



Integrated Degradation Models in R Using iDEMO

Ya-Shan Cheng
Academia Sinica

Chien-Yu Peng
Academia Sinica

Abstract

Degradation models are widely used to assess the lifetime information for highly reliable products with quality characteristics whose degradation over time can be related to reliability. The performance of a degradation model largely depends on an appropriate model description of the product's degradation path. The cross-platform package **iDEMO** (integrated degradation models) is developed in R and the interface is built using the Tcl/Tk bindings provided by the **tcltk** and **tcltk2** packages included with R. It is a tool to build a linear degradation model which can simultaneously consider the unit-to-unit variation, time-dependent structure and measurement error in the degradation paths. The package **iDEMO** provides the maximum likelihood estimates of the unknown parameters, mean-time-to-failure and q -th quantile, and their corresponding confidence intervals based on the different information matrices. In addition, degradation model selection and goodness-of-fit tests are provided to determine and diagnose the degradation model for the user's current data by the commonly used criteria. By only enabling user interface elements when necessary, input errors are minimized.

Keywords: degradation tests, first passage time, highly reliable products, graphical user interface, mean-time-to-failure, quantile, R, Tcl/Tk.

1. Introduction

High quality products are designed and manufactured to function for a long time before they fail. Hence, with only a relatively short period of time available for internal life testing, it is a great challenge for manufacturers to obtain reliability information for their products. Although there are helpful techniques such as censoring and/or accelerating a product's lifetime by testing at a higher level of stress, these techniques offer little assistance for highly reliable products. The major obstacle is the rather difficult problem of obtaining sufficient time-to-failure data to efficiently estimate a product's lifetime. Under such a restraint, if quality characteristics (QCs) do exist whose degradation over time (also called degradation

paths) can be related to reliability, an alternative approach can then be used to collect sufficient degradation data to more accurately predict the product's lifetime distribution. General references for degradation models can be found in Nelson (1990), Meeker and Escobar (1998) and Chao (1999). Other important research applications of degradation models are in areas such as micro-electronic components, laser diode, food and drugs, metal fatigue testing, plasma display panels, liquid crystal display and digital light processing projectors and other dependable systems.

The successful performance of a degradation test strongly depends on the appropriateness of the modeling of a product's degradation path. Conventionally, random effects and stochastic process formulations are two well-known approaches presented in the literature. The first approach assumes that variations in degradation data are mainly from the mean degradation paths of test units. General references for this approach are Lu and Meeker (1993), Lu, Park, and Yang (1997), Bae and Kvam (2004), and Park and Bae (2010). However, when the unit-to-unit variations of test units are insignificant, the second approach considers a fixed effect mean degradation path and variations in degradation data that are mainly described by time-dependent stochastic processes. Among them, the Wiener (Brownian motion) and gamma processes are two of the most well-known models. Typical examples using this approach are also included in Doksum and Høyland (1992), Tseng and Peng (2004), Lawless and Crowder (2004) and Peng (2012). Apparently, the stochastic process formulation has a suitable treatment for time-dependent structures within degradation paths. However, the unit-to-unit variation with time-dependent structures and measurement error may occur simultaneously in the degradation paths.

Recently, Peng and Tseng (2009) incorporated the random effects and the stochastic process model into a degradation model, which can simultaneously consider the unit-to-unit variation with time-dependent structure and measurement error. However, using the previous degradation model to fit degradation data may be challenging for those who use R (R Development Core Team 2012) only occasionally. To the best of our knowledge, there is no R package oriented to the computation of the lifetime information for such a degradation model. Hence, the package **iDEMO** introduced in this paper fills this gap and the user requires almost no knowledge of R syntax. The cross-platform package **iDEMO** is developed in R for statistical computing and the graphical user interface (GUI) is built using the Tcl/Tk bindings provided by the **tcltk** and **tcltk2** packages included with R. For R command-line interface, the user needs to remember the function names and arguments of commands. The GUI of **iDEMO** makes analysis of degradation data easier for engineers and researchers, shortening the period of time it takes to learn how to use R and decreasing the chances of syntax and typing errors. Furthermore, we use the disabled property in the Tcl/Tk widgets to minimize input errors. Outputs and figures are directly exported to the R console for further manipulation. The R package implementing the methodology described in this paper is available from the Comprehensive R Archive Network at <http://CRAN.R-project.org/package=iDEMO>. More discussions and applications of GUI for R can be found in Valero-Mora and Ledesma (2012), Fox and Carvalho (2012), Fellows (2012), Rödiger, Friedrichsmeier, Kapat, and Michalke (2012) and the references given therein.

The remainder of this paper is organized as follows. Section 2 briefly introduces the degradation model and the goals that **iDEMO** is designed to meet. Section 3 describes the functions of **iDEMO** and the usage of **iDEMO** is illustrated through the analysis of real data. Finally, concluding remarks and future possible extensions of the package are given.

2. Statistical overview: Degradation models

Peng and Tseng (2009) proposed the following model to describe the degradation data. Let $Y(t)$ and $L(t)$, $t \geq 0$, denote the observed and the true values of the QC of a product at time t , respectively. Assume that there exists a transformation function such that

$$M_0 : \begin{cases} Y(t) = L(t) + \sigma_\epsilon \epsilon, \\ L(t) = \Theta t + \sigma_B B(t), \end{cases}$$

where Θ follows a normal distribution with mean η and variance σ_η^2 (denoted by $\mathcal{N}(\eta, \sigma_\eta^2)$) representing the unit-to-unit variation of the products; ϵ is the measurement error with $\mathcal{N}(0, 1)$; σ_B is a diffusion coefficient; and $B(t)$ denotes standard Brownian motion representing a time-correlated structure. The random effect Θ , the standard Brownian motion $B(t)$ and the measurement error ϵ are assumed to be mutually independent. For this degradation model, similar to Lu and Meeker (1993, p. 164), we assume that $P\{\Theta \leq 0\}$ is negligible in order to avoid the certain probability of getting non-feasible degradation slopes. The advantage of degradation model M_0 is that it allows us to simultaneously consider the unit-to-unit variation with time-dependent structure and measurement error. It is easily seen that if we set the cases (i) $\sigma_B = 0$, (ii) $\sigma_\eta = \sigma_\epsilon = 0$ and (iii) $\sigma_\epsilon = 0$ in the degradation model M_0 , then the model M_0 reduces to (i) the conventional random-effect model M_1 , (ii) the Wiener process model M_2 and (iii) the degradation model M_3 without measurement error, respectively, which had been proposed in Peng and Tseng (2009). In addition to the models M_0 to M_3 , two more degradation models M_4 and M_5 are now available in **iDEMO** as shown in Table 1. Note that the model M_4 is the traditional regression and used as a benchmark to compare with the other competing degradation models. Furthermore, in order for statistical inference to be possible, the identifiability of the degradation model M_0 is needed and can be established as follows.

Proposition 1. *If the measure frequencies are greater than 3, then the degradation model M_0 is identifiable.*

The proof of Proposition 1 is presented in Appendix A. This means that different values of the parameter in the degradation model M_0 must generate different probability distributions of the observable variables. See Lehmann and Casella (1998), and Casella and Berger (2001) for details.

Model	Variation sources			
	η	σ_η	σ_B	σ_ϵ
M_0	✓	✓	✓	✓
M_1	✓	✓		✓
M_2	✓		✓	
M_3	✓	✓	✓	
M_4	✓			✓
M_5	✓		✓	✓

Table 1: Degradation models in **iDEMO**.

2.1. Lifetime information

Let ω denote the critical level for the degradation path of model M_0 . The product's lifetime T can then be suitably defined as the first passage time when the true degradation path $L(t)$ crosses the critical level ω ; i.e., $T = \inf\{t | L(t) \geq \omega\}$. The probability density function (PDF) and cumulative distribution function (CDF) of the lifetime distribution T are respectively

$$f_T(t) = \sqrt{\frac{\omega^2}{2\pi t^3(\sigma_\eta^2 t + \sigma_B^2)}} \exp\left\{-\frac{(\omega - \eta t)^2}{2t(\sigma_\eta^2 t + \sigma_B^2)}\right\}, \quad t \geq 0, \quad (1)$$

and

$$F_T(t) = \Phi\left(\frac{\eta t - \omega}{\sqrt{\sigma_B^2 t + \sigma_\eta^2 t^2}}\right) + \exp\left(\frac{2\eta\omega}{\sigma_B^2} + \frac{2\sigma_\eta^2\omega^2}{\sigma_B^4}\right) \Phi\left(-\frac{2\sigma_\eta^2\omega t + \sigma_B^2(\eta t + \omega)}{\sigma_B^2\sqrt{\sigma_B^2 t + \sigma_\eta^2 t^2}}\right), \quad (2)$$

where $\Phi(\cdot)$ is the CDF of $\mathcal{N}(0, 1)$. Note that if the drift rate is $\eta \geq 0$ for the degradation models M_2 and M_5 , we readily perceive that $\lim_{t \rightarrow \infty} F_T(t) = 1$ as derived from (2). The q -th quantile, $t(q)$, of a product's lifetime can then be evaluated by solving $F_T(t(q)) = q$. However, if $\eta \geq 0$ for the degradation models M_0 , M_1 and M_3 , and $\eta < 0$ for the degradation models M_0 - M_3 and M_5 , the density function in (1) is improper with the positive probability at infinity and the first passage time has a defective distribution. i.e.,

$$\begin{aligned} \mathbb{P}\{T = \infty\} &= 1 - \lim_{t \rightarrow \infty} F_T(t) \\ &= 1 - \Phi\left(\frac{\eta}{\sigma_\eta}\right) - \exp\left(\frac{2\eta\omega}{\sigma_B^2} + \frac{2\sigma_\eta^2\omega^2}{\sigma_B^4}\right) \Phi\left(-\frac{2\sigma_\eta\omega}{\sigma_B^2} - \frac{\eta}{\sigma_\eta}\right). \end{aligned}$$

For instance, by using L'Hospital rule for the model M_1 , it is easily shown that

$$\lim_{\sigma_B \rightarrow 0} \lim_{t \rightarrow \infty} F_T(t) = \Phi(\eta/\sigma_\eta).$$

For this scenario, it is impossible to estimate each and every q -th quantile of a product's lifetime because of $\lim_{t \rightarrow \infty} F_T(t) \in (0, 1)$. If $q \in (0, \lim_{t \rightarrow \infty} F_T(t))$, the q -th quantile of a product's lifetime can still be estimated. Otherwise, instead of the q -th quantile, the product's mean-time-to-failure (*MTTF*) is an alternative measure associated with the product's lifetime information. For more discussion of the negative drift rate, see [Whitmore \(1986\)](#), [Chhikara and Folks \(1989\)](#) and the references therein.

For the model M_0 , *MTTF* is given by

$$\mathbb{E}(T) = \frac{\sqrt{2}\omega}{\sigma_\eta} \mathcal{D}\left(\frac{\eta}{\sqrt{2}\sigma_\eta}\right) \quad (3)$$

$$\approx \frac{\omega}{\eta}, \quad \text{as } \eta \gg \sigma_\eta, \quad (4)$$

where $\mathcal{D}(z) = \exp(-z^2) \int_0^z \exp(x^2) dx$ is Dawson's integral for all real z . See [Peng \(2008\)](#), [Peng and Tseng \(2009\)](#) and [Peng and Hsu \(2012\)](#) for more details.

In practical application however, the parameters in the degradation model M_0 are unknown. Therefore, in order to assess the product's lifetime information, the likelihood function of the degradation model M_0 can be derived in the following section.

2.2. Log-likelihood function and model selection criteria

Assume that n units are tested, and the degradation measurements of each unit are available at time $t_1, \dots, t_m$ in a degradation test. The sample path of the i -th unit at time t_j is given by

$$Y_i(t_j) = \Theta_i t_j + \sigma_B B(t_j) + \sigma_\epsilon \epsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m.$$

For simplicity, let $\mathbf{t} = (t_1, \dots, t_m)^\top$, $\mathbf{Y}_i = (Y_i(t_1), \dots, Y_i(t_m))^\top$ and $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_n^\top)^\top$. Thus $\mathbf{Y}_i$ follows a multivariate normal distribution with mean vector $\eta \mathbf{t}$, and covariance matrix

$$\Sigma = \sigma_\eta^2 \mathbf{t} \mathbf{t}^\top + \Omega, \quad (5)$$

where

$$\Omega = \sigma_B^2 \mathbf{Q} + \sigma_\epsilon^2 \mathbf{I}_m, \quad \mathbf{Q} = [\min\{t_i, t_j\}]_{1 \leq i, j \leq m}, \quad (6)$$

and $\mathbf{I}_m$ is an identity matrix of order m . The log-likelihood function of $\boldsymbol{\vartheta} = (\eta, \sigma_\eta^2, \sigma_B^2, \sigma_\epsilon^2)^\top$ is given by

$$\mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}) = -\frac{nm}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \eta \mathbf{t})^\top \Sigma^{-1} (\mathbf{y}_i - \eta \mathbf{t}).$$

Hence, the maximum likelihood estimates (MLEs) $\hat{\boldsymbol{\vartheta}}$ of all unknown parameters can be found numerically by using the function `optim()` in R. Moreover, let r and $\mathcal{L}(\hat{\boldsymbol{\vartheta}}|\mathbf{y})$ be the number of parameters of a degradation model and the value of the log-likelihood function, respectively. For degradation models M_0 to M_5 , Akaike information criterion (AIC), Bayesian information criterion (BIC) and Hannan-Quinn criterion (HQC) are adopted for model selection and defined as $\text{AIC} = -2\mathcal{L}(\hat{\boldsymbol{\vartheta}}|\mathbf{y}) + 2r$, $\text{BIC} = -2\mathcal{L}(\hat{\boldsymbol{\vartheta}}|\mathbf{y}) + r \ln(n)$ and $\text{HQC} = -2\mathcal{L}(\hat{\boldsymbol{\vartheta}}|\mathbf{y}) + 2r \ln(\ln(n))$, respectively. The degradation model with minimum AIC (BIC or HQC) is chosen as the best model to fit the degradation data.

2.3. Information matrix type

For a given function g with continuous first partial derivatives, let $g(\boldsymbol{\vartheta})$ be the quantity of interest and $\nabla g(\boldsymbol{\vartheta}) = \partial g(\boldsymbol{\vartheta}) / \partial \boldsymbol{\vartheta}$. Then, under the large sample case, we have

$$\sqrt{n}(g(\hat{\boldsymbol{\vartheta}}) - g(\boldsymbol{\vartheta})) \xrightarrow{d} \mathcal{N}\left(0, \text{Avar}(g(\hat{\boldsymbol{\vartheta}}))\right),$$

where

$$\text{Avar}(g(\hat{\boldsymbol{\vartheta}})) = (\nabla g(\boldsymbol{\vartheta}))^\top \mathcal{I}^{-1}(\boldsymbol{\vartheta}) (\nabla g(\boldsymbol{\vartheta})),$$

and $\mathcal{I}(\boldsymbol{\vartheta})$ denotes expected Fisher's information matrix (FIM). That is

$$\mathcal{I}(\boldsymbol{\vartheta}) = \begin{bmatrix} \mathcal{I}_{\eta, \eta} & \mathcal{I}_{\eta, \sigma_\eta^2} & \mathcal{I}_{\eta, \sigma_B^2} & \mathcal{I}_{\eta, \sigma_\epsilon^2} \\ & \mathcal{I}_{\sigma_\eta^2, \sigma_\eta^2} & \mathcal{I}_{\sigma_\eta^2, \sigma_B^2} & \mathcal{I}_{\sigma_\eta^2, \sigma_\epsilon^2} \\ & & \mathcal{I}_{\sigma_B^2, \sigma_B^2} & \mathcal{I}_{\sigma_B^2, \sigma_\epsilon^2} \\ & & & \mathcal{I}_{\sigma_\epsilon^2, \sigma_\epsilon^2} \end{bmatrix},$$

where $\mathcal{I}_{\vartheta_k, \vartheta_l} = -\mathbb{E}(\partial^2 \mathcal{L}(\boldsymbol{\vartheta} | \mathbf{y}_i) / \partial \vartheta_k \partial \vartheta_l)$ for $1 \leq k, l \leq 4$. For instance, the approximate variance of the estimated *MTTF* ($\widehat{MTTF}$) and the estimated q -th quantile ($\hat{t}(q)$) of the product can be obtained respectively as

$$\text{Avar}(\widehat{MTTF}) = (\nabla MTTF)^\top \mathcal{I}^{-1}(\boldsymbol{\vartheta}) (\nabla MTTF)$$

and

$$\text{Avar}(\hat{t}(q)) = \frac{1}{(f_T(t(q)))^2} (\nabla F_T(t(q)))^\top \mathcal{I}^{-1}(\boldsymbol{\vartheta}) (\nabla F_T(t(q))).$$

Hence, the $(1-\alpha)\%$ confidence intervals (CIs) of the *MTTF* and q -th quantile can be obtained directly.

In practical applications, if we cannot be certain that the degradation model is correctly specified, then computing valid CIs (or variances) of the interesting quantities becomes a more practical issue. Hence, instead of using the FIM (i.e., $n\mathcal{I}(\boldsymbol{\vartheta})$), we adopt the observed information matrix (OIM) proposed by [Boldea and Magnus \(2009\)](#) to construct the valid CIs of the unknown parameters, *MTTF* and q -th quantile whether the degradation model is correctly specified or not. More specifically, we can estimate the FIM by

$$\mathcal{I}_1 = \sum_{i=1}^n \hat{\boldsymbol{\xi}}_i \hat{\boldsymbol{\xi}}_i^\top,$$

where $\hat{\boldsymbol{\xi}}_i = (\partial \mathcal{L}(\boldsymbol{\vartheta} | \mathbf{y}_i) / \partial \boldsymbol{\vartheta})|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}}$ is the score vector corresponding to the single observation $\mathbf{y}_i$ evaluated at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}$, or by the Hessian matrix

$$\mathcal{I}_2 = - \sum_{i=1}^n \frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta} | \mathbf{y}_i)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \Big|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}},$$

based on second-order derivatives, or by the robust matrix

$$\mathcal{I}_3^{-1} = \mathcal{I}_2^{-1} \mathcal{I}_1 \mathcal{I}_2^{-1}.$$

When the underlying model is correctly specified, the inverses $\mathcal{I}_1^{-1}$ and $\mathcal{I}_2^{-1}$ are consistent estimators of the asymptotic variance of $\hat{\boldsymbol{\vartheta}}$. Moreover, the inverse $\mathcal{I}_3^{-1}$ is also a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\vartheta}}$, whether the degradation model is correctly specified or not. See [White \(1982\)](#) and [Boldea and Magnus \(2009\)](#) for more details.

2.4. Overview of iDEMO functionality

The package **iDEMO** is designed not only for advanced users of R, but also for beginners. Analysis can be done using a user-friendly interface and **iDEMO** was designed to meet the following targets:

- (i) Degradation model selection is designed to choose the best one among the degradation models M_0 to M_5 for the user's current data based on AIC, BIC and HQC.
- (ii) For each degradation model M_0 to M_5 , we provide the MLEs of unknown parameters, product's *MTTF*, q -th quantile and the corresponding $(1-\alpha)\%$ CIs based on the FIM and/or OIM.

- (iii) Use the least square estimate (LSE) to obtain the pseudo failure time (PFT) and then diagnose the goodness of fit for the degradation model with PFTs by using the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests.
- (iv) Six plot types are available in **iDEMO**: degradation paths, degradation data summary, the PDF and CDF of the lifetime distribution and the probability-probability (PP) and quantile-quantile (QQ) plots for a specific degradation model with PFTs.

We will address the functions in **iDEMO** sequentially to achieve these goals in the following section.

3. iDEMO interface and functions

3.1. Data input format

In practical applications, the user needs to prepare a data file to run the program **iDEMO**. The data set should be saved as a text file (e.g., .txt, .dat, .prn and so on). The data format has two parts. The first part is the measuring time used in an experiment, which is placed in the first column of the data set. The second part contains the corresponding observations (QC) of each test unit column by column. For example, consider the data file `laser.txt` (Meeker and Escobar 1998) residing in the `etc` subdirectory of **iDEMO**. The few lines of the file are as follows:

```
time    UN1    ... UN15
250    0.47    ... 0.51
500    0.93    ... 0.83
.       .      ... .
.       .      ... .
.       .      ... .
4000  10.94    ... 6.62
```

There are 15 tested laser devices and the QC of a laser device is its operating current. The measured frequency of its operating current is 250 hours and the experiment was terminated at 4000 hours. When the operating current reaches a predefined threshold level, $\omega = 10$, the device is considered to have failed. In the following subsections, we use the laser data to illustrate the stepwise tutorial of **iDEMO**.

3.2. Starting the iDEMO

Once R is running, simply load **iDEMO** by typing the command `library("iDEMO")` into the R console. Typing the command `run.gui()` in the R console, the **iDEMO** window shown in Figure 1 appears. The appearance of the screen images may differ under different operating systems. As shown in Figures 1 and 2, three tabbed notebook widgets (**Basic information**, **Parameter estimation** and **Lifetime information**) are designed to provide required/optional information for the degradation data analysis.

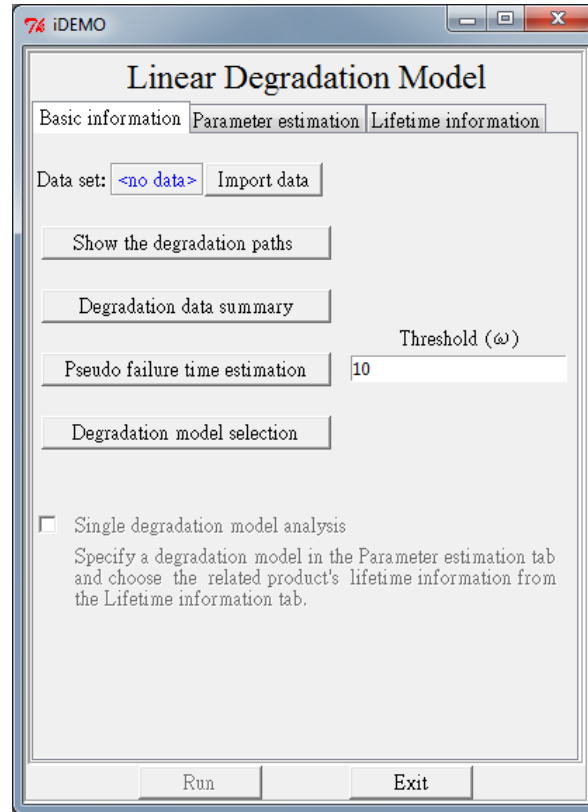


Figure 1: The initial **iDEMO** window (the basic information tab).

3.3. Basic information

Import data

Pressing the **Import data** button results in the following **Open** file window as shown in Figure 3. Select the data file and press the **Open** button. The filename will then appear on the **Data set** column and the data set that the user entered is now the active data set in the R system. To check the correctness of a data set used in GUI, type the filename (e.g., **laser**) of the data set in the R console. The data set will then show in the R console.

Note that clicking the button **Show the degradation paths** will produce the plot of the degradation paths. Figure 4 shows the plot of operating current (QC) over time for the laser data.

Changing the data set

If no data set is imported in GUI, the text **<no data>** will be shown on the **Data set** column. When the user clicks on the **Data set** column, a blank **Name list** window will appear in Figure 5. To change the data set, assume that we have imported two data sets (**data1** and **data2**). By clicking on the **Data set** column (the **Name list** window shown in Figure 5), choosing one (e.g., **data1**) and pressing the **OK** button, the data set (**data1**) can then be used in GUI. Repeating the same procedure, the data set can be changed from **data1** to **data2**.

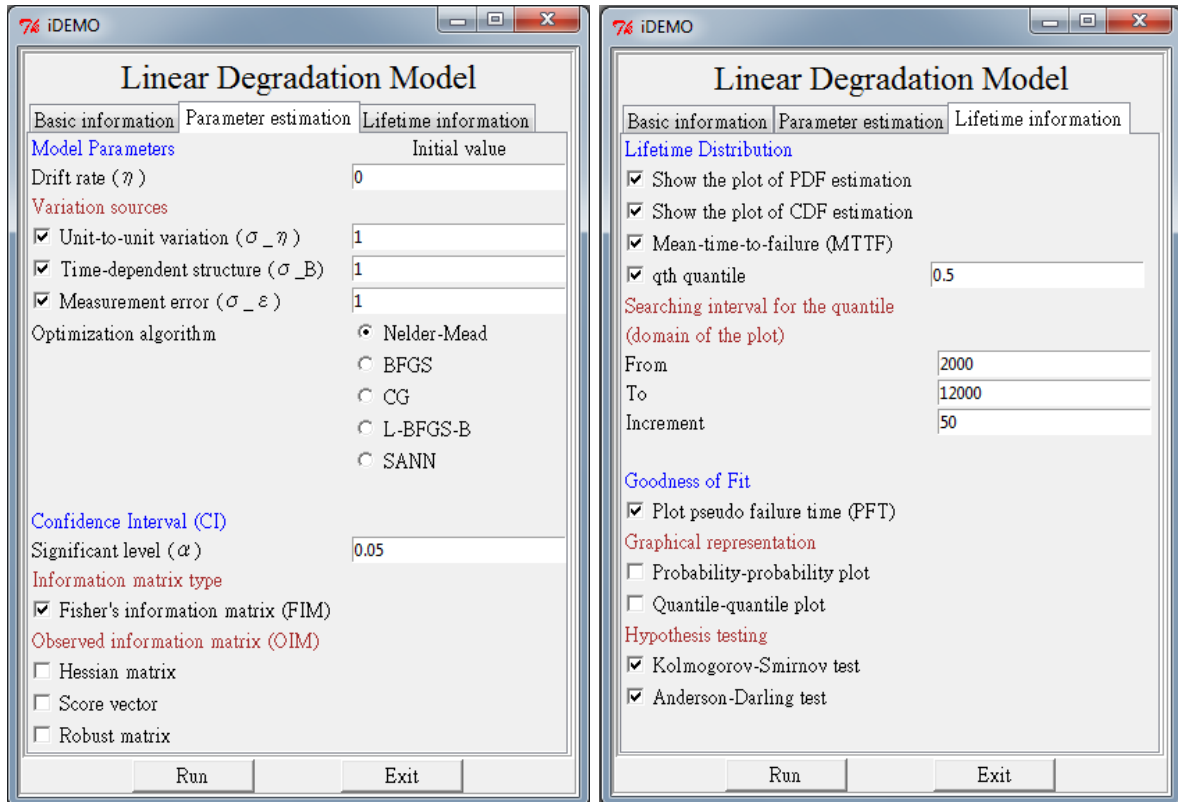


Figure 2: The parameter estimation tab (left) and lifetime information tab (right).

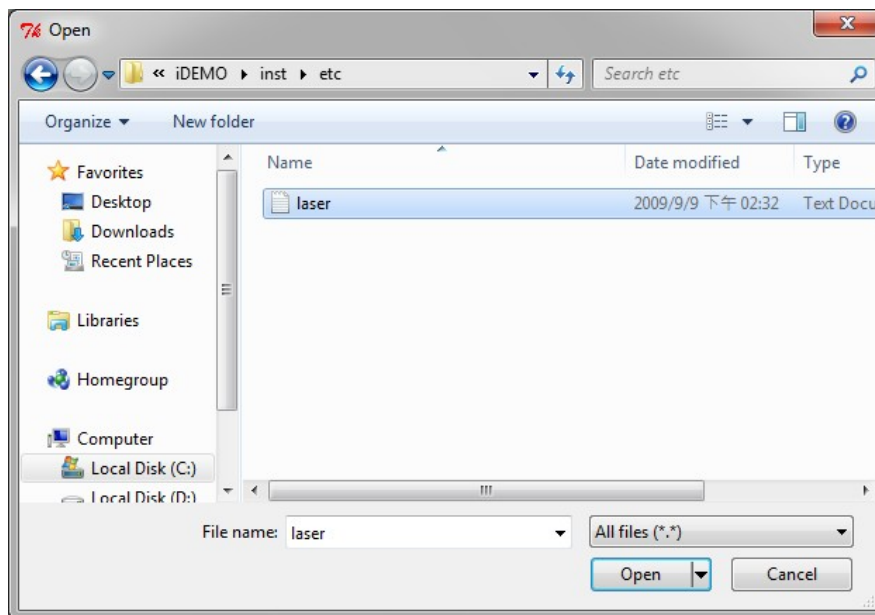


Figure 3: Open file window.

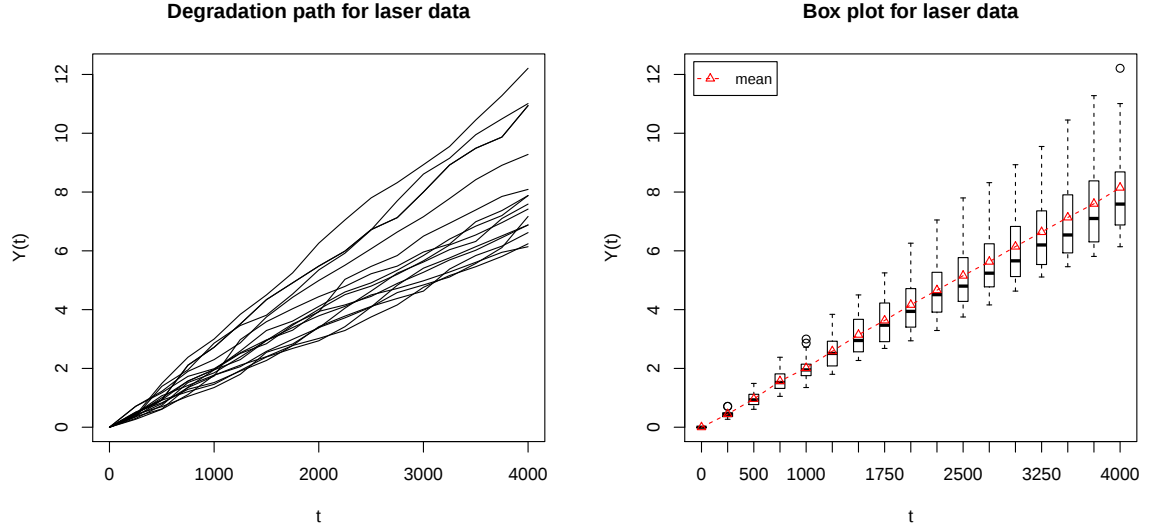


Figure 4: Laser data plot of degradation paths (left) and box plot of the degradation paths for each measurement (right).

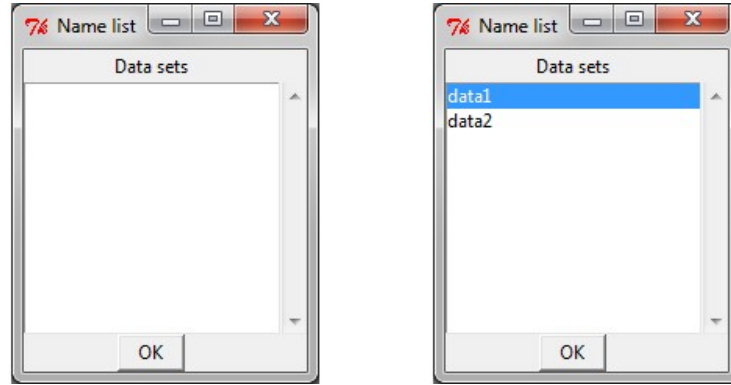


Figure 5: No data set (left) and two data sets (right).

Degradation data summary

Clicking the button **Degradation data summary** will show the box plot of the degradation paths for each measurement. Each box contains the extreme of the lower whisker, the lower hinge, the median, the mean, the upper hinge and the extreme of the upper whisker.

For the laser data, Figure 4 shows the box plot of QC over time.

Pseudo failure time estimation

We first obtain the n slopes ($\hat{\eta}$) after fitting a simple linear regression without the intercept term (i.e., LSE) for each degradation path. The threshold ω should be predefined (the default value is 10 and must be a positive value). The PFT estimation can then be estimated by $\omega/\hat{\eta}$. This means that the eventual crossing times may be beyond the termination time,

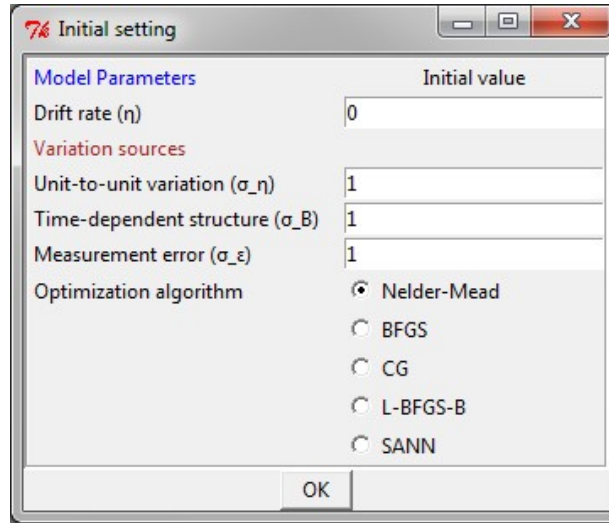


Figure 6: The initial setting window.

where we assume that the degradation test has ended. After clicking the **Pseudo failure time estimation** button, the PFT estimation for the test units will be listed and saved as `PFT.name` (e.g., `PFT.laser`) in R.

For the laser data, click the **Pseudo failure time estimation** button to obtain the PFT as follows.

Pseudo Failure Time Estimation

Threshold = 10

UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
3707.412	4173.362	5621.186	5983.330	5432.730	3612.619	6142.627	6414.984
UN9	UN10	UN11	UN12	UN13	UN14	UN15	
5066.927	3307.686	5268.051	4948.661	4780.164	5819.925	6121.213	

Degradation model selection

If it is not clear which model should be used, **Degradation model selection** can help users to choose an appropriate degradation model. Six combinations of variations in the degradation model can be used in **iDEMO**. After pressing the button **Degradation model selection**, an **Initial setting** window will appear as shown in Figure 6. The default values of η , σ_η , σ_B and σ_ϵ are 0, 1, 1 and 1, respectively. There are five optimization algorithms used in **iDEMO** for searching the MLE: Nelder-Mead (default), BFGS, CG, L-BFGS-B and SANN. The user can input different initial values of the unknown parameters (and/or the optimization algorithm) in the **Initial setting** window to ensure the correctness of the model selection result. Clicking the button **OK** will list the results of the model selection in the R console, including the optimization algorithm, MLEs of the unknown parameters, value of the log-

likelihood function, AIC, BIC and HQC for the models M_0 to M_5 . In addition, ranks are calculated for each criterion.

For the laser data, clicking the button **Degradation model selection** and the button **OK** in the **Initial setting** window will produce the following result of the model selection.

Data: laser

Please wait...

	eta	sigma_eta	sigma_B	sigma_epsilon	log-likelihood
1	0.002036889	0.0004200927	0.01042436	3.089111e-02	69.280101
2	0.002046628	0.0004464373	NA	2.060065e-01	-7.032381
3	0.002037167	NA	0.01265713	NA	45.567703
4	0.002037161	0.0004180486	0.01079381	NA	69.188414
5	0.002046602	NA	NA	1.098698e+00	-363.135454
6	0.002037183	NA	0.01265728	3.495128e-06	45.567703

	AIC(rank)	BIC(rank)	HQC(rank)
1	-130.5602 (2)	-127.728 (2)	-130.5904 (2)
2	20.0648 (5)	22.1889 (5)	20.0421 (5)
3	-87.1354 (3)	-85.7193 (3)	-87.1505 (3)
4	-132.3768 (1)	-130.2527 (1)	-132.3995 (1)
5	730.2709 (6)	731.687 (6)	730.2558 (6)
6	-85.1354 (4)	-83.0113 (4)	-85.158 (4)

Optimization Algorithm: Nelder-Mead

It is clearly seen that the degradation model M_3 is substantially better than the other models in terms of the three described criteria for the model selection.

Single degradation model analysis

Checking the box **Single degradation model analysis**, the **Run** button, the functions in the **Parameter estimation** and **Lifetime information** tabs will be enabled. According to the result of the degradation model selection, the user can specify a degradation model in the **Parameter estimation** tab and choose the related product's lifetime information from the **Lifetime information** tab.

3.4. Parameter estimation

Model parameters

Since the drift rate η is always needed, only select the variation parameters to decide which degradation model fits the data set. Then give each of them an initial value to search for the MLE. The initial values and algorithm are the same settings as described in **Degradation model selection**. i.e., the default value of $(\eta, \sigma_\eta, \sigma_B, \sigma_\epsilon)$ and the default algorithm are (0, 1, 1, 1) and Nelder-Mead, respectively. We recommend using different initial values (and/or the

Function	Purpose
<code>idemo.PDF</code>	PDF of the lifetime distribution.
<code>idemo.CDF</code>	CDF of the lifetime distribution.
<code>idemo.MTTF</code>	mean of the lifetime distribution.
<code>idemo.FIM</code>	FIM of the degradation model.
<code>OIM.hessian</code>	OIM calculated by Hessian matrix.
<code>OIM.score</code>	OIM calculated by score vector.
<code>OIM.robust</code>	OIM calculated by robust matrix.

Table 2: Statistical functions exported by **iDEMO**.

optimization algorithm) to obtain more stable and more accurate estimates of the unknown parameters. In addition, the parameter estimates of outputs will show up in both R console and the **Initial value** edit boxes and replace the initial values automatically. Note that if no variation or only variation σ_η is checked, the functions of CI and the functions in the **Lifetime information** tab will be disabled. This is because it is meaningless to have a model that contains no variation; for the latter, the covariance matrix of the corresponding degradation model is singular. Furthermore, if only variation σ_ϵ is checked (i.e., the traditional regression model M_4), according to the definition of the product's lifetime, it does not make sense to have a product's lifetime without variation. Therefore, the functions in the **Lifetime information** tab will be disabled. Otherwise, six combinations of variations in the degradation model can be arbitrarily chosen.

Confidence interval

The value of the significant level must be between 0 and 1; the default value is 0.05. There are four kinds of information matrices to choose from when constructing the CI. Since the variations σ_η , σ_B and σ_ϵ are positive parameters, to avoid obtaining the negative lower bound, we can use the log-transformation to obtain the corresponding CIs. Note that several functions listed in Table 2 are exported by **iDEMO**. Users can use the `idemo.FIM`, `OIM.hessian`, `OIM.score` and `OIM.robust` functions with specified values for further studies. Furthermore, if no information matrix is selected, the edit box of the significant level of the CI will be disabled. This means that the output will not contain any CI for the parameters, *MTTF* and q -th quantile.

For the laser data, based on the model M_3 , check all the boxes in the **Parameter estimation** tab except for the box **Measurement error** (σ_ϵ). The outputs of analysis can be obtained as follows.

Data: laser

Please wait, analyzing degradation data...

```
#####
      Model: Random effect + Brownian motion
#####
```

MLE and 95% confidence interval of parameters

```
eta = 0.002037161180864866
```

	LCI	UCI	LCI.ln	UCI.ln
FIM	0.001808653	0.002265669	0.001821003	0.002278978
OIM(Hessian matrix)	0.001808653	0.002265669	0.001821003	0.002278978
OIM(score vector)	0.001640347	0.002433975	0.001676603	0.002475259
OIM(robust matrix)	0.001808652	0.002265670	0.001821002	0.002278979

```
sigma_eta = 0.000418048640525997
```

	LCI	UCI	LCI.ln	UCI.ln
FIM	0.0002434043	0.0005926929	0.0002752933	0.0006348308
OIM(Hessian matrix)	0.0002434096	0.0005926877	0.0002752968	0.0006348228
OIM(score vector)	0.0001287850	0.0007073123	0.0002092773	0.0008350867
OIM(robust matrix)	0.0002558451	0.0005802522	0.0002836089	0.0006162171

```
sigma_B = 0.01079380695743448
```

	LCI	UCI	LCI.ln	UCI.ln
FIM	0.009796528	0.01179109	0.009841213	0.01183861
OIM(Hessian matrix)	0.009796565	0.01179105	0.009841246	0.01183857
OIM(score vector)	0.009987568	0.01160005	0.010016943	0.01163092
OIM(robust matrix)	0.009348863	0.01223875	0.009441404	0.01233993

```
log-likelihood = 69.1884136321707
```

```
Optimization Algorithm: Nelder-Mead
```

Note that LCI (UCI) is the lower (upper) bound of the CI of the corresponding parameter. LCI.ln and UCI.ln are the lower and upper bound of the CI with log-transformation, respectively.

3.5. Lifetime information

Lifetime distribution

Checking the boxes, **Show the plot of the PDF estimation**, and, **Show the plot of the CDF estimation**, will show the plots of the PDF and CDF estimation of the lifetime distribution. Users can choose which plots they would like to view by checking boxes from the GUI. Note that because the range of the CDF is between 0 and 1, the pairwise CIs of the CDF estimation are with logit-transformation. In addition, the reliability and failure (or hazard) rate functions can be obtained from the PDF and CDF in (1) and (2). For example, in R, type

```
R> library("iDEMO")
R> Eta <- 0.002
R> sEta <- 0.00042
```

```
R> sB <- 0.0108
R> W <- 10
R> reliability.fun <- 1 - idemo.CDF(t = 4000, Eta, sEta, sB, W)
R> reliability.fun

[1] 0.8649452

R> failure.rate.fun <- idemo.PDF(t = 4000, Eta, sEta, sB, W)/reliability.fun
R> failure.rate.fun
```

```
[1] 0.0003461305
```

For other parameter settings of *MTTF*, in R, type

```
R> idemo.MTTF(Eta = 0.001, sEta = 0.0005, W = 10)
```

```
[1] 12799.76
```

Checking the box, **Mean-time-to-failure** (*MTTF*), will list the MLE of the exact and approximate product's *MTTF* evaluated by (3) and (4). Note that if the estimated drift rate $\hat{\eta}$ is negative, it is meaningless to calculate the product's (exact and approximate) *MTTF*. This is because the slope of the mean degradation path should be positive, otherwise it cannot cross the (prefixed) positive threshold. As for the estimated q -th quantile, suppose that we would like to estimate the median lifetime $t(0.5)$; check the box ***qth quantile*** and type 0.5 (the default value) in the ***qth quantile*** edit box. Furthermore, we need to specify an interval in the **Searching interval for the quantile** edit box to search for the target value (q -th quantile). The default interval is from 2000 to 12000 in increments of 50. If the target quantile is not found in the searching interval after running the program, there are two possible settings that should be noticed: (i) the interval is too narrow to search for the target quantile; on the other hand, the inputted target quantile is too large or too small, and (ii) the value of the threshold may be specified incorrectly. Note that this searching interval is also used as the domain of the plots of the PDF and CDF estimation as well as the QQ plot.

For the laser data, based on the model M_3 and checking all boxes in **Lifetime Distribution** of **Lifetime information** tab, the results of the exact *MTTF*, approximate *MTTF* and median lifetime, along with their corresponding 95% CIs are listed as follows.

MTTF and 95% confidence interval

MTTF(exact)= 5150.077

	LCI	UCI	LCI.ln	UCI.ln
FIM	4463.290	5836.863	4507.114	5884.761
OIM(Hessian matrix)	4463.289	5836.864	4507.113	5884.763
OIM(score vector)	3710.412	6589.741	3894.126	6811.102
OIM(robust matrix)	4650.610	5649.543	4674.065	5674.565

MTTF(approx.)= 4908.792

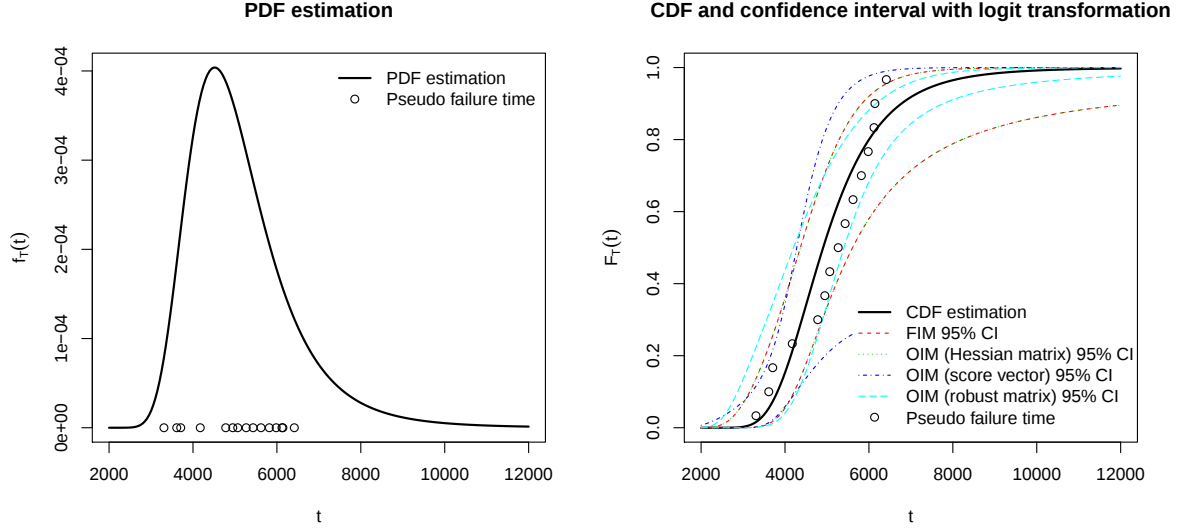


Figure 7: The plot of PDF estimation (left) and plot of CDF estimation (right).

	LCI	UCI	LCI.ln	UCI.ln
FIM	4358.173	5459.410	4387.931	5491.480
OIM(Hessian matrix)	4358.173	5459.410	4387.931	5491.480
OIM(score vector)	3952.620	5864.964	4039.982	5964.442
OIM(robust matrix)	4358.172	5459.412	4387.930	5491.481

qth-quantile and 95% confidence interval

$t(0.5) = 4908.792$

	LCI	UCI	LCI.ln	UCI.ln
FIM	4358.173	5459.410	4387.931	5491.480
OIM(Hessian matrix)	4358.173	5459.410	4387.931	5491.480
OIM(score vector)	3952.620	5864.964	4039.982	5964.442
OIM(robust matrix)	4358.172	5459.412	4387.930	5491.481

$\lim_{t \rightarrow \infty} F_T(t) = 0.9999995$

It is easily seen that the CIs evaluated by FIM and OIM calculated by the Hessian matrix are almost the same for the MLEs of the parameters, $MTTF$ and median lifetime. This means that the sample size of the laser data is large in comparison to the number of parameters. Furthermore, these results show that even if the degradation model is mis-specified, the CIs computed from the OIM calculated by the robust matrix are narrower than those from the FIM. Note that for the degradation model M_3 , $P\{\Theta \leq 0\} = 5.495135 \times 10^{-7}$ is negligible, although the density function of the lifetime distribution is improper (i.e., $\lim_{t \rightarrow \infty} F_T(t) = 0.9999995$). Figure 7 shows the plots of the PDF and CDF estimation of the lifetime distribution, respectively.

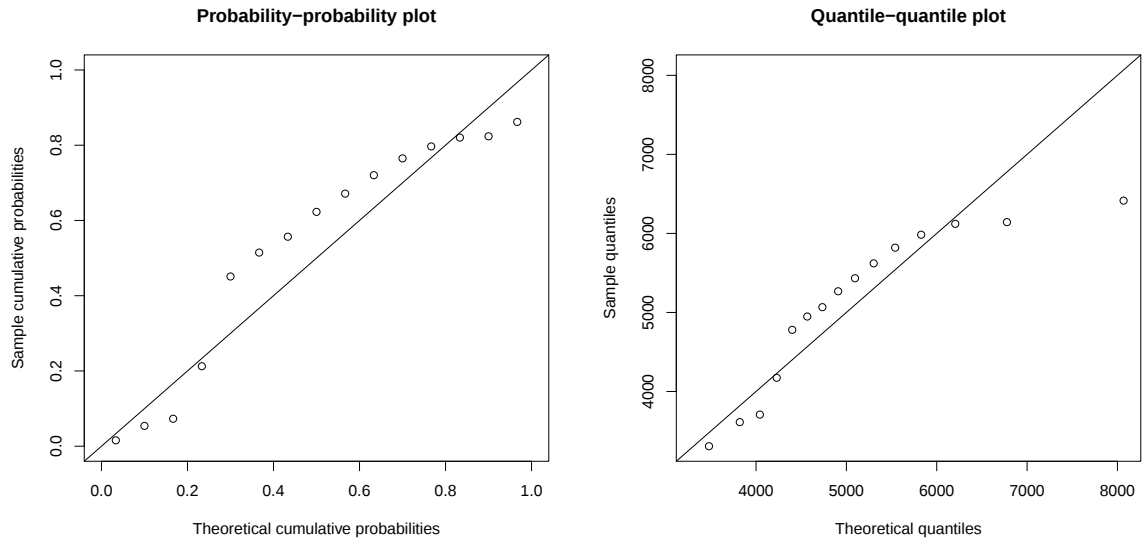


Figure 8: PP plot (left) and QQ plot (right).

Goodness of fit

To diagnose the goodness of fit for the specified degradation model with PFTs, we provide two graphical representations (i.e., PP and QQ plots) and two well-known hypothesis testings (i.e., Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests). If the box `Plot the pseudo failure time (PFT)` is not checked, the functions of the PP and QQ plots as well as the KS and AD tests will be disabled. This means that no goodness-of-fit outputs will be listed in the R console. Note that if the estimated drift rate $\hat{\eta}$ is negative or the PFT estimation contains the negative values (i.e., $\hat{\eta} < 0$), the functions in the `Goodness of Fit` will not work because it is meaningless to do so.

For the laser data, based on the model M_3 and checking all the boxes in the `Goodness of Fit of Lifetime information` tab, the results of the goodness of fit are obtained as follows.

Goodness-of-fit tests

Kolmogorov-Smirnov test
 statistic = 0.1844500
 p-value = 0.6224386

Anderson-Darling test
 statistic = 0.7970605
 p-value = 0.4808687

Figure 8 shows the PP and QQ plots, respectively. The p values of the KS and AD tests and the graphical representations demonstrate that the degradation model M_3 may be an appropriate model for fitting the PFTs.

3.6. Odds and ends

After finishing the above settings, press the **Run** button to submit the computational job. If the inputted information is invalid or the data files are incorrectly formatted, there is a warning or an error message shown in the R console, which informs the user to correct the format. If the inputted data passes the examination, the program **iDEMO** starts to perform the analysis and the message

Please wait, analyzing degradation data...

will show up in the command line. A prompt sign will appear immediately, but the computation is still proceeding. Please wait until the results appear. Click the **Exit** button to quit the GUI.

4. Conclusion and future development

In this paper we have presented the R package **iDEMO** for analysis of degradation tests. All the features of the GUI in **iDEMO** have been explained and illustrated through the laser data set. The GUI is designed not only for non-R users but also for advanced users. For the latter, some useful functions built into the package are provided to do further studies. In addition, the disabled property used in the GUI can effectively avoid input errors. The latest version of **iDEMO** can be downloaded at <http://www.stat.sinica.edu.tw/chienyu/> or <http://www.idemo.tw/>.

Further development in **iDEMO** for improving the flexibility and applications will focus on two directions.

- For practical applications, if the degradation path of a product degrades very slowly, engineers usually use higher-level stress variables (such as temperature, voltage, electric current etc.) to accelerate the degradation path. Typical experiments are accelerated degradation tests (ADTs), step-stress ADT, progressive-stress ADT and so on. See [Liao and Tseng \(2006\)](#), [Tseng and Peng \(2007\)](#), and [Peng and Tseng \(2010\)](#) for details.
- For the monotonic degradation path, instead of the Gaussian process fitted in the present paper, we may use other types of processes to fit data like the Hougaard process ([Hougaard 1986](#)), which includes stable, gamma and inverse Gaussian processes as special cases. An intuitive advantage of the Hougaard process is that they are strictly increasing, which seems more reasonable for the laser data.

Computational details

All computations and graphics in this paper have been obtained using the R version 2.15.0. Several utility packages have been created to help in the analyzing process. For instance, package **gsl** ([Hankin 2006](#)) provides the Dawson's integral for evaluating the *MTTF*, package **tktk2** ([Grosjean 2012](#)) creates the tabbed notebook widget and package **ADGofTest** ([Bellosta 2011](#)) calculates the *p* value of the AD statistic based on the algorithm developed by [Marsaglia and Marsaglia \(2004\)](#).

Acknowledgments

This work was partially supported by the National Science Council (Grant No: NSC-99-2118-M-001-002 and 100-2118-M-001-004) and National Center for Theoretical Sciences (NCTS) of Taiwan, Republic of China. We are grateful to Li-Yao Jill Wang and Tsuey-Hwa Hu for their technical support and many of the users of the package for very helpful suggestions. We would also like to thank the two anonymous referees and the special volume Editors for their observant and valuable comments which made this paper more readable and had improved the package.

References

- Bae SJ, Kvam PH (2004). “A Nonlinear Random Coefficients Model for Degradation Testing.” *Technometrics*, **46**(4), 460–469.
- Bellosta CJG (2011). *ADGofTest: Anderson-Darling GoF Test*. R package version 0.3, URL <http://CRAN.R-project.org/package=ADGofTest>.
- Boldea O, Magnus JR (2009). “Maximum Likelihood Estimation of the Multivariate Normal Mixture Model.” *Journal of the American Statistical Association*, **104**(488), 1539–1549.
- Casella G, Berger RL (2001). *Statistical Inference*. 2nd edition. Duxbury Press, Belmont.
- Chao MT (1999). “Degradation Analysis and Related Topics: Some Thoughts and a Review.” *The Proceedings of the National Science Council A*, **23**(5), 555–566.
- Chhikara RS, Folks L (1989). *The Inverse Gaussian Distribution. Theory, Methodology, and Applications*. Marcel Dekker, New York.
- Demidenko E (2004). *Mixed Models: Theory and Applications*. John Wiley & Sons, New York.
- Doksum KA, Høyland A (1992). “Model for Variable-Stress Accelerated Life Testing Experiments Based on Wiener Processes and the Inverse Gaussian Distribution.” *Technometrics*, **34**(1), 74–82.
- Fellows I (2012). “**Deducer**: A Data Analysis GUI for R.” *Journal of Statistical Software*, **49**(8), 1–15. URL <http://www.jstatsoft.org/v49/i08/>.
- Fox J, Carvalho MS (2012). “The **RcmdrPlugin.survival** Package: Extending the R Commander Interface to Survival Analysis.” *Journal of Statistical Software*, **49**(7), 1–32. URL <http://www.jstatsoft.org/v49/i07/>.
- Grosjean P (2012). *tcltk2: Tcl/Tk Additions*. R package version 1.2-1, URL <http://CRAN.R-project.org/package=tcltk2>.
- Hankin RKS (2006). “Special Functions in R: Introducing the **gsl** Package.” *R News*, **6**, 24–26. URL <http://CRAN.R-project.org/doc/Rnews/>.

- Hougaard P (1986). “Survival Models for Heterogeneous Populations Derived from Stable Distributions.” *Biometrika*, **73**(2), 387–396.
- Lawless J, Crowder M (2004). “Covariates and Random Effects in a Gamma Process Model with Application to Degradation and Failure.” *Lifetime Data Analysis*, **10**(3), 213–227.
- Lehmann EL, Casella G (1998). *Theory of Point Estimation*. 2nd edition. Springer-Verlag, New York.
- Liao CM, Tseng ST (2006). “Optimal Design for Step-Stress Accelerated Degradation Tests.” *IEEE Transactions on Reliability*, **55**(1), 59–66.
- Lu CJ, Meeker WQ (1993). “Using Degradation Measures to Estimate a Time-to-Failure Distribution.” *Technometrics*, **35**(2), 161–174.
- Lu JC, Park J, Yang Q (1997). “Statistical Inference of a Time-to-Failure Distribution Derived from Linear Degradation Data.” *Technometrics*, **39**(4), 391–400.
- Marsaglia G, Marsaglia JCW (2004). “Evaluating the Anderson-Darling Distribution.” *Journal Statistical Software*, **9**(2), 1–5. URL <http://http://www.jstatsoft.org/v09/i02/>.
- Meeker WQ, Escobar LA (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons, New York.
- Nelson W (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*. John Wiley & Sons, New York.
- Park JI, Bae SJ (2010). “Direct Prediction Methods on Lifetime Distribution of Organic Light-Emitting Diodes from Accelerated Degradation Tests.” *IEEE Transactions on Reliability*, **59**(1), 74–90.
- Peng CY (2008). “The First Negative Moment in the Sense of the Cauchy Principal Value.” *Statistics & Probability Letters*, **78**(13), 1765–1774.
- Peng CY (2012). “A Note on Optimal Allocations for the Second Elementary Symmetric Function with Applications for Optimal Reliability Design.” *Naval Research Logistics*, **59**(3-4), 278–284.
- Peng CY, Hsu SC (2012). “A Note on a Wiener Process with Measurement Error.” *Applied Mathematics Letters*, **25**(4), 729–732.
- Peng CY, Tseng ST (2009). “Mis-Specification Analysis of Linear Degradation Models.” *IEEE Transactions on Reliability*, **58**(3), 444–455.
- Peng CY, Tseng ST (2010). “Progressive-Stress Accelerated Degradation Test for Highly-Reliable Products.” *IEEE Transactions on Reliability*, **59**(1), 30–37.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

- Rödiger S, Friedrichsmeier T, Kapat P, Michalke M (2012). “**RKward**: A Comprehensive Graphical User Interface and Integrated Development Environment for Statistical Analysis with R.” *Journal of Statistical Software*, **49**(9), 1–34. URL <http://www.jstatsoft.org/v49/i09/>.
- Tseng ST, Peng CY (2004). “Optimal Burn-in Policy by Using an Integrated Wiener Process.” *IIE Transactions*, **36**(12), 1161–1170.
- Tseng ST, Peng CY (2007). “Stochastic Diffusion Modelling of Degradation Data.” *Journal of Data Science*, **5**(3), 315–333.
- Valero-Mora PM, Ledesma R (2012). “Graphical User Interfaces for R.” *Journal of Statistical Software*, **49**(1), 1–8. URL <http://www.jstatsoft.org/v49/i01/>.
- White H (1982). “Maximum Likelihood Estimation of Misspecified Models.” *Econometrica*, **50**(1), 1–25.
- Whitmore GA (1986). “Normal-Gamma Mixtures of Inverse Gaussian Distributions.” *Scandinavian Journal of Statistics*, **13**(3), 211–220.

A. Proof of Proposition 1

Proof. Without loss of generality, we only consider one degradation path (i.e., $n = 1$). Then it is easily seen that $\mathbf{Y} \sim \mathcal{N}(\eta \mathbf{t}, \boldsymbol{\Sigma})$. Now, if $(\eta \mathbf{t}, \boldsymbol{\Sigma}) = (\eta^* \mathbf{t}, \boldsymbol{\Sigma}^*)$, we have the equal mean function $\eta \mathbf{t} = \eta^* \mathbf{t}$, implying $\eta = \eta^*$. In addition, for the equal variance function $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^*$ from (5), we have

$$(\sigma_\eta^2 - \sigma_\eta^{*2}) \mathbf{t} \mathbf{t}^\top + (\sigma_B^2 - \sigma_B^{*2}) \mathbf{Q} + (\sigma_\epsilon^2 - \sigma_\epsilon^{*2}) \mathbf{I}_m = \mathbf{0}, \quad (7)$$

where the matrices $\mathbf{Q}$ and $\mathbf{I}_m$ are given in (6). Note that the off-diagonal elements of $\mathbf{Q}$ and $\mathbf{t} \mathbf{t}^\top$ are the functions of $\{t_i\}_{i=1}^m$. Thus for the off-diagonal entry (i, j) , we have the following result:

$$(\sigma_\eta^2 - \sigma_\eta^{*2}) t_j + (\sigma_B^2 - \sigma_B^{*2}) = 0, \quad 1 \leq i \neq j \leq m.$$

If $m \geq 4$ the above simultaneous equations will have the unique solution $\sigma_\eta^2 - \sigma_\eta^{*2} = 0$ and $\sigma_B^2 - \sigma_B^{*2} = 0$. Now, substituting $\sigma_\eta^2 = \sigma_\eta^{*2}$ and $\sigma_B^2 = \sigma_B^{*2}$ into (7), we have $\sigma_\epsilon^2 = \sigma_\epsilon^{*2}$. From the Proposition 10 of Demidenko (2004), this completes the proof. $\square$

Affiliation:

Chien-Yu Peng
 Institute of Statistical Science
 Academia Sinica
 No. 128, Section 2, Academia Road
 Taipei, 11529, Taiwan R.O.C.
 Telephone: +886/2/2783-5611 ext. 303
 Fax: +886/2/2783-1523
 E-mail: chienyu@stat.sinica.edu.tw
 URL: <http://www.stat.sinica.edu.tw/chienyu/>