in the sample depicting individual daily egg-laying counts are shown in gray. The mean daily egg-laying count, obtained by local linear smoothing is shown in red.



**Figure 6. Cost of reproduction: Quantifying the association between longevity and reproduction.** Predicted conditional intensity functions for daily egg-laying at different age-at-death levels ranging from 35 days (blue) to 90 days (red), using global Cox point process regression for a cohort of medflies who survived past 35 days (see (6) in the Appendix).

age-at-death, the daily egg count tends to peak sharply around 15 days, with a steady decline thereafter. Flies who survive longer tend to have relatively lower egg counts at all ages until 30 days, and especially at younger ages, coupled with enhanced reproductive activity beyond 30 days after emergence. This pattern reflects the observed cost of reproduction for female medflies.

#### **Predicting concurrent reproductive potential**

How the current reproductive activity of a medfly at various ages relates to change in reproductive activity reveals the underlying dynamics of the reproductive process, complementing the dynamic model in [17]. Denote the egg-laying count on day  $t$  by  $X(t)$  with mean  $\mu_X(t) = E[X(t)]$  and consider difference quotients

$$Y(t) = \frac{X(t+1) - X(t)}{(t+1) - t},$$

which serve to approximate the

derivative  $(d/dt)X(t)$  that reflects reproductive activity change at age  $t$ . We aim to study the dependence of  $Y(t)$  on  $X(t)$  by fitting the concurrent regression model [18],

$$E[Y(t)|X(t)] = \beta_0(t) + \beta_1(t)(X(t) - \mu_X(t)), \quad t \in [10, 35] \quad (1).$$

The estimated intercept function  $\beta_0(t)$  represents the estimated mean derivative of the egg-laying trajectories. It is positive prior to 15 days, indicating that egg-laying is ramping up, then turns negative, indicating a decline

in egg-laying that accelerates after 30 days, reflecting the depletion of the reproductive potential of female medflies (see Figure 7).

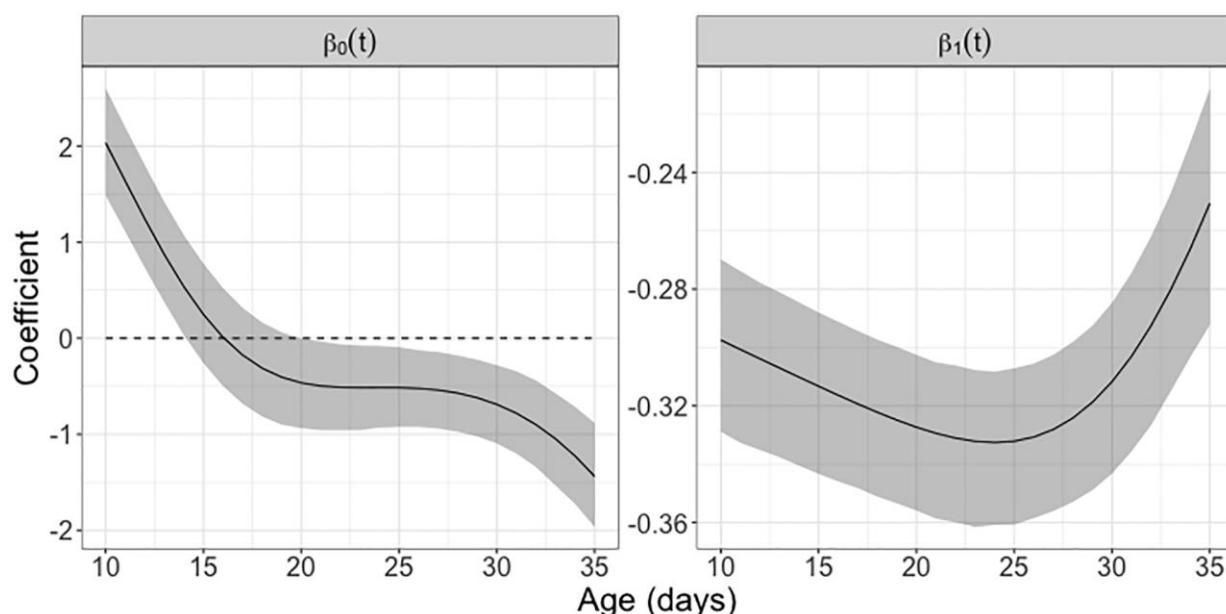
The estimated regression coefficient function  $\beta_1(t)$  acts on the difference between an individual's and the average egg-laying trajectory and takes negative values throughout; pointwise 95% bootstrap confidence intervals show that it is significantly negative from 10 to 35 days. This indicates a statistically significant effect whereby above-average egg-layers experience a more rapid decline than below-average egg-layers and conforms with a process that is self-regulating through a dynamic regression to the mean effect [39].

To fit the varying coefficient model and obtain the 95% bootstrap pointwise confidence bands, we used the functions *ConcurReg* and *GetCI Sparse*, available in the R package *fdaconcur* [40].

#### **Analysis of longitudinally monitored age-specific behavior patterns**

##### **Longitudinal compositional representation of medfly behaviors**

For illustration, we use data on age-specific behavioral patterns continuously recorded for 51 Mediterranean fruit flies under controlled laboratory conditions



**Figure 7. Study of concurrent reproductive potential using varying coefficient model.** The intercept function  $\beta_0(t)$  represents the (approximate) mean reproductive rate associated with daily egg-laying counts in female Mediterranean fruit flies that survived past 35 days. The regression coefficient function  $\beta_1(t)$  depicts the time-varying effect of the accumulated egg-laying count on the immediate reproductive rate. The estimated coefficient functions are based on the fitted concurrent regression (1). The 95% pointwise confidence bands are based on 1000 bootstrap replicates, using concurrent regression (1). If zero is not located within the confidence band this indicates pointwise (not simultaneous) statistical significance in the corresponding age interval.

(temperature:  $25 \pm 2^\circ\text{C}$ , relative humidity:  $65 \pm 5\%$ , light:dark cycle: 14:10). Each medfly was placed in an individual cage ( $12 \times 5 \times 7.5$  cm transparent plastic cup). The behavioral patterns of each fruit fly were observed instantaneously 12 times each day through its lifetime (no censoring). In this analysis, we focus on a cohort of flies surviving past 41 days and the three behaviors of flying  $Z_1(t)$ , walking  $Z_2(t)$  and resting  $Z_3(t)$  at age, respectively; see [8] for further details.

In a preprocessing step, the three behaviors were transformed into proportions per time unit, where the proportions always sum up to 1 and are non-negative, thus forming compositional data; for example, 80% resting, 10% flying and 10% walking during one hour of observations, reflecting the observed proportions in time among just these three behaviors (other behaviors are not considered). We then obtained square-root transformed compositional proportions

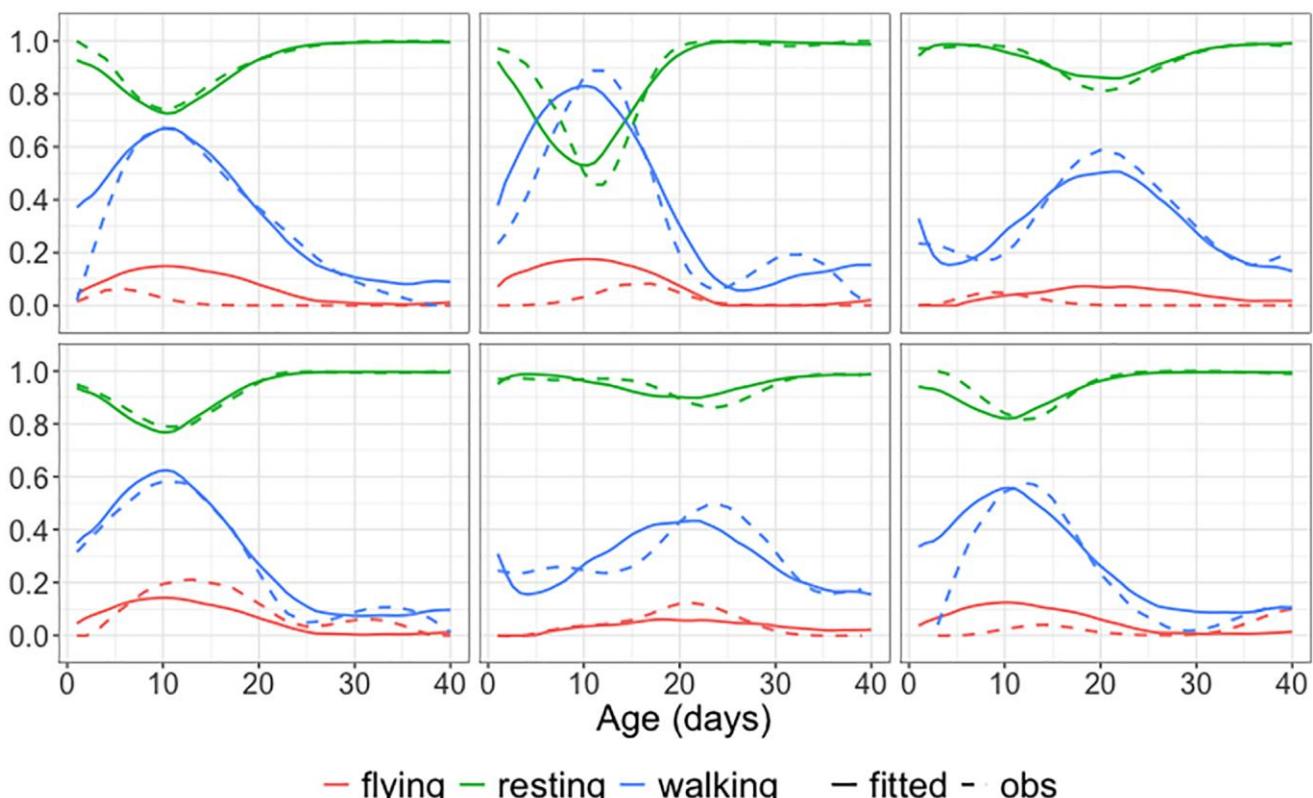
$$X(t) = [\sqrt{Y_1(t)}, \sqrt{Y_2(t)}, \sqrt{Y_3(t)}], \quad \text{where}$$

$Y_j(t) = Z_j(t) / (Z_1(t) + Z_2(t) + Z_3(t))$  for  $j = 1, 2, 3$ . For further details on compositional data and their representation on the positive orthant of a sphere, see [41–43]. This approach then leads to longitudinal data that are situated on a sphere.

### Spherical functional principal component analysis for longitudinal behavior data

Following [22] and [44], we first map the data to linear tangent spaces using Riemannian log maps centered at the Fréchet mean curve and then carry out a regular multivariate functional principal component analysis on the linear tangent space of the log-mapped data; spherical functional principal components, eigenfunctions and finite-truncated representations of the log-mapped data are first obtained on the tangent space and then mapped back to the original spherical space by applying Riemannian exp maps. An implementation is available through the R function *RFPCA*, available on GitHub at <https://github.com/CrossD/RFPCA>.

Figure 8 displays the observed and fitted trajectories  $X(t)$  for six randomly selected medflies, using the first three components obtained for the spherical functional principal component analysis. The close alignment between fitted and observed trajectories indicates a good fit. Figure 9 further illustrates the mean function and the first three eigenfunctions; each of these consists of three functions corresponding to the three behaviors. Figure 9 indicates that resting and walking were commonly observed, while flying occurred more rarely.



**Figure 8. Observed and fitted trajectories for age-varying behavioral data.** Observed data (solid lines) and spherical functional principal component analysis fitted behavioral trajectories (dashed lines) for six randomly selected medflies, with three selected components. The close alignment demonstrates a good fit.

The first three eigen components explain more than 96% of the total variation. The first eigenfunction reflects an overall contrast between resting and moving (flying or walking) throughout the observation period. The second eigenfunction reflects a contrast between early (0–15 days) and late (16–40 days) age periods, while the third eigenfunction features a contrast for both resting and walking between the middle age period (15–25 days) and the early/late age periods. Invoking modes of variation through this approach facilitates a comprehensive analysis of complex, time-varying behavioral patterns.

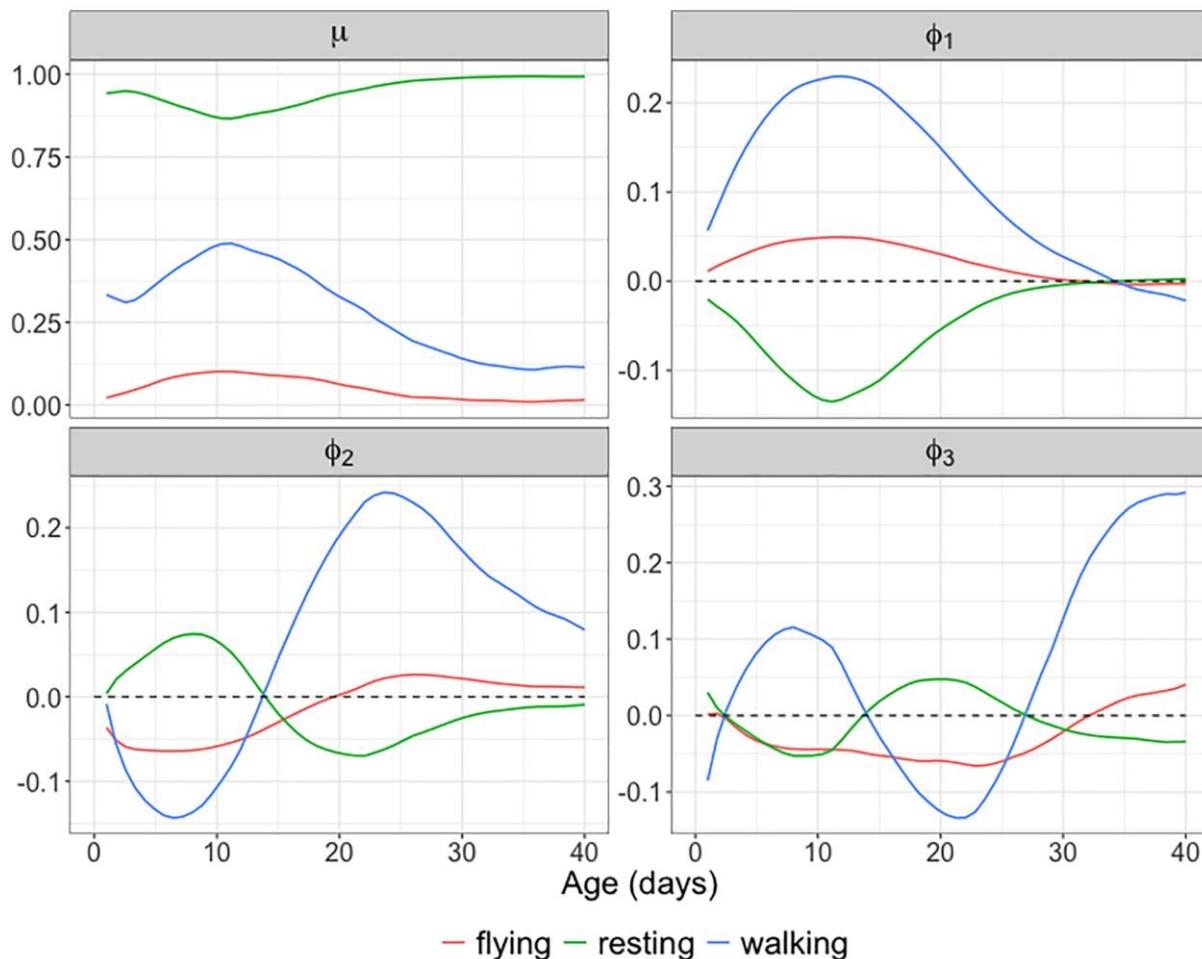
## Software

Several R packages are available for computational implementations of the methods illustrated in this perspective and beyond.

The R package *fdapace* [45] provides implementations of Functional Data Analysis and Empirical Dynamics. Central to this package is Functional Principal

Component Analysis, a key dimension reduction tool for functional data analysis, used for both sparsely and densely sampled random trajectories and time courses via the Principal Analysis by Conditional Estimation (PACE) algorithm. This core algorithm yields covariance and mean functions, eigenfunctions and principal component scores for functional data and their derivatives, covering both dense (functional) as well as sparse (longitudinal) sampling designs. Additionally, it provides fitted continuous trajectories with confidence bands for sparse designs, even for subjects with very few longitudinal observations, thereby presenting a viable and flexible alternative to more classical random effects modeling of longitudinal data. There is also a Matlab version (PACE) [25, 26, 46, 47].

The R package *frechet* [29] provides an implementation of statistical methods for random objects in various metric spaces. The core functionality of this package is Fréchet regression for random objects with Euclidean predictors, enabling regression analysis for non-Euclidean



**Figure 9. Mean function and eigenfunctions for longitudinal behavioral data.** The estimated mean function  $\hat{\mu}$  and the first three estimated spherical eigenfunctions  $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$  for the behavioral fly data using spherical functional principal component analysis. The first three eigenfunctions explain 96.54% of the total variation, with the individual components explaining 80.55%, 12.23% and 3.76%, respectively.

responses under mild conditions. As a special case, *frechet* includes the implementation of global and local Cox point process regression with point processes as responses and an intrinsically non-parametric intensity function [15]. This package is particularly useful for handling data such as distributions in 2-Wasserstein space, covariance matrices endowed with power metrics (including the Frobenius metric as a special case), and spherical data, among other data types [10].

The R package *fdaconcur* [40] provides an implementation of concurrent or varying coefficient regression methods for functional data, accommodating both densely and sparsely observed data. It includes the construction of pointwise confidence bands and models the influence of past predictor values through a smooth history index function [18].

The R package *fdarep* [33] is designed for repeatedly observed and thus dependent functional data, aiming at a framework where curves are recorded repeatedly for each subject in a sample. In particular, *fdarep* provides an implementation of two-dimensional functional principal component analysis (FPCA), Marginal FPCA, and Product FPCA for repeated functional data [14, 48].

## DISCUSSION

While we focus here on longitudinal data that relate to movement activity, egg-laying activity and behaviors that are continuously monitored, other longitudinal trajectories and their relationship with remaining lifetime distributions could be of equal interest. One issue is that one may encounter missing data or irregular designs with sporadic rather than continuous monitoring. Such irregular designs can be handled with the same methods as described above, and the software we refer to is also sufficiently flexible to handle irregularly recorded data. This is exemplified in [49], where functional data analysis methodology is shown to be superior to classical random effects models that traditionally have been used to handle missing or irregular and sparsely sampled longitudinal data. Various alternative implementations of functional principal component analysis are also available [50, 51].

Another monitoring design that is sometimes encountered generates functional snippets, where individuals are continuously monitored only over a subject-specific limited range of ages that is a subset of their lifespan, but not over the entire lifespan. This scenario is common for human data, where often only a relatively short period of monitoring is feasible and in any study where budgetary or logistic limitations allow only for a time-limited monitoring period. The age when an individual enters the monitoring period varies

by subject, as subjects of all ages are enrolled at inception of the study. Since the resulting functional snippet data convey less information compared to longitudinal monitoring over the entire lifespan and do not allow for direct auto-covariance estimation, which is a fundamental step for functional data analysis [13], only more restrictive approaches are available that require additional prior information and stronger assumptions [52–54].

For human data, the level and impact of movement activity on various health conditions was the topic of various studies [55–57]. Special care is needed to avoid confounding with factors such as socio-economic background that may bias the conclusions. Human studies have been predominantly cross-sectional rather than longitudinal and thus their emphasis is on cohort mean effects rather than individual random effects, where the latter is the emphasis of this article; for the relation between individual and cohort longevity see [58].

The statistical methodology illustrated in this article, centered around longitudinal and functional data analysis, distributional data analysis [59] and the more general framework of metric statistics and Fréchet regression [60], has also been employed for the analysis of samples of mortality trajectories (without monitoring data) [61–63]. Such data can be viewed as realizations of a stochastic mortality process and are easily obtained from various sources, including the Human Mortality Database (Max Planck Institute for Demographic Research, available at <http://www.mortality.org>).

## Data availability statement

The data supporting this study's findings are available from James R. Carey upon reasonable request (email address: [jrcarey@ucdavis.edu](mailto:jrcarey@ucdavis.edu)).

## APPENDIX

### Global Fréchet regression

We apply global Fréchet regression [10] with the remaining lifetime distribution observed at a given current alive age  $a$  as response. Here the space of distributions  $W$  is equipped with the 2-Wasserstein distance

$$d_W^2(F_1, F_2) = \int_0^1 \{F_1^{-1}(u) - F_2^{-1}(u)\}^2 du, \quad F_1, F_2 \in W$$

where  $F_1^{-1}$  and  $F_2^{-1}$  are the corresponding quantile functions of distribution functions  $F_1$  and  $F_2$ . Global Fréchet regression aims at conditional Fréchet means,

$$F_{x,\oplus} = \operatorname{argmin}_{v \in w} E[d_W^2(F, v) | X = x]. \quad (2)$$

Assume the random pairs  $\{(X_i, F_i)\}_{i=1}^n$  are realizations of  $(X, F)$ , where  $X$  is a scalar or vector predictor. In practice, we estimate the remaining lifetime distribution  $F_i$  from the sample of surviving medflies. The global Fréchet regression estimate is

$$\hat{F}_{x,\oplus} = \operatorname{argmin}_{v \in w} \frac{1}{n} \sum_{i=1}^n \hat{s}(X_i, x) d_W^2(\hat{F}_i, v), \quad (3)$$

where  $\hat{s}(X_i, x) = 1 + (X_i - \bar{X})^T \hat{\Sigma}^{-1} (x - \bar{X})$ ,  $\bar{X} = n^{-1} \sum_{i=1}^n X_i$  and  $\hat{\Sigma} = n^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$  are weights derived from a linear regression model. One can then obtain estimates for density and hazard function of the remaining lifetime distribution,

$$\hat{f}_{x,\oplus}(u) = \frac{\partial \hat{F}_{x,\oplus}(u)}{\partial u}, \quad \hat{\lambda}_{x,\oplus}(u) = \frac{\hat{f}_{x,\oplus}(u)}{1 - \hat{F}_{x,\oplus}(u)}, \quad u \in (0, 90). \quad (4)$$

### Product functional principal component analysis

Product functional principal component analysis (FPCA) serves as a dimension reduction tool for a function-valued stochastic process. If  $X(s, t)$  denote the activity level at hour  $s \in S = [0, 24]$  within day  $t \in T = [0, 32]$ , the mean and covariance functions of the underlying process are  $\mu(s, t) = E[X(s, t)]$  and  $G((s_1, t_1), (s_2, t_2)) = E[X(s_1, t_1)X(s_2, t_2) - \mu(s_1, t_1)\mu(s_2, t_2)]$ ,  $s_1, s_2 \in S, t_1, t_2 \in T$ .

Writing  $\{\psi_j(s)\}_{j=1}^\infty$  for the eigenfunctions of the operator in  $L^2(S)$  with marginal kernel  $G_S(s, u) = \int_0^{32} G((s, t), (u, t)) dt$  and  $\{\phi_k(t)\}_{k=1}^\infty$  for the eigenfunctions of the operator in  $L^2(T)$  with marginal kernel  $G_T(t, u) = \int_0^{24} G((s, t), (s, u)) ds$ , one can represent  $X(s, t)$  through a product FPCA as follows,

$$X(s, t) - \mu(s, t) = \sum_{j=1}^\infty \sum_{k=1}^\infty \chi_{jk} \phi_k(t) \psi_j(s). \quad (5)$$

Here  $\chi_{jk} = \int_0^{32} \int_0^{24} (X(s, t) - \mu(s, t)) \psi_j(s) \phi_k(t) ds dt$  are the principal component scores that are used to summarize the two-dimensional process. These components are zero mean uncorrelated random variables representing the fluctuations of the process  $X(s, t)$  around the mean function  $\mu(s, t)$ .

### Global Cox point process model

Writing  $N(t)$  for the point process that represents egg-laying at age  $t$ , we model  $(N, \Lambda)$  as a doubly stochastic Poisson process, where we postulate an underlying stochastic (positive integrable) intensity process  $\Lambda(t)$ , such that  $N|\Lambda = \lambda$  is a non-homogeneous Poisson process with intensity function  $\lambda$  [15, 64]. Given a realization of the latent process  $\Lambda = \lambda$ , the expected number of eggs laid up to time  $t$  is  $E[N(t)|\Lambda = \lambda] = \int_0^t \lambda(u) du$ . Conditional on observing  $N(T) = m > 0$  events with associated intensity function  $\Lambda = \lambda$ , the random event times  $T_1, \dots, T_m$  at which egg-laying occurs are independently and identically distributed.

To regress the infinite-dimensional object  $\Lambda$  on age-at-death  $X$ , we fit a global Cox point process regression model, targeting

$$m_\oplus(x) = \operatorname{argmin}_{\lambda \in \Omega} E[d^2(\Lambda, \lambda) | X = x], \quad (6)$$

where  $(\Omega, d)$  forms a metric space of intensity functions and  $\Lambda \in \Omega$  represents the random intensity function of egg-laying events. Intensity functions  $\Lambda \in \Omega$  admit a one-to-one decomposition into two components, namely, the intensity factor  $\tau = \int_0^T \Lambda(s) ds$  and the density function  $f(t) = \frac{\Lambda}{\tau}$ . Thus, the intensity function space  $\Omega$  can be viewed as the product metric space  $\Omega = D \times \Omega_s$ , where  $D$  denotes the space of density functions over  $[0, T]$  and  $\Omega_s = (0, \infty)$  denotes the space of intensity factors. We utilize the  $\ell^2$  type product metric  $d$  between intensities  $\Lambda_1 = (f_1, \tau_1)$  and  $\Lambda_2 = (f_2, \tau_2)$ , which is given by

$$d(\Lambda_1, \Lambda_2) = (d_W^2(f_1, f_2) + d_E^2(\tau_1, \tau_2))^{1/2},$$

Here  $d_E$  is the Euclidean metric and  $d_W$  is the Wasserstein metric between probability distributions. While the intensity function of egg-laying events remains unobservable, its components, namely the density function and the intensity factor, can be estimated from available data.

### AUTHOR CONTRIBUTIONS

All authors participated in the writing and finalization of the paper. Methodology and data analysis were jointly developed by Su I Iao, Poorbita Kundu and Hans-Georg Müller, where the latter developed the concepts and framing of the paper. The medfly data were provided by James Carey.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

## FUNDING

This work was funded in part by a grant from the Center for the Economics and Demography of Aging, UC Berkeley (NIH 2P30AG012839) and in part by NSF grant DMS-2310450.

## REFERENCES

1. Fishman EI, Steeves JA, Zipunnikov V, Koster A, Berrigan D, Harris TA, Murphy R. Association between Objectively Measured Physical Activity and Mortality in NHANES. *Med Sci Sports Exerc.* 2016; 48:1303–11.  
<https://doi.org/10.1249/MSS.0000000000000885>  
PMID:[26848889](#)
2. Tabacu L, Ledbetter M, Leroux A, Crainiceanu C, Smirnova E. Quantifying the varying predictive value of physical activity measures obtained from wearable accelerometers on all-cause mortality over short to medium time horizons in NHANES 2003–2006. *Sensors (Basel).* 2020; 21:4.  
<https://doi.org/10.3390/s21010004>  
PMID:[33374911](#)
3. Yang Y, Wang MC. Analyzing wearable device data using marked point processes. *Biometrics.* 2021; 77:54–66.  
<https://doi.org/10.1111/biom.13269>  
PMID:[32282947](#)
4. Hilden P, Schwartz JE, Pascual C, Diaz KM, Goldsmith J. How many days are needed? Measurement reliability of wearable device data to assess physical activity. *PLoS One.* 2023; 18:e0282162.  
<https://doi.org/10.1371/journal.pone.0282162>  
PMID:[36827427](#)
5. Schrack J, Zhang T, Wanigatunga A, Zipunnikov V, and Freedman V. Accelerometer-derived physical activity in the national health and aging trends study. *Innovation in Aging.* 2023; 7:165–6.  
<https://doi.org/10.1093/geroni/igad104.0542>
6. Chen H, Müller HG, Rodovitis VG, Papadopoulos NT, Carey JR. Daily activity profiles over the lifespan of female medflies as biomarkers of aging and longevity. *Aging Cell.* 2024; 23:e14080.  
<https://doi.org/10.1111/ace.14080>  
PMID:[38268242](#)
7. Carey JR, Liedo P, Müller HG, Wang JL, Chiou JM. Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of Mediterranean fruit fly females. *J Gerontol A Biol Sci Med Sci.* 1998; 53:B245–51.  
<https://doi.org/10.1093/gerona/53a.4.b245>  
PMID:[18314553](#)
8. Carey JR, Papadopoulos N, Kouloussis N, Katsoyannos B, Müller HG, Wang JL, Tseng YK. Age-specific and lifetime behavior patterns in *Drosophila melanogaster* and the Mediterranean fruit fly, *Ceratitis capitata*. *Exp Gerontol.* 2006; 41:93–7.  
<https://doi.org/10.1016/j.exger.2005.09.014>  
PMID:[16297586](#)
9. Müller HG, Zhang Y. Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics.* 2005; 61:1064–75.  
<https://doi.org/10.1111/j.1541-0420.2005.00378.x>  
PMID:[16401280](#)
10. Petersen A, Müller HG. Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics.* 2019; 47:691–719.  
<https://doi.org/10.48550/arXiv.1608.03012>
11. Castro PE, Lawton WH, Sylvestre EA. Principal modes of variation for processes with continuous sample curves. *Technometrics.* 1986; 28:329–37.  
<https://doi.org/10.2307/1268982>
12. Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc Series B Stat Methodol.* 1991; 53:233–43.  
<https://www.jstor.org/stable/2345738>
13. He G, Müller HG, Wang JL. Extending correlation and regression from multivariate to functional data. *Asymptotics in statistics and probability.* 2000; 197–210.  
<https://doi.org/10.1515/9783110942002-015>
14. Chen K, Delicado P, Müller HG. Modelling function-valued stochastic processes, with applications to fertility dynamics. *J R Stat Soc Series B Stat Methodol.* 2017; 79:177–96.  
<https://doi.org/10.1111/rssb.12160>
15. Gajardo Á, Müller HG. Cox point process regression. *IEEE Trans Inf Theory.* 2021; 68:1133–56.  
<https://doi.org/10.1109/TIT.2021.3126466>
16. Cardot H, Ferraty F, Mas A, Sarda P. Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics.* 2003; 30:241–55.  
<https://doi.org/10.1111/1467-9469.00329>
17. Müller HG, Carey JR, Wu D, Liedo P, Vaupel JW. Reproductive potential predicts longevity of female Mediterranean fruitflies. *Proc Biol Sci.* 2001; 268:445–50.  
<https://doi.org/10.1098/rspb.2000.1370>  
PMID:[11296855](#)

18. Şentürk D, Müller HG. Functional varying coefficient models for longitudinal data. *J Am Stat Assoc.* 2010; 105:1256–64.  
<https://doi.org/10.1198/jasa.2010.tm09228>

19. Zhang Y, Müller HG, Carey JR, Papadopoulos NT. Behavioral trajectories as predictors in event history analysis: male calling behavior forecasts medfly longevity. *Mech Ageing Dev.* 2006; 127:680–6.  
<https://doi.org/10.1016/j.mad.2006.04.001>  
 PMID:[16730781](#)

20. Anotaux M, Marchal J, Châline N, Desquillet L, Leborgne R, Gilbert C, Pasquet A. Ageing alters spider orb-web construction. *Animal Behaviour.* 2012; 84:1113–21.  
<https://doi.org/10.1016/j.anbehav.2012.08.017>

21. Overman KE, Choi DM, Leung K, Shaevitz JW, Berman GJ. Measuring the repertoire of age-related behavioral changes in *Drosophila melanogaster*. *PLoS Comput Biol.* 2022; 18:e1009867.  
<https://doi.org/10.1371/journal.pcbi.1009867>  
 PMID:[35202388](#)

22. Dai X, Müller HG. Principal component analysis for functional data on Riemannian manifolds and spheres. *The Annals of Statistics.* 2018; 46:333461.  
<https://doi.org/10.1214/17-AOS1660>

23. Davidian M, Lin X, Wang JL. Introduction: Emerging issues in longitudinal and functional data analysis. *Statistica Sinica.* 2004; 14:613–14.

24. Horvath L, Kokoszka P. *Inference for Functional Data with Applications*. Springer, NY. 2012.  
<https://doi.org/10.1007/978-1-4614-3655-3>

25. Chen K, Zhang X, Petersen A, Müller HG. Quantifying infinite-dimensional data: Functional data analysis in action. *Stat Biosci.* 2017; 9:582–604.  
<https://doi.org/10.1007/s12561-015-9137-5>

26. Wang JL, Chiou JM, Müller HG. Functional data analysis. *Annual Review of Statistics and Its Application.* 2016; 3:257–95.  
<https://doi.org/10.1146/annurev-statistics-041715-033624>

27. Müller HG, Wang JL, Capra WB. From lifetables to hazard rates: The transformation approach. *Biometrika.* 1997; 84:881–92.  
<https://www.jstor.org/stable/2337658>

28. Wang JL, Müller HG, Capra WB. Analysis of oldest-old mortality: Lifetables revisited. *The Annals of Statistics.* 1998; 26:126–63.  
<https://doi.org/10.1214/aos/1030563980>

29. Chen Y, Zhou Y, Chen H, Gajardo Á, Fan J, Zhong Q, Dubey P, Han K, Bhattacharjee S, Zhu C, Iao SI, Kundu P, Petersen A, Müller HG. *frechet: Statistical analysis for random objects and non-Euclidean data.* R package version 0.3.0. 2023.

30. Kleffe J. Principal components of random variables with values in a separable Hilbert space. *Mathematische Operationsforschung und Statistik.* 1973; 4:391–406.  
<https://doi.org/10.1080/02331887308801137>

31. Coffey N, Harrison AJ, Donoghue OA, Hayes K. Common functional principal components analysis: a new approach to analyzing human movement data. *Hum Mov Sci.* 2011; 30:1144–66.  
<https://doi.org/10.1016/j.humov.2010.11.005>  
 PMID:[21543128](#)

32. Han K, Hadjipantelis PZ, Wang JL, Kramer MS, Yang S, Martin RM, Müller HG. Functional principal component analysis for identifying multivariate patterns and archetypes of growth, and their association with long-term cognitive development. *PLoS One.* 2018; 13:e0207073.  
<https://doi.org/10.1371/journal.pone.0207073>  
 PMID:[30419052](#)

33. Kundu P, Zhu C, Chen K, Delicado P, Iao SI, Zhou H, Chen H, Cui M, Müller HG. *fdarep: Two-dimensional FPCA, marginal FPCA, and product FPCA for repeated functional data.* R package version 0.1.1. 2024.

34. Müller HG, Wang JL, Capra WB, Liedo P, Carey JR. Early mortality surge in protein-deprived females causes reversal of sex differential of life expectancy in Mediterranean fruit flies. *Proc Natl Acad Sci U S A.* 1997; 94:2762–5.  
<https://doi.org/10.1073/pnas.94.6.2762>  
 PMID:[9122268](#)

35. Harshman LG, Zera AJ. The cost of reproduction: the devil in the details. *Trends Ecol Evol.* 2007; 22:80–6.  
<https://doi.org/10.1016/j.tree.2006.10.008>  
 PMID:[17056152](#)

36. Papadopoulos NT, Liedo P, Müller HG, Wang JL, Molleman F, Carey JR. Cost of reproduction in male medflies: the primacy of sexual courting in extreme longevity reduction. *J Insect Physiol.* 2010; 56:283–7.  
<https://doi.org/10.1016/j.jinsphys.2009.10.014>  
 PMID:[19896949](#)

37. Meng X, Hu J, Plant RE, Carpenter TE, Carey JR. Distinctive egg-laying patterns in terminal versus non-terminal periods in three fruit fly species. *Exp Gerontol.* 2021; 145:111201.  
<https://doi.org/10.1016/j.exger.2020.111201>  
 PMID:[33316371](#)

38. Carey JR, Liedo P, Müller HG, Wang JL. A simple graphical technique for displaying individual fertility data and cohort survival: case study of 1000

Mediterranean fruit fly females. *Functional Ecology*. 1998; 12:359–63.  
<https://doi.org/10.1046/j.1365-2435.1998.00197.x>

39. Dubey P, Müller HG. Modeling time-varying random objects and dynamic networks. *Journal of the American Statistical Association*. 2022; 117:2252–67.  
<https://doi.org/10.1080/01621459.2021.1917416>

40. Iao SI, Bhattacharjee S, Chen Y, Zhu C, Chen H, Zhou Y, Gajardo Á, Kundu P, Zhou H, Müller HG. fdaconcur: Concurrent regression and history index models for functional data. R package version 0.1.3. 2024.

41. Aitchison J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1982; 44:139–77.  
<https://www.jstor.org/stable/2345821>

42. Scealy JL, Welsh AH. Regression for compositional data by using distributions defined on the hypersphere. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2011; 73:351–75.  
<https://doi.org/10.1111/j.1467-9868.2010.00766.x>

43. Scealy JL, Welsh AH. Colours and cocktails: Compositional data analysis 2013 Lancaster lecture. *Australian & New Zealand Journal of Statistics*. 2014; 56:145–69.  
<https://doi.org/10.1111/anzs.12073>

44. Dai X, Lin Z, Müller HG. Modeling sparse longitudinal data on Riemannian manifolds. *Biometrics*. 2021; 77:1328–41.  
<https://doi.org/10.1111/biom.13385>  
PMID:33034049

45. Zhou Y, Chen H, Iao SI, Kundu P, Zhou H, Bhattacharjee S, Carroll C, Chen Y, Dai X, Fan J, Gajardo Á, Hadjipantelis PZ, Han K, et al. fdapace: Functional data analysis and empirical dynamics. R package version 0.6.0. 2024.

46. Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*. 2005; 100:577–90.  
<https://doi.org/10.1198/016214504000001745>

47. Yao F, Müller HG, Wang JL. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*. 2005; 33:2873–903.  
<https://doi.org/10.48550/arXiv.math/0603132>

48. Chen K, Müller HG. Modeling repeated functional observations. *Journal of the American Statistical Association*. 2012; 107:1599–609.  
<https://doi.org/10.1080/01621459.2012.734196>

49. Chen Y, Dubey P, Müller HG, Bruchhage M, Wang JL, Deoni S. Modeling sparse longitudinal data in early neurodevelopment. *Neuroimage*. 2021; 237:118079.  
<https://doi.org/10.1016/j.neuroimage.2021.118079>  
PMID:34000395

50. Chiou JM, Chen YT, Yang YF. Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*. 2014; 24:1571–96.  
<http://dx.doi.org/10.5705/ss.2013.305>

51. Cui E, Li R, Crainiceanu CM, Xiao L. Fast Multilevel Functional Principal Component Analysis. *J Comput Graph Stat*. 2023; 32:366–77.  
<https://doi.org/10.1080/10618600.2022.2115500>  
PMID:37313008

52. Delaigle A, Hall P, Huang W, Kneip A. Estimating the covariance of fragmented and other related types of functional data. *Journal of the American Statistical Association*. 2021; 116:1383–401.  
<https://doi.org/10.1080/01621459.2020.1723597>

53. Lin Z, Wang JL, Zhong Q. Basis expansions for functional snippets. *Biometrika*. 2021; 108:709–26.  
<https://doi.org/10.1093/biomet/asaa088>

54. Lin Z, Wang JL. Mean and Covariance Estimation for Functional Snippets. *J Am Stat Assoc*. 2022; 117:348–60.  
<https://doi.org/10.1080/01621459.2020.1777138>  
PMID:35757778

55. Varma VR, Dey D, Leroux A, Di J, Urbanek J, Xiao L, Zipunnikov V. Re-evaluating the effect of age on physical activity over the lifespan. *Prev Med*. 2017; 101:102–8.  
<https://doi.org/10.1016/j.ypmed.2017.05.030>  
PMID:28579498

56. Hamm JM, Parker K, Lachman ME, Mogle JA, Duggan KA, McGrath R. Increased frequency of light physical activity during midlife and old age buffers against cognitive declines. *J Behav Med*. 2024; 47:622–34.  
<https://doi.org/10.1007/s10865-024-00478-2>  
PMID:38429599

57. Ai M, Morris TP, Noriega de la Colina A, Thovinakere N, Tremblay-Mercier J, Villeneuve S, H Hillman C, Kramer AF, Geddes MR, and PREVENT-AD Research Group. Midlife physical activity engagement is associated with later-life brain health. *Neurobiol Aging*. 2024; 134:146–59.  
<https://doi.org/10.1016/j.neurobiolaging.2023.11.004>  
PMID:38091752

58. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979; 16:439–54.  
PMID:510638

59. Petersen A, Zhang C, Kokoszka P. Modeling probability density functions as data objects. *Econometrics and Statistics*. 2022; 21:159–78.  
<https://doi.org/10.1016/j.ecosta.2021.04.004>

60. Dubey P, Chen Y, Müller HG. Metric statistics: Exploration and inference for random objects with distance profiles. *The Annals of Statistics*. 2024; 52:757–92.  
<https://doi.org/10.1214/24-AOS2368>

61. Chiou JM, Müller HG. Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *Journal of the American Statistical Association*. 2009; 104:572–85.  
<https://doi.org/10.1198/jasa.2009.0023>

62. Shang HL. Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method. *The Annals of Applied Statistics*. 2016; 10:1639–72.  
<https://doi.org/10.1214/16-AOAS953>

63. Stefanucci M, Mazzuco S. Analysing cause-specific mortality trends using compositional functional data analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 2022; 185:61–83.  
<https://doi.org/10.1111/rssa.12715>

64. Cox DR, Isham V. *Point Processes*. CRC Press. 1980.  
<https://doi.org/10.1201/9780203743034>