

Modelos y análisis para datos de degradación

SERGIO YÁNEZ C.*
RONALD ANDRÉS GRANADA G.**
MARIO CESAR JARAMILLO E.***

Resumen

La degradación es una debilidad que eventualmente puede causar la falla (e. g., el desgaste que sufren los neumáticos de un automóvil). Cuando es posible medirla, esta puede proporcionar mayor información que los datos de tiempo de falla, para propósitos de determinación y mejoramiento de la confiabilidad de un producto. Este artículo es de carácter divulgativo y desarrolla técnicas que son propuestas por Meeker y Escobar (1998). Se cree importante hacer conocer este tópico, hoy en la frontera de la Teoría de Confiabilidad (Lawless 2000)). En este trabajo se compara el análisis clásico aproximado de degradación con el modelo de degradación explícito. Estos últimos modelos implican la utilización de modelos físicos de degradación a los cuales se les introduce efectos aleatorios. Se implementan las técnicas para la estimación de modelos mixtos en S-PLUS siguiendo a Pinheiro y Bates (2000) y se utiliza el bootstrap para intervalos de confianza.

Palabras Clave: Teoría de confiabilidad, datos de degradación, bootstrap, modelos de efectos mixtos, efectos aleatorios.

*Profesor asociado escuela de Estadística, Universidad Nacional de Colombia sede Medellín; e-mail: syc@geo.net.co

**Estudiante de la maestría en Estadística, Universidad Nacional de Colombia sede Medellín; e-mail: ronaldgra@yahoo.com

***Profesor asistente Escuela de Estadística, Universidad Nacional de Colombia sede Medellín; e-mail: mcjarami@perseus.unalmed.edu.co

Abstract

Degradation is a weakness that eventually can cause failure (e.g. car tire wear). When it is possible to measure degradation, such measures often provide more information than failure-time data for purposes of assessing and improving product reliability. This is a paper which mainly pretends to divulge techniques that had been developed by Meeker & Escobar (1998). We think it is worth to make this topics known, because they are in the research frontier of the Reliability Theory (Lawless 2000). We compare in this work the explicit degradation models with the approximate degradation analysis. The explicit degradation model requires specific models developed by engineers and physical scientists, which are treated as mixed models with random effects. To obtain ML estimates we use S-PLUS following Pinheiro & Bates (2000), and also use bootstrap confidence intervals.

Keywords: *Reliability theory, degradation data, bootstrap, mixed effects models, random effects.*

1. Modelos para datos de degradación

Cuando se quiere estudiar la confiabilidad de un producto en particular, un tipo de información que se puede recolectar, son los datos de la degradación física del producto como una función del tiempo (e.g., El desgaste de una llanta para automóvil, el crecimiento de una grieta en un disco de una turbina). Sin embargo en muchos casos no se puede medir físicamente la degradación, pero si se pueden tener medidas de la evolución del desempeño del producto y por ende de su “degradación” en el tiempo en términos de desempeño, por ejemplo el porcentaje de incremento en la transconductancia de un semiconductor (Lu, Park & Yang 1997). Los datos mencionados son catalogados como “datos de degradación” (Ver Figura 1). Si los datos de degradación están disponibles, su uso puede tener importantes ventajas prácticas. En particular:

- Los datos de degradación, especialmente en aplicaciones con pocas o ninguna falla, proporcionan mucha más confiabilidad que con el método tradicional de censura de los datos de tiempo de falla.
- Las pruebas aceleradas, son usadas en forma común para obtener rápidamente información de las pruebas de confiabilidad. La observación directa de los procesos físicos de degradación puede permitir un modelamiento directo de los mecanismos causantes de las fallas, proporcionando más

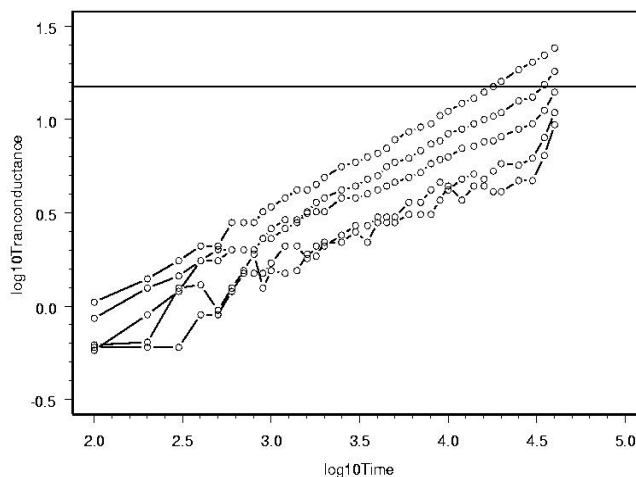


Figura 1: Porcentaje de incremento en la Transconductancia vs tiempo en segundos para cinco semiconductores

creíbles y precisas estimaciones de confiabilidad, y una base firme para cuando se necesite extrapolación.

1.1. Trayectorias de la degradación a la falla

Un gran número de fallas pueden ser descritas principalmente por procesos de degradación. La figura 2 muestra tres tipos generales de curvas de degradación en unidades arbitrarias de degradación y tiempo: lineal, convexa, cóncava. La línea horizontal es el nivel de degradación (o nivel aproximado), al cual puede considerarse que una componente falla.

La literatura especializada dispone de un gran número de modelos que dan cuenta de curvas de degradación. Dichos modelos han sido desarrollados por ingenieros y físicos basados en los principios básicos que describen los procesos de degradación. Usualmente dichos modelos, empiezan con una descripción determinista de los procesos de degradación, en la forma de ecuaciones diferenciales (e.g., Modelo de Paris para el crecimiento de las grietas por fatiga del material ver Figura 3) o sistemas de ecuaciones diferenciales (ver Meeker & Escobar 1998). La aleatoriedad puede ser introducida apropiadamente, usando distribuciones de probabilidad para describir la variabilidad de los parámetros

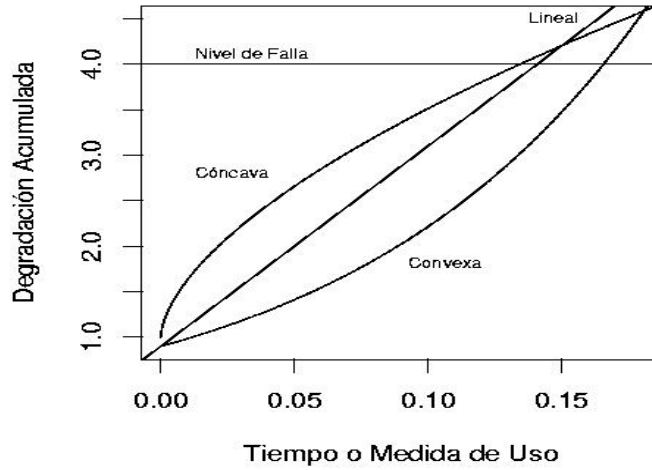


Figura 2: Posibles formas de curvas de degradación univariadas.

del modelo, como tasas, “constantes” o propiedades del material.

1.2. Trayectoria general del modelo de degradación

La trayectoria real de degradación de una unidad particular sobre el tiempo se denota por $D(t)$, $t > 0$. en algunas aplicaciones, valores de $D(t)$ son tomados como puntos discretos en el tiempo t_1, t_2, \dots, t_m y la degradación muestral observada y_{ij} de la unidad i al tiempo t_j es:

$$y_{ij} = D_{ij} + \varepsilon_{ij}, i = 1, \dots, n, j = 1, \dots, m_i, \quad (1)$$

donde $D_{ij} = D(t_{ij}, \beta_{1i}, \dots, \beta_{ki})$ es la trayectoria real para la i -ésima unidad al tiempo t_{ij} y $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ es la desviación residual de la unidad i en el tiempo t_j . El número total de inspecciones en la unidad i es denotado m_i . El tiempo t_j puede ser tiempo real o tiempo operacional o alguna otra medida apropiada de uso (e.g., millones de ciclos, kilómetros, segundos, horas etc.). Para la i -ésima unidad $(\beta_{1i}, \dots, \beta_{ki})$ es el vector de k parámetros desconocidos, algunos de los cuales pueden ser considerados aleatorios y otros fijos. Generalmente las trayectorias muestrales tienen $k = 1, 2, 3$ ó 4 (ver Meeker & Escobar 1998).

Las escalas de \mathbf{y} y \mathbf{t} pueden ser escogidas como sugieran la teoría física y los datos, para simplificar la forma de $D(t, \beta_1, \dots, \beta_k)$. Los modelos de degradación,

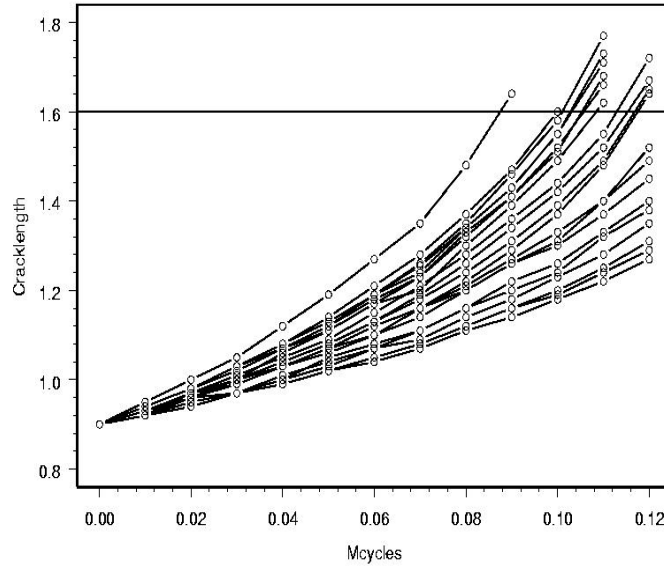


Figura 3: Trayectorias de degradación para los datos de grietas por fatiga en la Aleación A

requieren no sólo la especificación de la forma de la función $D(t, \beta_1, \dots, \beta_k)$, sino que requieren también la especificación de cuales de los β_1, \dots, β_k son aleatorios (diferentes de unidad a unidad) y cuales son fijos (comunes para todas las unidades). Por conveniencia, se modelará la variabilidad unidad a unidad en β_1, \dots, β_k con una distribución normal multivariada con vector de medias μ_β y matriz de covarianzas Σ_β .

Generalmente se asume que los parámetros aleatorios β_1, \dots, β_k son independientes de las desviaciones ε_{ij} . Otro supuesto muy común es que σ_ε es constante; sin embargo en algunas ocasiones es necesario modelarla para cada nivel de la variable de uso, con el fin de corregir una posible heterocedasticidad.

1.3. Parámetros del modelo de degradación

Si bien los valores de β_1, \dots, β_k para las unidades individuales, pueden ser de interés en algunas aplicaciones (para predecir la degradación futura de una unidad particular, basada en unas pocas observaciones iniciales), nos concentraremos en el uso de los datos de degradación, para hacer inferencias acerca de

la población, proceso o predicciones acerca de la degradación de futuras unidades. En este caso los parámetros principales del modelo son μ_β , Σ_β y también la desviación estándar del error σ_ε . Por facilidad de notación podemos escribir $\theta_\beta = (\mu_\beta, \Sigma_\beta)$ para denotar los parámetros sobre toda la población o proceso.

2. Estimación de los parámetros del modelo de degradación

La verosimilitud para los parámetros aleatorios del modelo de degradación puede ser expresada como:

$$L(\theta_\beta, \sigma_\varepsilon | DATOS) = \prod_{i=1}^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{j=1}^{m_i} \frac{1}{\sigma_\varepsilon} \phi_{nor}(\zeta_{ij}) \right] \times f_\beta(\beta_{1i}, \dots, \beta_{ki}; \theta_\beta) d\beta_{1i}, \dots, d\beta_{ki} \quad (2)$$

donde $\zeta_{ij} = [y_{ij} - D(t_{ij}, \beta_{1i}, \dots, \beta_{ki})]/\sigma_\varepsilon$ y $f_\beta(\beta_{1i}, \dots, \beta_{ki}; \theta_\beta)$ es la densidad Normal Multivariada. Cada evaluación de la verosimilitud puede, en general requerir aproximación numérica de n integrales de dimensión k (donde n es el número de trayectorias muestrales y k el número de parámetros aleatorios en cada trayectoria). Maximizar (2) con respecto a $(\mu_\beta, \Sigma_\beta, \sigma_\varepsilon)$, directamente, aún con las capacidades computacionales de hoy, es muy difícil a menos que $D(t)$ sea una función lineal. Pinheiro & Bates (2000) describen y comparan esquemas de estimación que proporcionan estimaciones aproximadas de *ML* para $\theta_\beta = (\mu_\beta, \Sigma_\beta)$ también como para las unidades específicas $\beta_{1i}, \dots, \beta_{ki}$, $i = 1, \dots, n$. Ellos implementan los programas para estimar los parámetros en *S-PLUS* y en *R* como las funciones `lme` y `nlme` (linear and non linear mixed effects); las cuales son utilizadas en este trabajo.

3. Modelos que describen la degradación y la falla

Algunos productos tienen una pérdida gradual del desempeño (e.g., disminución de la luz emanada de una lámpara fluorescente). Entonces la falla puede ser definida como un nivel específico de degradación (Digamos 60% de la salida inicial). Esto es lo que nosotros definimos como “falla suave”.

Un valor fijo de D_f puede ser usado para denotar el nivel crítico de la trayectoria de degradación donde se asume que ocurre la falla. El tiempo de

falla T es definido como el tiempo en el cual la trayectoria real de degradación $D(t)$ cruza el nivel crítico de degradación D_f .

Un modelo específico para $D(t)$ y D_f define una distribución del tiempo de falla. En general, esta distribución puede ser escrita como una función de los parámetros del modelo de degradación. Supongamos que la unidad falla al tiempo t , si la degradación acumulada alcanza el nivel de fallo D_f a un tiempo t :

$$P(T < t) = F(t) = F(t, \theta_\beta) = P[D(t, \beta_1, \dots, \beta_k) \geq D_f] \quad (3)$$

Para un D_f fijo, la distribución de T depende de la distribución de los $\beta_{1i}, \dots, \beta_{ki}$, la cual depende de la forma básica de los parámetros en θ_β . En algunos casos simples es posible escribir la expresión para $F(t)$ en una forma cerrada. En general, sin embargo la expresión en una forma cerrada puede no existir. Para la mayoría de los modelos prácticos, especialmente cuando $D(t)$ es no lineal y más de uno de los parámetros $\beta_{1i}, \dots, \beta_{ki}$, sean aleatorios, es necesario evaluar $F(t)$ numéricamente o vía Monte Carlo utilizando el siguiente algoritmo (ver Meeker & Escobar 1998):

1. Generar N realizaciones simultáneas $\check{\beta}_1, \dots, \check{\beta}_k$ de β_1, \dots, β_k de una distribución normal multivariada con media $\hat{\mu}_\beta$ y matriz de varianza-covarianza $\hat{\Sigma}_\beta$, donde N es un número grande (e.g., $N = 100000$).
2. Calcular los N tiempos de falla simulados correspondientes a las N realizaciones de $\check{\beta}_1, \dots, \check{\beta}_k$ en $D(t, \beta_1, \dots, \beta_k)$, encontrando los tiempos de cruce para cada una (puede ser necesario utilizar un método de búsqueda de raíces numéricas dependiendo de la forma de $D(t)$).
3. Para cualquier valor deseado de t , usamos

$$F(t) \approx \frac{\text{Número de Tiempos de Cruce Simulados} \leq t}{N}$$

como una evaluación de $F(t)$.

El error potencial en esta aproximación Monte Carlo se puede fácilmente evaluar por el uso de la distribución binomial como $\sqrt{F(t)(1-F(t))/N}$; que es un número bastante pequeño para la escala de los N que utilizamos.

3.1. Intervalos de confianza bootstrap

Puesto que no existe un método simple para calcular los errores estándar de $F(t)$, se puede usar el método bootstrap percentil con corrección de sesgo, des-

crito en Meeker & Escobar (1998) o más ampliamente por Efron & Tibshirani (1993), implementado con el siguiente algoritmo:

1. Usar los datos observados de las n trayectorias muestrales para calcular las estimaciones de $\hat{\theta}_\beta$ y $\hat{\sigma}_\epsilon^2$.
2. Use el algoritmo de Monte Carlo con $\hat{\theta}_\beta$ como entrada para calcular $\hat{F}(t)$ en los valores deseados de t .
3. Generar un gran número B (e.g., $B = 4000$) de muestras bootstrap que emulen la muestra original y calcule las correspondientes estimaciones bootstrap de $\hat{F}^*(t)$ acorde con los siguientes pasos.
 - a) Generar, de $\hat{\theta}_\beta$, n realizaciones simuladas de los parámetros aleatorios $\beta_1^*, \dots, \beta_{ki}^*$, $i = 1, \dots, n$.
 - b) Usando el mismo esquema que en experimento original, calcular las n trayectorias simuladas como

$$y_{ij}^* = D(t_{ij}, \beta_{1i}^*, \dots, \beta_{ki}^*) + \epsilon_{ij}^*$$

hasta el tiempo planeado de detención t_c , donde los residuales son simulados independientemente de una distribución $N(0, \hat{\sigma}_\epsilon^2)$.

- c) Use las n trayectorias simuladas para estimar los parámetros del modelo, dando las estimaciones bootstrap de $\hat{\theta}_\beta^*$.
 - d) Use el algoritmo Monte Carlo con $\hat{\theta}_\beta^*$ como entrada para calcular las estimaciones bootstrap de $\hat{F}^*(t)$ en los valores deseados de t .
4. Para cada valor deseado de t , los intervalos de confianza bootstrap para $F(t)$ serán calculados usando los siguientes pasos.
 - a) Organice las B estimaciones bootstrap $\hat{F}^*(t)_1, \dots, \hat{F}^*(t)_B$ en orden ascendente dando $\hat{F}^*(t)_{(b)}$, $b = 1, \dots, B$.
 - b) Siguiendo a Efron & Tibshirani (1993), los límites inferior y superior de los intervalos de confianza puntuales aproximados del $100(1-\alpha)\%$ para la distribución $F(t)$ son

$$\left[\underset{\sim}{F}(t), \tilde{F}(t) \right] = \left[\hat{F}^*(t)_{(l)}, \hat{F}^*(t)_{(u)} \right]$$

donde

$$\begin{aligned} l &= B \times \Phi_{nor} \left[2\Phi_{nor}^{-1}(q) + \Phi_{nor}^{-1}(\alpha/2) \right] \\ u &= B \times \Phi_{nor} \left[2\Phi_{nor}^{-1}(q) + \Phi_{nor}^{-1}(1 - \alpha/2) \right] \end{aligned}$$

y q es la proporción de los B valores de $\hat{F}^*(t)$ que son menores que $F(t)$ (si se usa $q = 0.5$ es equivalente al método bootstrap percentil).

4. Aplicación de la metodología para trayectorias de degradación lineal

Los datos que se tratarán para ilustrar esta metodología fueron inicialmente trabajados por Takeda & Susuki (1983) y posteriormente por Lu, Park & Yang (1996). Dichos datos hacen referencia a un tipo especial de degradación, la hot-carrier-induced (HCI). Este tipo de degradación (ver Figura 1) se desarrolla gradualmente y cambia el desempeño de los semiconductores metalizados y oxidados (MOS). El desempeño de los semiconductores puede ser medido en términos simples por el comportamiento de una variable llamada Transconductancia que tiene la siguiente propiedad: a medida que aumenta, el semiconductor pierde sus cualidades físicas es decir va perdiendo su resistencia natural a que fluya la corriente eléctrica y por lo tanto puede generar errores en las aplicaciones donde se utilice. En el experimento planteado por Takeda & Susuki (1983) se midió la degradación como el aumento en porcentaje de la transconductancia para cada uno de cinco dispositivos para 35 tiempos diferentes medidos en segundos de exposición a una corriente bajo condiciones constantes.

Las variables monitoreadas en los tiempos sucesivos fueron la degradación de la transconductancia (porcentaje de $\Delta G_m / \Delta G_{m0}$), la variación en el cambio del voltaje (ΔV_T), o la corriente lineal (porcentaje de $\Delta I_D / \Delta I_{D0}$). Donde G_m es la transconductancia, G_{m0} es la máxima transconductancia del dispositivo probado, $\Delta G_m = G_{m0} - G_m$. I_{D0} es la “drain current” original, y ΔI_D es la diferencia entre las corrientes “unstressed and stressed” (para más detalles ver Lu, Park & Yang 1996).

De la teoría física podemos llegar a un modelo que relaciona el porcentaje de incremento en la transconductancia como una función del tiempo t^* de la siguiente manera. $\log_{10}(\Delta I_D / I_D) \approx [\log_{10}(K_1) + n \log_{10}(K_2)] + \eta \log_{10}(t^*)$ donde η , K_1 y K_2 son constantes que dependen de las condiciones del experimento y que no dependen de t^* . Si definimos $y_{ij} = \log_{10}(\Delta I_D / I_D)$ es decir el logaritmo en base diez del porcentaje de incremento en la transconductancia para la unidad i al tiempo t_{ij}^* , además $\beta_0 = \log_{10}(K_1) + \eta \log_{10}(K_2)$, $\beta_1 = \eta$, y $t_{ij} = \log_{10}(t_{ij}^*)$; utilizando el modelo general para las trayectorias de degradación propuesto en (1) tenemos el siguiente modelo de degradación lineal.

$$y_{ij} = D_{ij} + \varepsilon_{ij}, i = 1, \dots, 5, j = 1, \dots, 35, \text{ donde } D_{ij} = \beta_{0i} + \beta_{1i} t_{ij} \quad (4)$$

En el modelo anterior los parámetros se consideran aleatorios pues los dispositivos pasan a través de varios pasos en el proceso de fabricación, muchos factores cambian las propiedades iniciales y las tasas de degradación de los dispositivos. Entonces se consideraron β_0 y β_1 como parámetros aleatorios unidad a unidad para toda la población de los dispositivos; además estos parámetros son independientes de los ε_{ij} . Se asume entonces por facilidad y conveniencia que la distribución conjunta de los parámetros es normal bivariada (ver Pinheiro & Bates 2000, Meeker & Escobar 1998, Lu et. al 1997).

Para estimar los parámetros en (4) utilizaremos las funciones implementadas por Pinheiro y Bates en *S-PLUS*; En este caso la función `lmela` cual nos proporciona el vector de medias estimado de los parámetros y la respectiva matriz de varianzas-covarianzas para los mismos. Para este problema en particular después de ajustar el modelo (4) se encuentran problemas de heterocedasticidad, es decir, la varianza del término del error no es constante para todos los niveles del tiempo (t_{ij}); lo anterior se puede notar en la Figura 4 que nos muestra el gráfico de residuales para el modelo ajustado de efectos aleatorios. Para resolver el problema anterior, se debe tener en cuenta que un

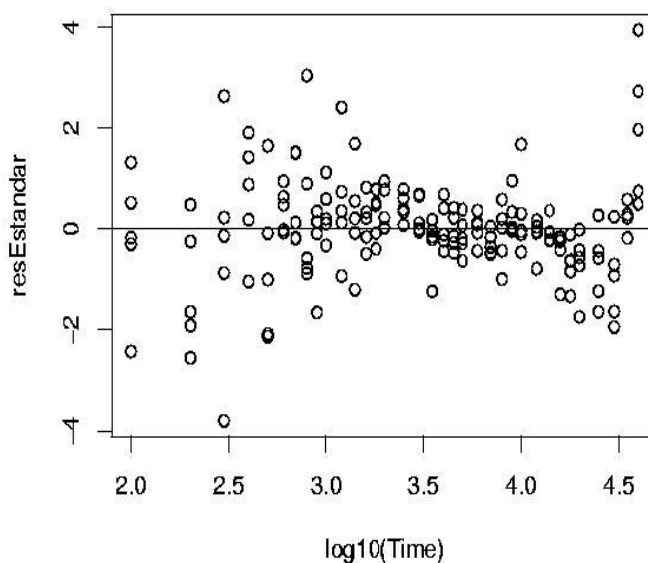


Figura 4: Gráfico de residuales vs $\log_{10}(\text{tiempo})$ para el modelo lineal de efectos aleatorios

proceso de degradación usualmente consiste de tres zonas; la zona de corrida inicial (gran variabilidad), la zona de desgaste normal (la variabilidad es pequeña y estable) y la zona catastrófica de desgaste (la variabilidad vuelve a incrementarse significativamente). Estas tres zonas pueden verse claramente en el grafico de residuales anterior; lo que sugiere que la desviación estándar de los errores debería ser modelada como una función del tiempo t_{ij} . Un modelo que se puede usar para modelar la desviación estándar es usar la transformación logarítmica, sugerida por Nelson (1984) y Lu et. al (1996).

$$\log(\sigma_{ij}) = \alpha_0 + \alpha_1 |t_{ij} - t_0|, \quad (5)$$

donde t_0 es un tiempo escogido de acuerdo al punto donde la desviación estándar se estabiliza. Cuando α_1 es positiva la desviación estándar aumenta a medida que t_{ij} se aleja de t_0 .

Adicionando esta función de la varianza al modelo anteriormente planteado se obtienen las siguientes estimaciones de los parámetros del modelo con la función `lme` del *S-PLUS*.

$$\hat{\mu}_\beta = \begin{pmatrix} -1,0091 \\ 0,4500 \end{pmatrix}, \hat{\Sigma}_\beta = \begin{pmatrix} 0,0075 & -0,0029 \\ -0,0029 & 0,0028 \end{pmatrix}$$

y $\alpha_0 = -4.235$ y $\alpha_1 = 1.4642$ cuando $t_0 = 3.65$. El coeficiente de correlación para los *betas* es $\rho = -0.605$. El gráfico de residuales estandarizados para el modelo con la estructura de varianza puede verse en la Figura 5. Los dos modelos planteados pueden ser comparados con los criterios que se muestran en el Cuadro 1. Entre mas pequeños sean los criterios AIC y BIC mejor es el modelo y por el contrario el *log* de la verosimilitud entre más grande es, mejor es el modelo. Los valores del Cuadro 1 muestran una mejora significativa en los tres criterios para el modelo con estructura de varianza y por lo tanto se utilizará éste para realizar el análisis de degradación explícito. La estimación de

Cuadro 1: Criterios para la comparación de los modelos.

Modelo	Parámetros	AIC	BIC	logLik
Sin estructura de varianza	6	-550.381	-531.392	281.191
Con estructura de varianza	7	-631.925	-609.771	322.962

$F(t)$ en este caso se puede obtener de varias formas, sin embargo se utilizará la siguiente forma cerrada para esta, la cual es explicada con más detalle por Lu, et. al 1996.

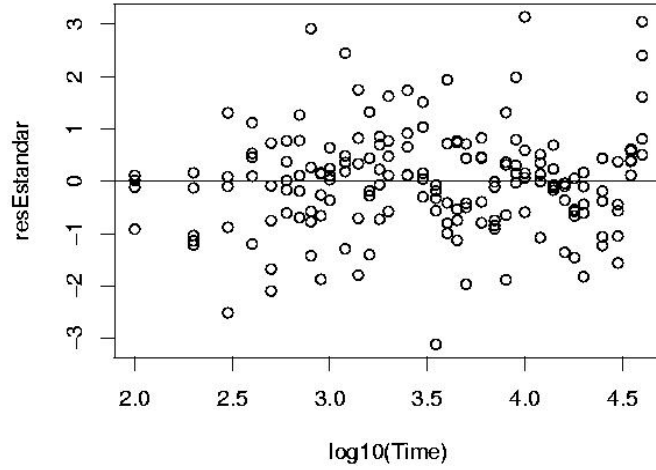


Figura 5: Residuales estandarizados vs $\log_{10}(\text{tiempo})$ para el modelo con estructura de varianza de los ε_{ij} .

$$P(T \leq t) = P\left(\frac{y_f - \beta_0}{\beta_1} \leq t\right) \approx P(y_f \leq \beta_0 + \beta_1 t) = \Phi\left\{\frac{(\mu_{\beta_0} + \mu_{\beta_1} t) - y_f}{(\sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1)^{1/2}}\right\},$$

donde $\Phi(\cdot)$ es la función de probabilidad acumulada para la distribución normal estándar y se utilizan los parámetros estimados por el modelo de efectos aleatorios (ver Figura 6). La cdf anterior puede obtenerse también con el algoritmo de simulación Monte Carlo que se presentó en la sección 3 y se obtienen resultados casi idénticos, por lo tanto cuando no se tiene a la mano una forma cerrada para $F(t)$, el algoritmo de simulación planteado en la sección 3 es una manera relativamente sencilla y buena de obtener la distribución de probabilidad empírica asociada con unos datos de degradación. Los intervalos de confianza bootstrap se construyeron utilizando el algoritmo propuesto por Meeker y Escobar que se describió en la sección 3.1. El Cuadro 2 muestra la estimación de $F(t)$ y su respectivo intervalo bootstrap del 95% para algunos tiempos de falla (en escala logarítmica de base diez) con el método explícito de análisis de datos de degradación.

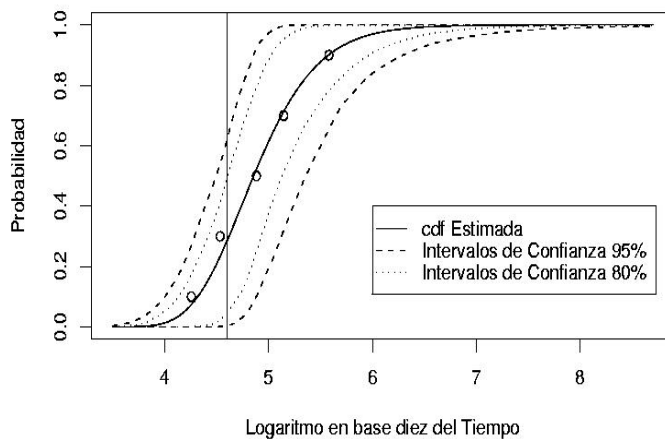


Figura 6: CDF estimada por la metodología de degradación explícita, estimación no paramétrica y los intervalos de confianza bootstrap del 95 % y el 80 % para cdf por degradación.

4.1. Comparación con el análisis clásico aproximado de tiempos de falla aplicado a datos de degradación

En análisis tradicional de tiempos de falla, se tienen los tiempos en que las componentes fallan como tal, es decir dejan de funcionar o el tiempo en que fueron censuradas. Para estos datos se puede estimar la cdf, la función de sobrevivencia, los cuantiles muestrales ya sea de manera paramétrica (ajustando una distribución a los datos) o de manera no paramétrica. Los resultados pueden variar de un método a otro.

Cuadro 2: Estimación de $F(t)$ por el método explícito e intervalos bootstrap del 95 %.

$Log_{10}(t)$	$F(t^*)$	LI del 95 % de confianza	LS del 95 % de confianza
4.34	0.10	0.00041	0.31463
4.86	0.50	0.13259	0.85095
5.54	0.90	0.72433	0.99931
5.81	0.95	0.84689	0.99998

En este artículo, se tienen datos que describen el comportamiento de la degradación de un dispositivo para ciertos niveles del tiempo. Si dichos datos alcanzan un nivel fijo de degradación, se puede considerar que la unidad o el dispositivo falla. Sin embargo, en algunos casos los datos pueden mostrar pocas o ninguna falla, por lo que es necesario extrapolar para obtener el tiempo estimado en que la componente alcanza el nivel de falla. En nuestro caso en particular tenemos trayectorias lineales, las cuales podemos prolongar fácilmente (ver Figura 7); obsérvese que este procedimiento no considera censuras, que es razonable en este caso por la falta de datos. Este procedimiento que acabamos de describir es lo que denominamos como análisis clásico aproximado de tiempos de falla.

Con esta extrapolación se obtiene los siguientes tiempos de falla que se pueden ver en el Cuadro 3 columna dos, dados en escala logarítmica base diez, la CDF estimada por métodos no paramétricos en la columna tres y el respectivo intervalo de confianza del 95 %, en columnas cuatro y cinco (e.g., para el dispositivo tres el tiempo de falla es 76225,5 segundos, al cual le corresponde una probabilidad acumulada de 0.6). A partir del Cuadro 3, el cual resume la in-

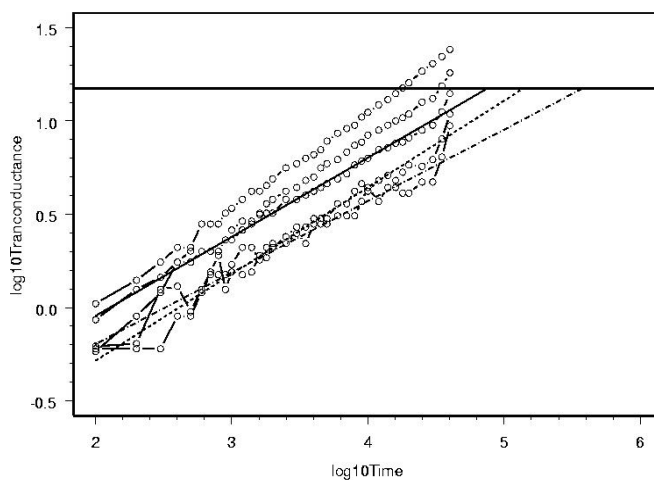


Figura 7: Extrapolación de las trayectorias de degradación para los datos de los semiconductores.

formación del análisis clásico de tiempos de falla por métodos no paramétricos se puede construir la Figura 8. Otra metodología del análisis clásico aproximado que podemos utilizar, consiste en ajustar una distribución a los datos de

Cuadro 3: Tiempos de falla obtenidos por extrapolación, $F(t)$ e intervalos de confianza del 95 %

Dispositivo	$Log_{10}(t)$	$F(t^*)$	LI del 95 %	LS del 95 %
1	4.2553	0.2	0.0272	0.6910
2	4.5326	0.4	0.1002	0.7996
3	4.8821	0.6	0.2004	0.8998
4	5.1451	0.8	0.3089	0.9728
5	5.5815	1.0	1.0000	1.0000

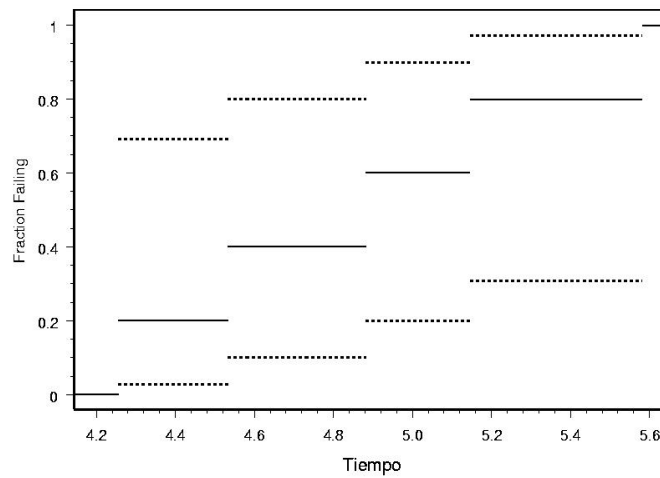


Figura 8: CDF no paramétrica para los tiempos de falla de los semiconductores e intervalos de confianza.

tiempos de falla obtenidos por extrapolación. Realizando este procedimiento paramétrico, encontramos que la distribución que más se ajusta a los datos es una distribución lognormal con parámetros estimados $\hat{\mu} = 1,581$, $\hat{\sigma} = 0,095$ ver gráfico de probabilidad lognormal, Figura 9. Se pudo notar como los diferentes

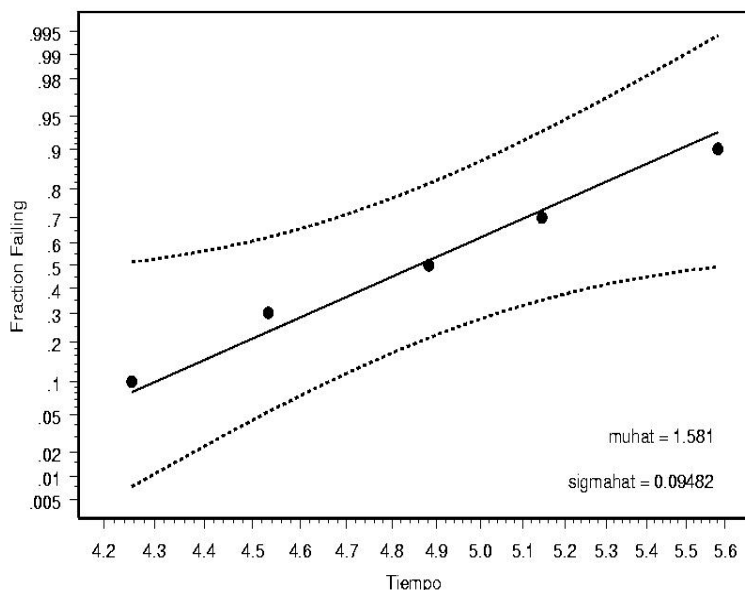


Figura 9: Gráfico de probabilidad lognormal para los tiempos de falla de los dispositivos.

métodos del análisis aproximado proporcionan resultados variables, de los cuales en la siguiente gráfica se comparan las cdf's estimadas por cada uno de los métodos. Se logra apreciar que la estimación paramétrica del análisis aproximado es bastante buena en relación al análisis explícito de degradación, sobre todo en el rango tiempo en que se desarrolló el experimento (ver Figura 10). En este caso la estimación paramétrica aproximada es bastante buena en relación con el resultado del análisis explícito. Esto debido posiblemente al comportamiento lineal de las trayectorias del modelo de degradación. Sin embargo si miramos más detenidamente la gráfica de las comparaciones en el rango de cinco a siete (en escala logarítmica en base diez) notamos cómo la diferencia empieza a ser significativa (ver Figura 11). Obsérvese en la Figura 11 que el ajuste por el método tradicional, sobreestima la función de probabilidad acumulada cdf para proyecciones fuera del rango de los datos en el experimento. Por ejemplo

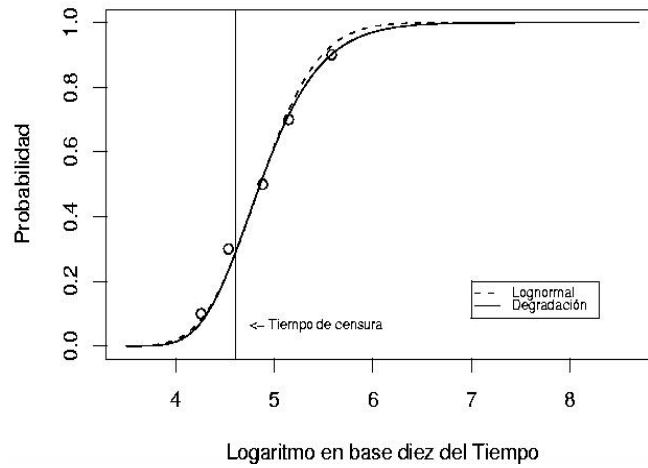


Figura 10: Comparación de las diferentes estimaciones de $F(t)$.

para un $\log_{10}(t) = 5.5263$ (mucho más lejos del tiempo final del experimento) el análisis de degradación nos proporciona una estimación de $F(t)$ de 0.886 mientras que la estimación paramétrica lognormal nos da una estimación de 0.913; es decir, un error relativo al análisis explícito de degradación del orden del 3%.

El Cuadro 4 nos muestra los diferentes valores estimados de $F(t)$ para algunos tiempos fuera del rango de tiempo del experimento, tanto con la metodología explícita de degradación y con el análisis clásico aproximado paramétrico.

Cuadro 4: Diferentes estimaciones de $F(t)$

$\text{Log}_{10}(t)$	CDF por Degradación	CDF lognormal
5.0000	0.6142	0.6168
5.5263	0.8864	0.9132
5.9474	0.9649	0.9838
6.3684	0.9899	0.9979
7.0000	0.9985	0.9999

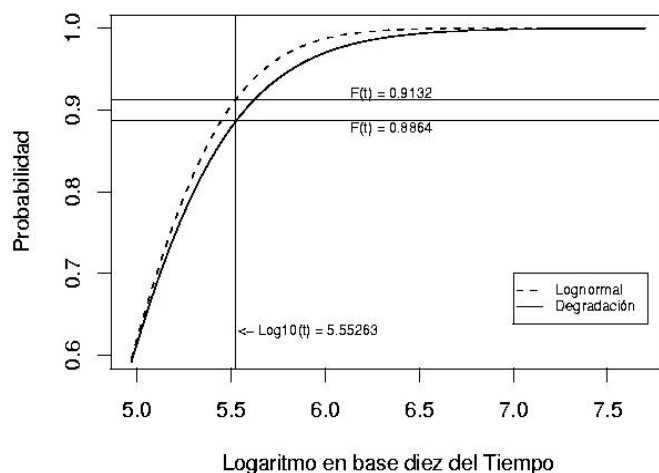


Figura 11: Diferencia de las estimaciones de la CDF para tiempos entre 10^5 y 10^7 segundos.

5. Conclusiones

Este trabajo presenta diferentes metodologías para el manejo de los datos de degradación y su posterior análisis de confiabilidad. Los datos de degradación pueden presentarse de muchas formas y siempre están ligados con la teoría física, por lo tanto, la metodología propuesta en este trabajo se vale de esta teoría y del análisis estadístico para obtener mejores estimaciones, tanto de los parámetros físicos que puedan ser de interés, como de las funciones de probabilidad del tiempo de falla que son fundamentales en un análisis serio de confiabilidad para un producto, unidad o componente de un sistema más complejo.

Las metodologías aproximadas para el análisis de los datos de degradación, presentan importantes desventajas, ya que no tienen en cuenta la física misma de la degradación y además son desarrollados con escasos datos, lo cual puede generar gran incertidumbre en los resultados obtenidos. Para los datos de degradación de semiconductores las metodologías aproximadas presentan buen comportamiento en el rango de tiempo de desarrollo del experimento; sin embargo cuando se trata de hacer extrapolación, estos métodos tienden a sobreestimar $F(t)$, posiblemente por la carencia de suficientes datos para realizar un análisis más “potente” y por que dejan a un lado toda la información del com-

portamiento de la degradación para cada una de las unidades que intervienen en el experimento.

La metodología de análisis explícito de datos de degradación es mucho más compleja de implementar, pues requiere el uso de diferentes metodologías estadísticas que no son de uso común y además pueden requerir un gran gasto computacional, tanto en la estimación de los parámetros del modelo como en la construcción de $F(t)$ y sus respectivos intervalos de confianza, obtenidos por la metodología bootstrap. Esta metodología sin embargo es más adecuada que las metodologías aproximadas cuando se quiere extrapolar la distribución de los tiempos de falla.

Bibliografía

- Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.
- Lawless, J. (2000), 'Statistics in reliability', *Journal of the American Statistical Association* **95**(451), 989–992.
- Lu, C. J. & Meeker, W. Q. (1993), 'Using degradation measures to estimate a time-to-failure distribution', *Technometrics* **35**, 161–174.
- Lu, J. C., Park, J. & Yang, Q. (1996), 'Statistical inference of a time-to-failure distribution derived from linear degradation data', *Technometrics* **39**, 391–400.
- Meeker, W. Q. & Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, Wiley, New York.
- Nelson, W. (1990), *Accelerated Testing: Statistical Methods, Tests Plans and Data Analysis*, Wiley, New York.
- Pinheiro, J. C. & M., B. D. (2000), *Mixed-Effects Models in S and S-PLUS*, Springer, New York.
- Takeda, E. & Susuki, N. (1983), 'An empirical model for device degradation due hot-carrier injection', *IEEE Electron Device Letters* **4**, 11–113.