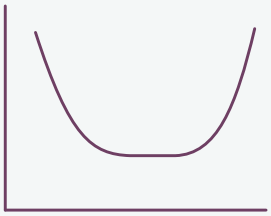
LESSON BOOK — VOLUME II



Reliability Engineering Volume II

Degradation and prognostics, power-system adequacy, observability and decisions

Sixteen lessons · 479 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Reliability Engineering

Volume II

Degradation and prognostics, power-system adequacy, observability and engineering decisions

A lesson sequence, written up as a book

Ricardo Marquez

October 14, 2026

How to read this book

This book is about one question: how an engineered system fails, how likely that is, what it costs, and what should be done about it. Each chapter expands one lesson of a sequence that builds the answer in the order the discipline itself uses,

function and failure → lifetime models → system logic → dependencies and risk → repair and availability → degradation and prognosis → power-system adequacy → measurements and credibility → decisions.

The sequence has nine modules and thirty-six lessons, published in two volumes. Volume I holds Parts I to V, Chapters 1 to 20: the classical reliability theory as the textbooks state it, from failure definitions and lifetime models through system logic, probabilistic risk assessment and repairable systems. This second volume holds Parts VI to IX, Chapters 21 to 36, which extend that theory to degradation and prognostics, power-system adequacy, the sufficiency of measurements, and engineering decisions. Chapter numbers run continuously across the two volumes; a reference to a chapter numbered 20 or below points to Volume I, whose results this volume builds on and does not repeat. The notation chapter is reproduced here so that this volume can be read with Volume I closed.

The book is written for a reader who builds or uses physics-based models of engineered systems and wants the concepts and the vocabulary of reliability engineering stated precisely. Every new quantity is therefore introduced in the same order: its name, its meaning, its units, its equation, and how it is estimated. Where the field uses a shorthand term, a *Jargon* box expands it once and the text then uses the expanded meaning. The vocabulary is the standard one of the discipline, as fixed by Rausand, Barros and Høyland's *System Reliability Theory*, Billinton and Allan's *Reliability Evaluation of Power Systems*, Abur and Expósito's *Power System State Estimation*, the prognostics and health management literature and the other references each chapter names; no term is coined where a standard one exists.

Three systems recur throughout, with growing complexity as the methods allow: an industrial cooling-water pump, a data center cooling plant with three chillers of which two are required, and an electric transmission network whose line ratings depend on the weather. The numbers in the examples are illustrative, chosen to make each relation concrete, and are labeled as such.

One distinction recurs in every chapter: the physical system, which fails or performs; the model, which represents it under stated assumptions; and the evidence, which is never complete. Three kinds of uncertainty are kept apart throughout: aleatory variability in which unit fails and when, epistemic uncertainty about the parameters of the model, and uncertainty about whether the model's structure is adequate at all.

Each chapter closes with a statement of what to retain, a table of its terms with their meaning, units and the section that defines each, and four exercises that can be solved from the chapter alone.

Contents

How to read this book **ii**

Notation and vocabulary **vii**

VI Prognostics and Health Management **1**

21 Degradation and damage accumulation **2**
21.1 Failure as a threshold crossing of a degradation state 2
21.2 Degradation mechanisms and deterministic models 5
21.3 Fatigue, crack growth and thermal aging 9
21.4 Stochastic degradation processes 15
21.5 Condition-based reliability 21
21.6 Operating history, decisions and credibility 25
21.7 What to retain 29
21.8 Exercises 31

22 Condition monitoring and fault diagnosis **33**
22.1 Detection, diagnosis and prognosis 33
22.2 Physical models and condition indicators 36
22.3 Signal processing for condition monitoring 39
22.4 Bayesian diagnosis with competing hypotheses 44
22.5 Sensor faults, observability and the next measurement 50
22.6 Validation and application of diagnosis 56
22.7 What to retain 59
22.8 Exercises 61

23 Remaining useful life estimation **62**
23.1 Remaining useful life as a conditional random variable 62
23.2 Wiener-process remaining useful life 67
23.3 State and parameter uncertainty 69
23.4 Gamma processes and particle filters 75
23.5 Calibration and validation of RUL predictions 80
23.6 Decisions and credibility 82
23.7 What to retain 86
23.8 Exercises 87

24 Predictive maintenance and PHM decisions **89**
24.1 Predictive maintenance as a decision problem 89
24.2 Probability-based maintenance triggers 92
24.3 The cooling-pump replacement decision 94

24.4 Inspection value and scheduling . . . 97
24.5 Maintenance logistics and system availability . . . 103
24.6 Sequential decisions under uncertainty . . . 106
24.7 What to retain . . . 111
24.8 Exercises . . . 113

VII Power-System Reliability and Adequacy 115

25 Generation adequacy and reliability indices 116
25.1 Generation adequacy and the adequacy event . . . 116
25.2 Forced outages and the COPT . . . 119
25.3 Annual adequacy and uncertain demand . . . 125
25.4 Dependent outages and the construction of the COPT . . . 130
25.5 Adequacy, security and sensitivity . . . 134
25.6 Limits and interpretation of adequacy indices . . . 138
25.7 What to retain . . . 140
25.8 Exercises . . . 141

26 State enumeration and capacity outage models 143
26.1 Multistate generating units . . . 143
26.2 Exact enumeration of a multistate fleet . . . 147
26.3 COPT recursion and generating functions . . . 151
26.4 Correlated generator outages . . . 156
26.5 Monte Carlo adequacy simulation . . . 160
26.6 Choice of method and validation . . . 166
26.7 What to retain . . . 169
26.8 Exercises . . . 171

27 Transmission reliability and contingency analysis 173
27.1 Contingencies, security and transmission reliability . . . 173
27.2 The network model and single-outage analysis . . . 176
27.3 Distribution factors and AC security . . . 181
27.4 Probabilistic contingency analysis . . . 184
27.5 Uncertain ratings and corrective actions . . . 189
27.6 Risk attribution and the limits of the analysis . . . 195
27.7 What to retain . . . 200
27.8 Exercises . . . 202

28 Time-varying reliability and cascading failures 204
28.1 Cascades and static contingency analysis . . . 204
28.2 Time-dependent and weather-correlated failures . . . 207
28.3 Conductor thermal dynamics . . . 210
28.4 Protective controls and the cascade event tree . . . 213
28.5 Branching models and chronological simulation . . . 220
28.6 Risk-based decisions and model credibility . . . 225
28.7 What to retain . . . 228
28.8 Exercises . . . 229

VIII State Estimation, Observability, and Model Credibility 231

29 Observability and measurement sufficiency 232
29.1 States, measurements and observability . . . 232
29.2 Observability of the three-bus network . . . 235
29.3 Uncertainty, redundancy and bad data . . . 238
29.4 Sensor placement and measurement sufficiency . . . 245
29.5 Nonlinear, dynamic and parameter observability . . . 247
29.6 Observability in reliability decisions . . . 253
29.7 What to retain . . . 256
29.8 Exercises . . . 257

30 State estimation and bad-data detection 259
30.1 State estimation and its two outputs . . . 259
30.2 Weighted least-squares estimation . . . 262
30.3 Residual analysis and bad-data detection . . . 266
30.4 Critical measurements and topology errors . . . 272
30.5 Robust state estimation . . . 276
30.6 Bayesian state estimation and reliability decisions . . . 279
30.7 What to retain . . . 284
30.8 Exercises . . . 286

31 Identifiability, uncertainty and model adequacy 288
31.1 Parameter estimation and identifiability . . . 288
31.2 Sensitivity matrices and excitation . . . 293
31.3 Fisher information and Bayesian estimation . . . 296
31.4 Model adequacy and model discrepancy . . . 302
31.5 Bayesian model comparison . . . 305
31.6 Applications and the modeling workflow . . . 309
31.7 What to retain . . . 312
31.8 Exercises . . . 314

32 Bayesian updating and value of information 315
32.1 Bayesian updating of competing hypotheses . . . 315
32.2 Entropy and expected information gain . . . 319
32.3 Value of information for decisions . . . 322
32.4 The pump replacement decision . . . 324
32.5 Measurements for parameters and models . . . 329
32.6 Sequential decisions and model credibility . . . 336
32.7 What to retain . . . 342
32.8 Exercises . . . 344

IX Engineering Applications and Decisions 346

33 Reliability standards and engineering practice 347
33.1 Engineering assessments and the role of standards . . . 347
33.2 Transmission planning standards . . . 350
33.3 Functional safety and asset management . . . 353
33.4 Requirements and a two-pump assessment . . . 359
33.5 Verification, validation and traceability . . . 366
33.6 Risk registers, uncertainty and acceptance . . . 369

33.7 What to retain . . . 375
33.8 Exercises . . . 376

34 Data center power and cooling reliability 378
34.1 Data center service reliability . . . 378
34.2 Electrical infrastructure reliability . . . 381
34.3 Cooling infrastructure reliability . . . 384
34.4 Coupled failures and the integrated case . . . 387
34.5 Thermal ride-through and the coupled event tree . . . 392
34.6 Importance, decisions and a unified model . . . 397
34.7 What to retain . . . 402
34.8 Exercises . . . 403

35 Risk importance and decision optimization 405
35.1 Component importance and intervention value . . . 405
35.2 Risk worth measures and parameter sensitivity . . . 409
35.3 Counterfactual risk reduction . . . 413
35.4 An investment portfolio . . . 418
35.5 Value of an upgrade and value of information . . . 424
35.6 Uncertainty and the decision framework . . . 427
35.7 What to retain . . . 431
35.8 Exercises . . . 432

36 Integrated reliability assessment 435
36.1 The integrated assessment . . . 435
36.2 Physical model and measurement audit . . . 438
36.3 From diagnosis to risk and action . . . 443
36.4 The next measurement . . . 448
36.5 Credibility and the assessment report . . . 453
36.6 Synthesis of the two volumes . . . 458
36.7 What to retain . . . 461
36.8 Exercises . . . 462

Notation and vocabulary

Every symbol below is used with the same meaning in every chapter. Time is measured in operating hours unless a chapter states otherwise.

Symbol	Unit	Meaning
T	h	Random time to failure of a nonrepairable item, measured from a defined starting condition; $T \geq 0$.
t	h	A specified operating duration, or the running time variable.
H	h	Operating horizon or mission time over which a requirement must hold.
$R(t)$	–	Reliability (survivor) function, $\Pr(T > t)$.
$F(t)$	–	Cumulative failure probability, $\Pr(T \leq t) = 1 - R(t)$.
$f(t)$	h^{-1}	Failure density, $\mathrm{d}F/\mathrm{d}t$.
$\lambda(t)$	h^{-1}	Hazard function (instantaneous failure rate), $f(t)/R(t)$.
$\Lambda(t)$	–	Cumulative hazard, $\int_0^t \lambda(u)\,\mathrm{d}u$, so that $R(t) = \exp[-\Lambda(t)]$.
λ	h^{-1}	Constant failure rate of the exponential model.
μ	h^{-1}	Constant repair rate, 1/MTTR.
$R(t + \Delta t \mid t)$	–	Conditional reliability: survival over a further Δt given survival to t, $R(t + \Delta t)/R(t)$.
MTTF, MTBF, MTTR	h	Mean time to failure (nonrepairable, or operating duration between restorations); mean time between failures (repairable, including restoration); mean time to repair or restore.
MRL(t)	h	Mean residual life at age t, $\mathbb{E}[T - t \mid T > t]$.
A, $A(t)$, A_{mission}	–	Steady-state availability MTTF/(MTTF + MTTR); instantaneous (point) availability at time t; point availability at the end of a mission of length H, $A(H)$, for an item that starts in the operating state.
R_{system}, R_i	–	Reliability of the system; of component i.
k-out-of-n	–	A system of n units that functions when at least k of them function.
$\mathcal{R}$	R$/yr	Risk, $\sum_i \Pr(S_i)\, C(S_i)$ over a partition of outcomes S_i; expected annual loss when C is monetary.
$C(F)$	R$	Consequence of failure event F, conditional on its occurrence.
$X(t)$	varies	Uncertain system state or inputs at time t (load, weather, equipment condition).
$g(x)$	varies	Performance-margin function; $g \geq 0$ means the requirement is met.

Symbol	Unit	Meaning
$L(t)$, $C(t)$	varies	Required load or demand; available capacity, in the units of the function (m^3/h, kW, A).
$C_{\text{available}}(t)$	varies	Available capacity of a set of units, $\sum_i S_i(t)\, C_i(t)$.
$S_i(t)$	–	Status of unit i at time t: 1 available, 0 unavailable.
$T_{\max}$	°C	Upper temperature limit of a conductor, a room or a component.
I, $R(T)$	A, Ω	Current; temperature-dependent resistance of a conductor (never the reliability function, which is R).
mc_p	J/K	Heat capacity of the modeled conductor length.
q_{solar}, q_{conv}, q_{rad}	W	Solar heat gain; convective and radiative heat loss of a conductor.
S, O, D	–	FMEA severity, occurrence and detection ratings on a 1–10 scale.
RPN	–	Risk priority number, $S \cdot O \cdot D$.
β, η	–, h	Shape and scale parameters of the Weibull distribution.
σ	varies	Standard deviation; stress amplitude, by context.
N	–	Number of components, of load cycles, of trials, by context.
Statistical estimation (Chapters 6–8)		
$\hat{R}(t)$, $\hat{\lambda}$	–, h^{-1}	Estimates from data (Kaplan–Meier survival estimate; maximum likelihood estimate of the failure rate).
r, T_{tot}	–, h	Number of observed failures; total exposure (time on test) of all units, censored ones included.
$L(\theta)$, $\ell(\theta)$, $\mathcal{I}(\theta)$	varies	Likelihood, log-likelihood and Fisher information of a parameter θ.
Gamma(a, b)	–	Gamma distribution with shape a and rate b; prior and posterior of a failure rate.
μ, σ (lognormal)	–	Mean and standard deviation of $\ln T$ in the lognormal model; in that use μ is not the repair rate.
$\Gamma(\cdot)$, $\Phi(\cdot)$	–	Gamma function; standard normal cumulative distribution function.
System logic and probabilistic risk assessment (Chapters 9–15)		
$\phi(\mathbf{x})$, x_i	–	Structure function and component state indicators (1 operating, 0 failed; or the failure indicator, as stated).
q_i	–	Probability of basic event i (component failure or unavailability).
Q_{top}, Q_{sys}	–	Probability of the top event; system failure probability or unavailability.
C_j, P_j	–	Minimal cut set j; minimal path set j (chapter text says when C and P name components instead).
I_i^{B}, I_i^{FV}	–	Birnbaum and Fussell–Vesely importance of component i.
RAW_i, RRW_i	–	Risk achievement worth; risk reduction worth.
β (beta factor), α_k	–	Common-cause beta factor (fraction of a component's failure rate that is common cause; not the Weibull shape); alpha factor, the fraction of failure events involving exactly k components.

Symbol	Unit	Meaning
λ_T, λ_I, λ_C	h^{-1}	Total, independent and common-cause failure rates of one component in a common-cause group.
f_{IE}, f_{S_k}	yr^{-1}	Initiating-event frequency; frequency of accident sequence k.
$\Pr(S_k \mid I)$	–	Probability of sequence k given the initiating event.
LOLE, EUE	h/yr, MWh/yr	Loss-of-load expectation; expected unserved energy.
Repairable systems and availability (Chapters 17–20)		
$N(t)$, T_j, X_j	–, h, h	Number of failures of a repairable item by time t; the j-th failure epoch; the j-th interfailure time.
$W(t)$, $w(t)$	–, h^{-1}	Cumulative intensity (mean function) $\mathbb{E}[N(t)]$ of a counting process; its rate of occurrence of failures (ROCOF), $\mathrm{d}W/\mathrm{d}t$, which is not the hazard of one lifetime.
$M(t)$, $m(t)$	–, h^{-1}	Renewal function (mean number of renewals) and renewal density of a perfect-repair process.
b, η (power law)	–, h	Shape and scale of the power-law intensity $w(t) = (b/\eta)(t/\eta)^{b-1}$ of a nonhomogeneous Poisson process.
V_n	h	Virtual age of an item after its n-th repair (Kijima models).
$P_i(t)$, $\boldsymbol{\pi}$	–	Probability of being in state i at time t of a Markov model; its stationary distribution.
$\mathbf{Q}$, q_{ij}	h^{-1}	Generator (transition-rate) matrix of a continuous-time Markov chain and its off-diagonal rates; not the top-event probability Q_{top} nor the basic-event probability q_i.
ρ	–	Ratio of failure rate to repair rate, λ/μ; by context, a discount rate.
f_{F}, MDT	yr^{-1}, h	Frequency of entering the failed state; mean down time per failure, so that unavailability equals $f_{\mathrm{F}} \cdot \mathrm{MDT}$.
T_p, $C(T_p)$	h, R$/h	Preventive replacement age; long-run cost rate of the age-replacement policy.
c_p, c_f, c_D	R$, R$, R$/h	Cost of a preventive replacement, of a failure replacement, of one hour of downtime.
s, $\nu\tau$	–, –	Spare-parts stock level; expected spare demand over a lead time τ at demand rate ν.
Degradation, prognostics and health management (Chapters 21–24)		
$D(t)$, D_0, D_{crit}	varies	Degradation (damage) state, its initial value and the failure threshold; failure is the first crossing. Not the FMEA detection rating D.
μ_D, σ_D, $B(t)$	varies	Drift and diffusion of the Wiener degradation model $D(t) = D_0 + \mu_D t + \sigma_D B(t)$; standard Brownian motion. μ_D is not the repair rate.
α_D, β_D	–, varies	Shape rate and scale of a gamma degradation process.
T_{f}, $\mathrm{RUL}(t)$, $\mathrm{RUL}_\alpha(t)$	h, d	First-passage (failure) time; remaining useful life at time t, $T_{\mathrm{f}} - t$ given survival; its α-quantile.

Symbol	Unit	Meaning
e, z	varies, –	Residual, observed minus model-predicted value; standardized residual.
$H_0, H_1, \ldots$	–	Diagnostic hypotheses (healthy, each fault, each sensor fault) with posterior probabilities $\Pr(H_i \mid \mathbf{y})$.
$\mathcal{H}$	bit	Shannon entropy of the hypothesis probabilities; expected information gain of a test.
Y^+, Y^-	–	Positive and negative outcome of an inspection, characterized by sensitivity and specificity.
τ_{L}, p^*	d, –	Action lead time; decision threshold on the failure probability.
EVSI, EVPI	R$	Expected value of sample (test) information; of perfect information.
Power-system adequacy (Chapters 25 onward)		
L_t, C_t, S_t	MW	Hourly demand; available capacity; shortfall $(L_t - C_t)^+$.
FOR_i, U_i	–	Forced outage rate of unit i: its unavailability, a probability, not a rate.
LOLP_t, LOLH	–, h/yr	Loss-of-load probability in interval t; loss-of-load hours, the LOLE on an hourly basis.
X, $P(X \geq x)$	MW, –	Capacity on outage and its cumulative probability in a capacity outage probability table (COPT).
K_i, $\mathrm{PTDF}_{\ell,k}$, $\mathrm{LODF}_{\ell,m}$	MW, –, –	Rating of unit i; power transfer and line outage distribution factors of line ℓ.
θ_i, V_i, x_{ij}, P_{ij}, $P_{ij}^{\max}$	rad, pu, pu, MW	Bus voltage angle and magnitude; line reactance; DC line flow $(\theta_i - \theta_j)/x_{ij}$; line rating.
$\mathbf{B}$	pu	DC susceptance matrix of the network.
λ_{b}, N_{c}	–, –	Branching ratio (mean offspring number) of a cascade model; cascade size.
τ_{th}	min	Thermal time constant of a conductor.
State estimation and identifiability (Chapters 29–31)		
$\mathbf{x}$, $\mathbf{y}$	varies	State vector; measurement vector (written $\mathbf{z}$ in the power-system literature), $\mathbf{y} = h(\mathbf{x}) + \boldsymbol{\varepsilon}$.
$\mathbf{R}$, $\mathbf{W}$	varies	Measurement error covariance (bold; not the reliability function) and its inverse, the weight matrix.
$\mathbf{H}$, $\mathbf{G}$	varies	Measurement Jacobian $\partial h/\partial \mathbf{x}$ (bold; not the horizon) and the gain (information) matrix $\mathbf{H}^{\mathsf{T}}\mathbf{W}\mathbf{H}$.
$J(\mathbf{x})$, $\hat{\mathbf{x}}$	–, varies	Weighted least-squares objective; its minimizer, the state estimate.
$\mathbf{e}$, $\mathbf{S}$, $\boldsymbol{\Omega}$, e_i^{N}	varies	Residual $\mathbf{y} - h(\hat{\mathbf{x}})$; residual sensitivity matrix; residual covariance $\mathbf{SR}$; normalized residual $e_i/\sqrt{\Omega_{ii}}$.
$\boldsymbol{\theta}$, $\mathbf{S}_\theta$	varies	Parameter vector (bold; not the bus angles) and its sensitivity matrix $\partial h/\partial \boldsymbol{\theta}$.
$\sigma_1 \geq \sigma_2 \geq \ldots$, κ	–	Singular values of the scaled sensitivity matrix; condition number $\sigma_1/\sigma_{\min}$.
$\delta(\cdot)$, B_{12}	varies, –	Model discrepancy term; Bayes factor between models 1 and 2.
Value of information and engineering decisions (Chapters 32–36)		
$L(a, H)$, $\rho(a)$	R$	Loss of action a when hypothesis H holds; its expected loss under the current probabilities.

Symbol	Unit	Meaning
C_{now}, C_{sample}, C_{perfect}	R$	Minimum expected loss with current information; with the outcome of a test; with the true hypothesis known.
EIG, $I(H;Y)$, D_{KL}	bit or nat	Expected information gain of a test, equal to the mutual information between hypothesis and outcome; Kullback–Leibler divergence.
$\mathcal{R}(a)$, $\Delta\mathcal{R}(a)$	varies	Risk after intervention a; its reduction from the baseline $\mathcal{R}_0$.
x_i, c_i, B	–, R$/yr, R$/yr	Selection variable of project i (1 selected, 0 not); its annualized cost; the budget.
v, H_{rel}	R$/MWh, h/yr	Consequence value of unserved energy; hours per year for which an analyzed operating condition applies.
CVaR_α	varies	Conditional value at risk: the expected loss in the upper $(1-\alpha)$ tail.
A_{IT}, $Y(t)$	–	Availability of the required IT service; its indicator at time t.
C_{eff}, t_{crit}	MJ/K, min	Effective thermal capacitance of a room; time for its temperature to reach the limit after a loss of cooling.
λ_{DU}, T_{p}	h^{-1}, h	Dangerous undetected failure rate of a safety function; proof-test interval (not the replacement age of Chapter 18).
PFD_{avg}, PFH	–, h^{-1}	Average probability of dangerous failure on demand; average frequency of dangerous failure per hour.

Vocabulary

The book uses the standard vocabulary of reliability engineering and probabilistic risk assessment, as fixed by the textbooks named in the preface. Where a term has a common shorthand, the shorthand is given in parentheses once and the full term is used thereafter.

Item A component, subsystem or system treated as a unit of analysis.

Required function What an item must do for the system to count as operating; stated as a measurable performance criterion.

Failure The termination of an item's ability to perform its required function; a functional event, which need not coincide with physical breakage.

Fault The state of an item after failure, or a defect that may lead to failure; a fault can exist without a failure having occurred.

Failure mode The manner in which the failure is observed (no flow, low flow, leak, spurious trip).

Failure mechanism The physical, chemical or electrical process that produces the failure (fatigue, corrosion, wear, electromigration).

Failure cause The circumstance that activates the mechanism (overload, contamination, misuse).

Failure effect The consequence of the mode at the next level of the system hierarchy.

System boundary The set of items and interfaces inside the reliability model; what is outside is an input or an assumption.

Reliability Probability that an item performs its required function under specified conditions for a specified period.

Availability Probability that an item is in an operating state at a given time, or the long-run fraction of time it is operating, restoration included.

Maintainability Probability that a failed item is restored within a specified time under specified procedures.

Risk The combination of the probability of outcomes and their consequences.

RAMS Reliability, availability, maintainability and safety, considered together.

Hazard function (failure rate) The conditional failure intensity at age t given survival to t; constant for the exponential model, increasing with wear-out, decreasing with early failures.

Bathtub curve The hazard shape with an early decreasing part, a constant middle and an increasing wear-out part.

Mission reliability Reliability over a stated mission time without credit for repair.

Repairable system An item that is restored to operation after failure; described by counting and renewal processes rather than a single lifetime.

FMEA, FMECA Failure mode and effects analysis; the same with a criticality analysis.

Fault tree analysis (FTA) A deductive logic model of the combinations of events that produce a defined top event.

Minimal cut set A smallest set of component failures sufficient to cause the top event.

Common-cause failure Failure of several items from one shared cause, which defeats the independence that redundancy assumes.

Importance measure A quantity that ranks components by their contribution to system unreliability (Birnbaum, Fussell–Vesely, risk achievement worth, risk reduction worth).

Probabilistic risk assessment (PRA) The scenario-based quantification of risk: initiating events, event trees, fault trees and consequences.

Aleatory and epistemic uncertainty Variability inherent in the process, and lack of knowledge about its parameters or structure.

Censoring A lifetime observation that ends before failure, so that only a lower bound of T is known (right-censoring); interval and left censoring bound T from both sides or from above.

Kaplan–Meier estimator The nonparametric product-limit estimate of the reliability function from censored data.

Maximum likelihood estimate (MLE) The parameter value that maximizes the likelihood of the observed failures and censored records.

Credible interval A Bayesian interval that contains the parameter with stated posterior probability; distinct from a confidence interval.

Posterior predictive reliability The reliability of a new unit obtained by averaging the model over the posterior of its parameters.

Structure function The Boolean function that gives the system state from the component states.

Reliability block diagram A success-logic diagram in which the system operates when a path of operating blocks connects input to output.

Cut set, path set A set of component failures that causes system failure; a set of operating components that keeps the system operating. Minimal when no proper subset qualifies.

Rare-event approximation The sum of minimal-cut-set probabilities as an estimate of the top-event probability; an upper bound that is accurate when the probabilities are small.

Inclusion–exclusion The exact expansion of the probability of a union of events in alternating sums over intersections.

Coherent system A system in which no component failure can restore operation and every component matters.

Common-cause component group A set of redundant, similar components exposed to the same common causes; the unit to which beta and alpha factors apply.

Coupling factor The shared property (design, environment, procedure) through which one root cause affects several components.

Initiating event The event that starts an accident sequence; characterized by a frequency per year.

Event tree An inductive model that branches on the success or failure of each protection function after an initiating event.

Accident sequence, end state One path through an event tree; the consequence category in which it ends.

Frequency–consequence curve The annual frequency of exceeding each consequence level; the risk profile beyond its expectation.

Integrated PRA A risk assessment that links initiating-event frequencies, fault trees, event trees, consequences and uncertainty into one traceable model; organized by the risk triplet (what can go wrong, how likely, with what consequences).

Counting process, ROCOF The number of failures of a repairable item by time t; the rate of occurrence of failures, the derivative of its mean, as distinct from the hazard of one lifetime.

Homogeneous and nonhomogeneous Poisson process (HPP, NHPP) Counting processes with a constant, respectively time-varying, intensity; the NHPP with a power-law intensity is the Crow–AMSAA model.

Perfect, minimal and imperfect repair Repair that returns an item to as-good-as-new, to as-bad-as-old, or to a state between the two (virtual age).

Renewal process The counting process of a perfectly repaired item: independent identically distributed interfailure times.

Age replacement The maintenance policy that replaces an item at failure or at age T_p, whichever comes first; optimized through the renewal-reward theorem.

Continuous-time Markov chain (CTMC) A state model with exponential holding times and a generator matrix of transition rates; solved by the Kolmogorov forward equations or its stationary distribution.

Frequency and duration The description of unavailability as the product of the failure frequency and the mean down time.

Degradation, failure threshold The gradual loss of a property that fixes an item's ability to function; failure is the first crossing of the threshold by the degradation state.

Prognostics and health management (PHM) Detection, diagnosis and prognosis of equipment condition from measurements, and the maintenance decisions built on them.

Remaining useful life (RUL) The time from now until the degradation state reaches its threshold, a random variable conditional on the present evidence.

Condition indicator, residual A quantity computed from measurements that tracks condition; the difference between a measurement and its model prediction, which detection and diagnosis test.

Fault detection and isolation (FDI) Deciding that a fault is present, and which one; with bad-data detection, the discipline's names for separating physical faults from sensor faults.

Value of information The expected reduction in decision loss from a test or inspection before deciding; sample information (EVSI) and perfect information (EVPI).

Generation (resource) adequacy Whether installed resources suffice to meet demand at every hour under random outages; measured by LOLP, LOLE and EUE.

Forced outage rate (FOR) The unavailability of a generating unit, the fraction of time it is out of service on forced outage; a probability.

Capacity outage probability table (COPT) The distribution of capacity on outage for a fleet, built by recursive convolution of the units' two-state models.

Multistate unit, derated state A generating unit modeled with partial-capacity states between full availability and outage.

Contingency, N−1 criterion The outage of one or more network elements; the requirement that the system withstand the loss of any single element without violating limits.

Cascading failure A sequence of dependent outages in which each failure raises the stress on, and the failure probability of, the remaining elements.

Observability Whether the states of interest can be inferred from the available measurements; the rank condition on the measurement Jacobian.

State estimation The weighted least-squares (or robust, or Bayesian) inference of the system state from redundant measurements.

Critical measurement A measurement whose removal makes the system unobservable; its residual is always zero, so its error cannot be detected.

Bad-data detection and identification The chi-square test on the weighted residual sum, and the largest normalized residual test that names the suspect measurement.

Topology error A mismatch between the network model's switching state and the physical network, which a residual test can confuse with a sensor fault.

Identifiability Whether the parameters of a model can be determined from the measurements, structurally (from the model) and practically (given the noise and the design).

Model discrepancy The systematic difference between a model's prediction and the physical system, distinct from measurement error.

Expected information gain The expected reduction in the entropy of the hypothesis probabilities from a test; a measure of knowledge, not of decision value.

Counterfactual risk reduction The change in risk from a hypothetical intervention on the system, as distinct from conditioning on an observed state.

Portfolio optimization The selection of projects under a budget, with system risk recomputed for each combination because project benefits interact.

Verification, validation Whether a model is implemented and solved correctly; whether it represents the physical system adequately for its intended use.

Traceability The documented connection from a requirement through the model, its inputs and evidence to the result and the decision.

Safety integrity level (SIL) The band of required failure measures for a safety function, from 1 to 4.

Thermal ride-through The time a room's thermal inertia allows between a loss of cooling and the crossing of the temperature limit.

Part VI

Prognostics and Health Management

Chapter 21

Degradation and Damage Accumulation

21.1 Failure as a threshold crossing of a degradation state

Chapters 17 to 20 of Volume I represented equipment as operating, failed or under repair. They modeled failure and restoration rates, availability and the optimization of maintenance, and in all of them a failure was an event whose rate was estimated from the history of earlier failures. Those methods leave one question unanswered: what happens physically inside a component before it fails. A bearing gradually loses material; the paper insulation of a transformer accumulates thermal damage; a battery loses capacity; a welded joint develops a fatigue crack that grows with every load cycle. Such processes often evolve over months or years before they reach a critical condition, and if their evolution can be modeled, the failure can be predicted, the remaining life of the component estimated, and maintenance planned on the condition of the equipment rather than on its age. This chapter, which opens the part of the book on prognostics and health management, asks

whether equipment failure can be predicted from the evolving physical condition of the equipment, rather than only from historical failure frequencies.

The answer is a change of the object that is modeled. Instead of the failure event, the model describes the physical process that produces it: a degradation state that evolves under the operating conditions and a failure threshold that it eventually crosses. The chapter first states this framing; it then describes the physical mechanisms and the deterministic models of degradation, works the physics-based models of fatigue, crack growth and thermal aging, introduces the stochastic degradation processes whose first-passage times are lifetime distributions, conditions the reliability of one particular component on its measured condition, and ends with operating history, engineering decisions and the credibility of degradation models.

Jargon: degradation, degradation state, failure threshold, first-passage time, damage accumulation, PHM, condition monitoring, diagnosis, prognostics, remaining useful life, condition-based maintenance

Degradation is a progressive change in the physical or functional condition of an item that reduces its ability to perform its required function. The *degradation state* $D(t)$ is a measurable or inferred quantity that describes that condition at time t: a crack length, a wall penetration, a wear depth, a normalized damage index or a fraction of capacity lost. The *failure threshold* D_{crit} is the value of the degradation state at which the item no longer meets its required function. The *first-passage time* T_{f}, also called the threshold-crossing time, is the first time at which $D(t)$ reaches D_{crit}. *Damage accumulation* is the growth of a degradation state that does not recover, such as fatigue damage or metal loss. *Prognostics and health management* (PHM) is the discipline that monitors the condition of equipment, diagnoses its faults, predicts its future degradation and failure, and uses the prediction for maintenance and operating decisions. Its three activities have standard names: *condition monitoring*, the observation of indicators of condition; *diagnosis*, the determination of the mechanism responsible for an observed change; and *prognostics*, the prediction of future degradation and of the *remaining useful life* (RUL), the time from the present until the item can no longer perform its required function. *Condition-based maintenance* is maintenance whose timing is decided from the observed condition of the item rather than from its age or a fixed calendar.

Key idea

A failure mechanism that accumulates damage turns the time to failure into the first time a degradation state crosses a threshold. Modeling the state, its dependence on the operating history and its random variation gives a distribution of the time to failure for each individual component, conditioned on what has been measured about it.

21.1.1 From failure events to degradation trajectories

In the lifetime models of Chapter 5 of Volume I, the time to failure of an item is a random variable and the reliability function is the probability that it exceeds a stated duration,

$$R(t) = \Pr(T_{\text{f}} > t), \tag{21.1}$$

estimated from the failure times of a population of similar items. Nothing in that description refers to the physical condition of the item between its installation and its failure. A degradation model introduces that condition explicitly as a degradation state $D(t)$, measured directly or inferred from measurements. Examples are the length of a fatigue crack, the wear of a bearing, the depth of corrosion penetration in a pipe wall and the cumulative damage of an insulation. (D here is the degradation state of Chapter 4 of Volume I; it is not the FMEA detection rating D of that chapter.) The item fails when the degradation reaches a critical threshold, so that the time to failure is the first time at which the state reaches it:

$$T_{\text{f}} = \inf\{\, t \geq 0 : D(t) \geq D_{\text{crit}} \,\}. \tag{21.2}$$

The infimum is the greatest lower bound of the set of times at which the state is at or above the threshold, which for a continuous path is the first such time. T_{f} is called a *first-passage time* or *threshold-crossing time*. When failure is defined by the threshold, T_{f} is the random time to failure T of Volume I, and equation (21.1) gives its reliability function.

Chapter 4 stated the deterministic version of this model in Section 4.3.3: with a damage rate $\mathrm{d}D/\mathrm{d}t = k\, s(t)^n$, in which k is a degradation coefficient, $s(t)$ the intensity of the stress and n the stress sensitivity exponent, a constant stress gives the time to failure $(D_{\text{crit}} - D_0)/(ks^n)$,

and the cumulative failure probability over a horizon H is $\Pr(D(H) \geq D_{\text{crit}})$ when the damage never decreases (equation (4.11)). This chapter extends that model in three directions: to physical laws for particular mechanisms, to degradation that is random along its path, and to conditioning on measurements of one particular item.

21.1.2 Failure as a threshold crossing

Figure 21.1 draws the framing. The horizontal axis is operating time and the vertical axis the degradation state. The blue curve is one degradation path that starts at an initial value D_0 and increases, slowly at first and faster later; the dashed red line is the critical threshold D_{crit}; the dot marks the first crossing and the dashed vertical line below it the failure time T_{f}. The arrow at an intermediate time t is the remaining distance to the threshold, $D_{\text{crit}} - D(t)$, which a later section uses to predict the remaining life. Failure occurs where the path meets the threshold, not at a predefined age of the component: a component whose path rises faster fails earlier at the same age.

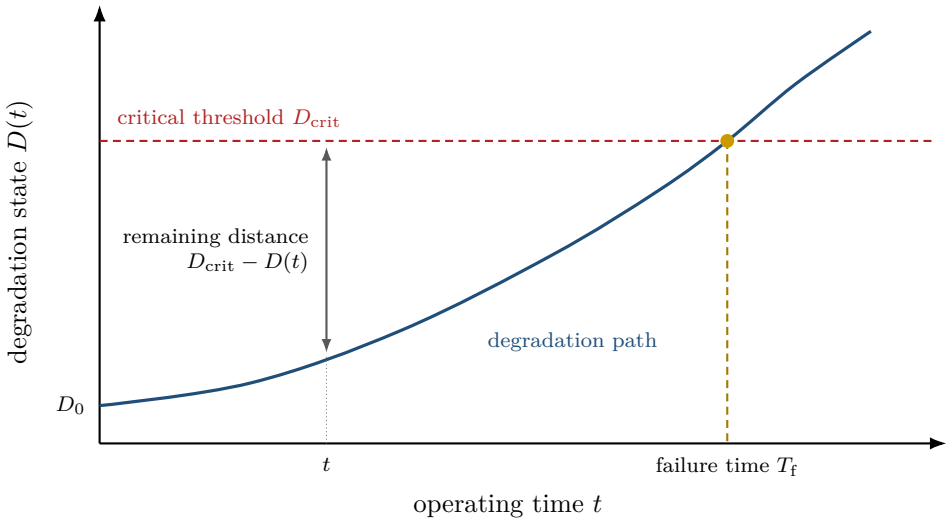


Figure 21.1: An illustrative increasing degradation path. Failure occurs at the first crossing of the critical threshold D_{crit}, the first-passage time T_{f} of equation (21.2), not at a predefined component age. The double arrow is the remaining distance to the threshold at an intermediate time t.

This changes the reliability problem. Instead of modeling the failure events directly, the model describes the physical process that produces them, in the chain

$$\text{operating conditions} \;\rightarrow\; \text{damage evolution} \;\rightarrow\; \text{threshold crossing} \;\rightarrow\; \text{failure}.$$

The first arrow is a physical law that turns loads, temperatures and cycles into a rate of damage; the second is the accumulation of that rate up to the threshold; the third is the definition of failure against the required function. The chain is the stress-strength model of Section 4.3.1 with a strength that falls over time. There, an item with strength C under a load L fails when the margin $g = C - L$ is negative, with probability $p_f = \Pr(L > C)$; here, degradation lowers the capacity of the item until the margin is consumed, and the threshold D_{crit} is the degradation state at which the margin reaches zero. Each link of the chain can be modeled deterministically, which gives a predicted failure time, or stochastically, which gives a distribution of failure times. The sections that follow develop both.

Principle 21.1. *Under a degradation model, the time to failure is the first-passage time of the degradation state through the failure threshold. The reliability function is the probability that the whole degradation path stays below the threshold up to the stated time.*

21.2 Degradation mechanisms and deterministic degradation models

This section states what degradation is relative to a required function, lists the physical mechanisms that produce it with their degradation states, and develops the deterministic models in which the state follows an ordinary differential equation: linear, nonlinear and dependent on the operating conditions.

21.2.1 Degradation and the required function

Degradation is a progressive change in a component's physical or functional condition. It becomes relevant to reliability when it reduces the component's ability to satisfy its required function, in the sense of Chapter 2 of Volume I. Much degradation happens while the component is still operating. A battery remains in service while it loses capacity, and it fails only when its available energy or power falls below what its application requires. A transmission-line conductor accumulates fatigue damage at its suspension clamps while it continues to carry its load. A cooling pump continues to deliver flow while the wear of its bearings and its vibration increase, and it fails, for the cooling-water system of Chapter 4 of Volume I, when it can no longer deliver the required $100\,\mathrm{m}^3/\mathrm{h}$ or when the bearing seizes.

The degradation variable must therefore connect to a meaningful failure criterion. A crack length matters because a critical length exists at which the remaining section fractures; a capacity fraction matters because the application sets a minimum energy; a vibration velocity matters only through the bearing condition it indicates and the seizure or flow loss that condition leads to. A degradation state without a threshold tied to the required function is a monitored quantity, not yet a reliability model.

21.2.2 Common physical mechanisms

The mechanisms that produce degradation were surveyed in Chapter 4; five of them recur in prognostics. *Fatigue*: repeated stress cycles initiate and propagate cracks, in rotating shafts, turbine components, transmission structures and mechanical connections. *Corrosion*: chemical or electrochemical reactions gradually remove material or degrade protective layers, and eventually threaten pressure containment or structural strength. *Wear*: contact, lubrication conditions, particle contamination and repeated loading damage surfaces and change friction, vibration and clearances. *Thermal and electrical aging*: temperature and electrical stress accelerate the deterioration of insulation, polymers and other materials, in transformers, motors and power electronics. *Electrochemical degradation*: batteries lose usable capacity and their internal resistance grows, through calendar aging, cycling, temperature and the operating conditions. Table 21.1 lists, for each mechanism, a degradation state that describes it, the stresses that drive it, a failure criterion and the model of this chapter that represents it.

The mechanisms differ, but many of them can be described by one mathematical structure, a first-order evolution law for the degradation state,

$$\frac{\mathrm{d}D}{\mathrm{d}t} = g_D(D, \mathbf{u}(t), \mathbf{z}(t), \boldsymbol{\theta}), \tag{21.3}$$

where $D(t)$ is the degradation state, $\mathbf{u}(t)$ the operating inputs and loading (speed, current, flow, cycles), $\mathbf{z}(t)$ the environmental conditions (temperature, humidity, water chemistry), $\boldsymbol{\theta}$ the

Table 21.1: Degradation mechanisms, their degradation states, driving stresses, failure criteria and the models of this chapter.

Mechanism	Degradation state $D(t)$	Driving stresses	Failure criterion	Model in this chapter
Fatigue	Miner damage sum; crack length a	stress range, number of cycles	$D = 1$; $a = a_c$	S-N curve and Miner rule; Paris law
Corrosion	wall penetration (mm)	water chemistry, temperature, flow, coating	maximum acceptable penetration	linear degradation; gamma process
Wear	bearing damage index; clearance	contact load, speed, lubrication, particles	seizure; flow below requirement	Wiener process; speed-dependent rate
Thermal and electrical aging	insulation damage; equivalent aging time	temperature, electric stress	end of insulation life	Arrhenius rate, cumulative thermal damage
Electro-chemical	capacity loss $1 - Q_{\mathrm{cap}}/Q_{\mathrm{cap}}(0)$; resistance growth	calendar time, cycles, temperature, state of charge	80 % capacity retention; power limit	calendar and cycling model

physical parameters of the model, and g_D the degradation evolution law, with the units of D per unit time. (g_D is not the performance margin g of Volume I.) The law should be supported by a physical mechanism or by adequate empirical evidence; equation (4.8) of Chapter 4 is the special case $g_D = k\, s(t)^n$, which depends on the stress alone.

Not every observed change in a sensor signal is evidence of progressive damage. The vibration of a pump may increase because its operating speed changed, not because its bearing deteriorated. The degradation state is a property of the component; a measurement responds to the state and to the operating conditions together. The distinction is taken up in Section 21.5.2 and, in full, in the treatment of fault diagnosis in Chapter 22.

21.2.3 Linear degradation

The simplest degradation model assumes a constant rate of damage accumulation,

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \kappa, \qquad \kappa > 0, \tag{21.4}$$

where κ is the degradation rate in the units of D per unit time. With the initial condition $D(0) = D_0$, integration gives

$$D(t) = D_0 + \kappa t. \tag{21.5}$$

Failure occurs when $D(t) = D_{\mathrm{crit}}$, so $D_0 + \kappa\, T_{\mathrm{f}} = D_{\mathrm{crit}}$, and solving for the failure time,

$$T_{\mathrm{f}} = \frac{D_{\mathrm{crit}} - D_0}{\kappa}. \tag{21.6}$$

Example 21.1 (Corrosion penetration)**.** A pipe wall corrodes at a penetration rate $\kappa = 0.08\,\mathrm{mm/yr}$. The penetration measured at the latest inspection is $D_0 = 0.60\,\mathrm{mm}$, and the maximum acceptable penetration, set by the wall thickness needed to contain the operating pressure, is $D_{\mathrm{crit}} = 2.0\,\mathrm{mm}$ (illustrative values). Then

$$T_{\mathrm{f}} = \frac{2.0\,\mathrm{mm} - 0.60\,\mathrm{mm}}{0.08\,\mathrm{mm/yr}} = \frac{1.40}{0.08}\,\mathrm{yr} = 17.5\,\mathrm{yr},$$

counted from the inspection. The deterministic model predicts another 17.5 years until the threshold is reached.

That prediction is not yet a probabilistic reliability estimate. It assumes that the corrosion rate and the threshold are known exactly and remain constant, whereas the rate of corrosion depends on the water chemistry, the temperature, the flow velocity, the condition of protective coatings and other factors that vary along the pipe and over the years. To compute a failure probability, the uncertain quantities and the random variation of the process must be represented, which Section 21.4 does for this same pipe with a gamma process.

21.2.4 Nonlinear degradation

Degradation often accelerates as damage accumulates, because the damage itself raises the rate: a longer crack concentrates more stress at its tip, a worn bearing vibrates more and wears faster. The simplest model of such acceleration makes the rate proportional to the state,

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \kappa D, \tag{21.7}$$

with κ now in units of inverse time. Separating the variables, $\mathrm{d}D/D = \kappa\,\mathrm{d}t$, and integrating from D_0 at time zero, $\ln(D/D_0) = \kappa t$, so

$$D(t) = D_0\,\mathrm{e}^{\kappa t}. \tag{21.8}$$

Failure occurs when $D_0\,\mathrm{e}^{\kappa T_\mathrm{f}} = D_\mathrm{crit}$, that is, at

$$T_\mathrm{f} = \frac{1}{\kappa}\,\ln\!\left(\frac{D_\mathrm{crit}}{D_0}\right). \tag{21.9}$$

The model requires $D_0 > 0$: a state that starts at zero never grows under equation (21.7), which is a reminder that accelerating models describe the growth of damage that already exists, not its initiation.

A more general law raises the state to a damage exponent n_D,

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \kappa D^{n_D}. \tag{21.10}$$

For $n_D > 1$ the rate increases strongly with damage. Separating the variables, $D^{-n_D}\,\mathrm{d}D = \kappa\,\mathrm{d}t$, and integrating,

$$\frac{D_0^{1-n_D} - D(t)^{1-n_D}}{n_D - 1} = \kappa t, \qquad n_D \neq 1,$$

so that the failure time is

$$T_\mathrm{f} = \frac{D_0^{1-n_D} - D_\mathrm{crit}^{1-n_D}}{(n_D - 1)\,\kappa}. \tag{21.11}$$

For $n_D > 1$ the solution $D(t) = [D_0^{1-n_D} - (n_D - 1)\kappa t]^{-1/(n_D-1)}$ grows without bound as t approaches $D_0^{1-n_D}/[(n_D - 1)\kappa]$, a finite time; in a physical system the threshold, or a different mechanism such as unstable fracture, is reached first. Behavior of this kind appears in some crack-growth and wear processes, although the actual laws are often more complicated: the Paris law of Section 21.3.4 is equation (21.10) with the crack length as the state and $n_D = m/2$.

Figure 21.2 compares the three laws on one example. The horizontal axis is operating time in hours and the vertical axis a normalized damage. All three paths start at $D_0 = 0.10$ and reach the threshold $D_\mathrm{crit} = 1$ at the same time, 900 h (the example of Section 4.3.3), which fixes their rates: $\kappa = 0.001\,\mathrm{h}^{-1}$ for the linear law, $\kappa = \ln(10)/900\,\mathrm{h} = 2.558 \times 10^{-3}\,\mathrm{h}^{-1}$ for the exponential law, and, from equation (21.11) with $n_D = 2$, $\kappa = (10 - 1)/900 = 0.01\,\mathrm{h}^{-1}$ for the quadratic law. The dotted horizontal line marks half the distance to the threshold, $D = 0.55$. The linear path reaches it at 450 h, the exponential path at 666 h and the quadratic path at 818 h. Models that agree on the failure time disagree strongly on the state along the way: under

the quadratic law, a component that shows less than a fifth of the critical damage at 450 h (the path is at 0.18) has only half of its life left, and the last half of the damage accumulates in the last 9 % of the life. An inspection program designed under the linear law would inspect too rarely for an accelerating mechanism.

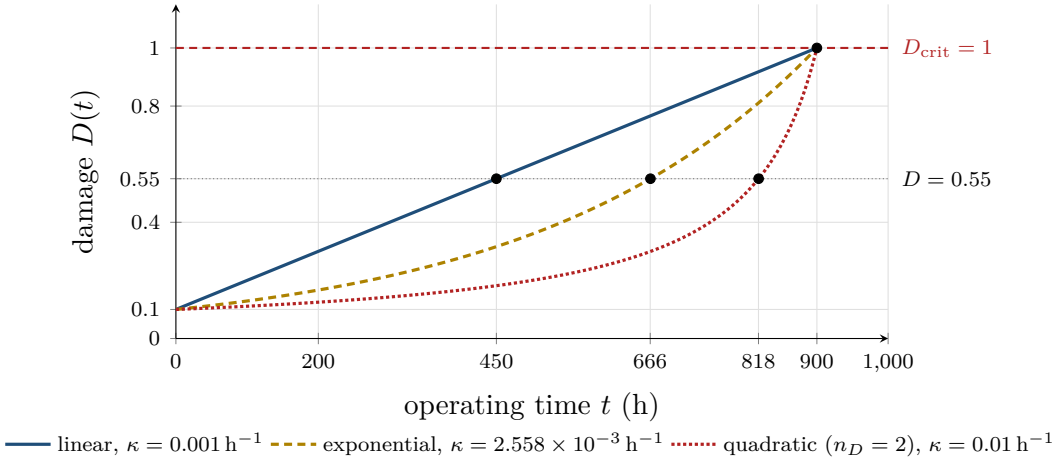


Figure 21.2: Linear, exponential and quadratic degradation laws, $D_0 + \kappa t$, $D_0 \mathrm{e}^{\kappa t}$ and $[D_0^{-1} - \kappa t]^{-1}$, with $D_0 = 0.10$ and rates chosen so that all three reach $D_{\mathrm{crit}} = 1$ at 900 h (illustrative). The marked points are the times at which each path reaches half the distance to the threshold: 450, 666 and 818 h.

Whether an accelerating law is physically appropriate depends on the mechanism. Uniform corrosion of a pipe wall at a steady water chemistry proceeds at a roughly constant rate; a fatigue crack accelerates; wear can do either, depending on whether the worn surface degrades the lubrication. The choice of g_D is a statement about the mechanism and has to be justified as such.

21.2.5 Damage that depends on operating conditions

A constant degradation rate may be reasonable under stable operating conditions, but industrial systems experience changing load and environment. A more useful model makes the rate a function of the conditions,

$$\frac{\mathrm{d}D}{\mathrm{d}t} = g_D\big(\sigma(t), T(t), I(t), \ldots\big), \tag{21.12}$$

where $\sigma(t)$ is the mechanical stress, $T(t)$ the absolute temperature and $I(t)$ the electric current. (In this chapter T with an argument or a subscript other than f is a temperature in kelvin; the failure time is always T_{f}.) Writing $\mathbf{z}(u)$ for the vector of conditions at time u and integrating from the initial state, the accumulated damage is

$$D(t) = D_0 + \int_0^t g_D\big(\mathbf{z}(u)\big)\,\mathrm{d}u. \tag{21.13}$$

The damage at time t depends on the whole history of the conditions, not on the elapsed time alone. Two nominally identical components can therefore age differently. One cooling pump may operate near its maximum load continuously while another runs only intermittently; their chronological ages are equal, but their accumulated physical damage can differ by a large factor. Section 21.6.1 returns to this point with a quantitative example.

Chapter 5 of Volume I reached a similar conclusion for the hazard function. There, Section 5.6.3 wrote a condition-dependent hazard $\lambda(t \mid \mathbf{z}(t)) = \lambda_0(t) \exp[\boldsymbol{\gamma}^{\mathsf{T}} \mathbf{z}(t)]$ and obtained the reliability by integrating it along the history of the conditions, equation (5.22). The two models

answer different questions. A condition-dependent hazard changes the probability of a failure event without describing the state of the component; a condition-dependent degradation law changes the state, and the probability of failure follows from where the state is relative to its threshold. The second can use a measurement of the state; the first cannot.

21.3 Physics-based damage models: fatigue, crack growth and thermal aging

Three mechanisms have physical damage laws developed far enough to predict failure from the loading history: fatigue, through the stress-life curve and the Miner rule; fatigue crack growth, through the Paris law; and thermal aging, through the Arrhenius rate. This section restates each from Chapter 4 of Volume I, applies it to a new example, and shows how each becomes a degradation state that can be updated as the operating history is recorded.

21.3.1 Fatigue as damage driven by load cycles

Fatigue damage arises from repeated loading, and it differs from ordinary wear in two respects. A metal component can fail under cyclic stresses that are each below its static strength: a transmission tower connection may withstand a particular wind-induced load once, but repeated cycles of the same load can initiate a crack and grow it. And the relevant exposure is often not elapsed time but the number of loading cycles N and their amplitudes, so that a fatigue model needs information about the amplitude and the number of stress cycles, not only about the operating hours.

An S-N curve relates the stress amplitude to the fatigue life. One common empirical relation is Basquin's law,

$$\sigma_a = \sigma'_f \, (2N_f)^{b_{\mathrm{B}}}, \tag{21.14}$$

where σ_a is the stress amplitude (half the stress range of a cycle), σ'_f the fatigue strength coefficient in the units of stress, N_f the number of cycles to failure, and $b_{\mathrm{B}} < 0$ the fatigue strength exponent (written b in fatigue texts; the subscript separates it from the S-N exponent b below). The factor $2N_f$ is the number of stress reversals, two per cycle; some formulations use cycles directly and absorb the factor into the fitted coefficient. Solving equation (21.14) for the life, $2N_f = (\sigma_a/\sigma'_f)^{1/b_{\mathrm{B}}}$, gives

$$N_f = \tfrac{1}{2}\, \sigma_f'^{\,b}\, \sigma_a^{-b}, \qquad b = -\frac{1}{b_{\mathrm{B}}} > 0,$$

which is the simpler engineering form used in Section 4.4.1,

$$N(\Delta\sigma) = K\, \Delta\sigma^{-b}, \tag{21.15}$$

with $K > 0$ and $b > 0$, written with the stress range $\Delta\sigma = 2\sigma_a$ of a fully reversed cycle and the factor 2^b absorbed into K. Basquin exponents of smooth metal specimens are typically between -0.05 and -0.12, which gives b between about 8 and 20; welded details, whose life is mostly crack growth, have b near 3, as Section 4.4.2 showed. The important result is that a higher cyclic stress produces fewer cycles to failure within the regime in which the curve applies.

21.3.2 The Miner rule and a cyclically loaded connection

Real components experience many different stress amplitudes. The Miner rule of Section 4.4.1, the standard linear damage accumulation rule, assigns to each cycle at stress level i the fraction

$1/N_i$ of the life at that level, where N_i is the constant-amplitude fatigue life. After n_i cycles at each of the levels, the accumulated damage is

$$D = \sum_i \frac{n_i}{N_i}, \tag{21.16}$$

and nominal fatigue failure is assumed when D reaches $D_{\text{crit}} = 1$. Each fraction n_i/N_i is the fraction of the nominal fatigue life consumed at load level i, and the sum is the degradation state of a fatigue model.

Example 21.2 (A cyclically loaded structural connection)**.** A structural connection has experienced, over its service so far, the stress cycles of Table 21.2, counted in three classes of stress range. For each class, testing or an appropriate design fatigue curve gives the number of cycles to failure at constant amplitude (illustrative values). The table reads row by row: the second column is the number of cycles n_i experienced in the class, the third the constant-amplitude life N_i, and the fourth the damage contribution n_i/N_i. The accumulated damage is

$$D = \frac{100\,000}{1\,000\,000} + \frac{40\,000}{200\,000} + \frac{10\,000}{50\,000} = 0.100 + 0.200 + 0.200 = 0.50.$$

Under the nominal Miner model, the connection has consumed 50 % of its critical damage. If the future load spectrum repeats the historical spectrum, another identical block of cycles adds $\Delta D = 0.50$, and the nominal prediction is failure at the end of one additional block.

Table 21.2: Miner damage of a cyclically loaded structural connection (illustrative values).

Stress range	Cycles experienced n_i	Cycles to failure N_i	Damage n_i/N_i
50 MPa	100 000	1 000 000	0.100
75 MPa	40 000	200 000	0.200
100 MPa	10 000	50 000	0.200
Total			0.500

The three lives of the table do not lie on one straight line of equation (21.15): between 50 and 75 MPa they imply an exponent $b = \ln 5/\ln 1.5 = 3.97$, and between 75 and 100 MPa an exponent $\ln 4/\ln(4/3) = 4.82$. A design fatigue curve may legitimately bend in this way; the Miner sum needs only the life at each level, not a single power law. Unlike the welded joint of Section 4.4.1, where the many small cycles dominated the damage, here the two higher classes contribute equally and together 80 % of it.

The rule has an important limitation: it does not account for the order of the loading cycles. High-amplitude loading early in life can affect crack initiation and growth differently from the same loading late in life, and the rule does not represent residual stresses, crack closure, corrosion fatigue or an uncertain threshold. The value $D = 0.50$ should therefore not be read as an experimentally verified statement that 50 % of the useful life remains. It is a model-dependent damage index, and measured damage sums at failure scatter widely around one.

21.3.3 Fatigue under time-varying operating conditions

In practice the stress cycles are not given in classes; they are extracted from a recorded or simulated stress history by a cycle-counting method, of which rainflow counting is the standard one. Rainflow counting pairs each local stress reversal with the reversal that closes a hysteresis loop, so that an irregular history becomes a list of complete cycles, each with a range and a mean. The damage calculation is then the chain

$$\text{stress history } \sigma(t) \;\rightarrow\; \text{stress cycle identification} \;\rightarrow\; \{n_i, \Delta\sigma_i\} \;\rightarrow\; D.$$

Figure 21.3 draws the chain for a wind-exposed structure such as a transmission tower. The boxes in the main row are the stages of the calculation, from left to right: the stress history at the critical detail, the cycle counting, the histogram of cycles by stress range, the S-N curve with the Miner sum, and the damage with the remaining life it implies. The note above the first box lists the inputs from which the stress history is computed: the mean wind speed, the turbulence intensity, the wind direction, the natural frequencies of the structure and its aerodynamic response. The note above the fourth box is the material fatigue property, the S-N curve of the detail. The dashed arrow below the row is the update: each new record of stress, or of wind from which stress is computed, adds its increment of damage to the state.

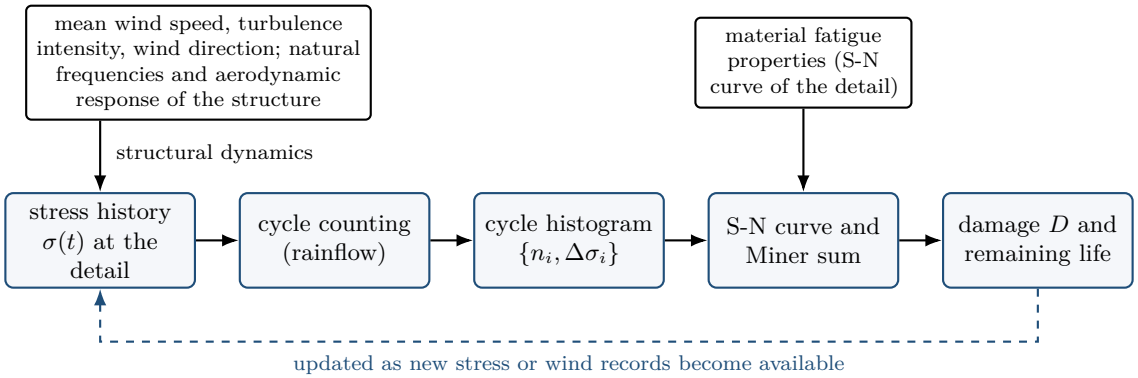


Figure 21.3: Fatigue damage of a wind-exposed structure computed from its operating history: the stress history at the critical detail is reduced to counted cycles, and the cycles to a Miner damage, which is updated with every new record.

Because the damage is a sum, it is updated by adding the increment of each new record. For the connection of Example 21.2, suppose that the next block of service, a windier period than the first, brings 30 % more cycles in every class (illustrative). Its damage increment is $1.3 \times 0.50 = 0.65$ instead of 0.50, and the threshold $D = 1$ is reached when $0.50/0.65 = 77\,\%$ of that block has elapsed, not at its end. This is a direct connection between meteorological modeling, structural dynamics and reliability analysis: the weather that sets the dynamic ratings of a transmission line also sets the fatigue damage of its structures.

21.3.4 Crack growth and the cycles to a critical length

The Miner rule summarizes damage as an accumulated scalar. When a crack is already present, its growth can be modeled directly. Let $a(N)$ be the crack length after N cycles. The Paris law of Section 4.4.2 states that the growth per cycle is a power of the stress-intensity-factor range,

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C_{\mathrm{P}}\,(\Delta K)^m, \qquad \Delta K = Y\,\Delta\sigma\,\sqrt{\pi a}, \tag{21.17}$$

where C_{P} and m are material and environmental parameters determined by experiment, ΔK is the stress-intensity-factor range in $\mathrm{MPa}\sqrt{\mathrm{m}}$, Y a dimensionless geometry correction factor and $\Delta\sigma$ the stress range. Substituting, $\mathrm{d}a/\mathrm{d}N = C_{\mathrm{P}}\,(Y\Delta\sigma\sqrt{\pi a})^m$: the growth rate depends on the current crack length, and as the crack grows the stress-intensity factor increases and the crack accelerates. The law is a physical explanation of nonlinear degradation.

Suppose that Y and $\Delta\sigma$ remain constant over the interval analyzed, and define

$$K_0 = Y\,\Delta\sigma\,\sqrt{\pi}, \tag{21.18}$$

in MPa. Then $\Delta K = K_0\sqrt{a}$ and

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C_{\mathrm{P}}\, K_0^m\, a^{m/2},$$

which is the power law of equation (21.10) with the crack length as the state, cycles in place of time and $n_D = m/2$. Rearranging, $\mathrm{d}N = \mathrm{d}a/(C_{\mathrm{P}} K_0^m a^{m/2})$, and integrating from the present crack length a_0 to the critical length a_c gives the remaining number of cycles,

$$N_{\mathrm{rem}} = \int_{a_0}^{a_c} \frac{\mathrm{d}a}{C_{\mathrm{P}}\,(K_0\sqrt{a})^m}. \tag{21.19}$$

For $m \neq 2$ the antiderivative of $a^{-m/2}$ is $a^{1-m/2}/(1-m/2)$, so

$$N_{\mathrm{rem}} = \frac{a_c^{\,1-m/2} - a_0^{\,1-m/2}}{C_{\mathrm{P}}\, K_0^m\,(1-m/2)}, \qquad m \neq 2, \tag{21.20}$$

which is equation (4.20) of Chapter 4 with numerator and denominator multiplied by -1. For $m = 2$ the integrand is $1/(C_{\mathrm{P}} K_0^2 a)$ and

$$N_{\mathrm{rem}} = \frac{\ln(a_c/a_0)}{C_{\mathrm{P}}\, K_0^2}, \qquad m = 2. \tag{21.21}$$

The expression connects the current crack length, the applied cyclic stress and the remaining fatigue life. With the illustrative steel parameters of Section 4.4.2, $C_{\mathrm{P}} = 10^{-11}$ meters per cycle per $(\mathrm{MPa}\sqrt{\mathrm{m}})^3$, $m = 3$, $Y = 1.12$, $\Delta\sigma = 100\,\mathrm{MPa}$ and $a_c = 10\,\mathrm{mm}$,

$$K_0 = 1.12 \times 100 \times 1.7725 = 198.5\,\mathrm{MPa}, \qquad C_{\mathrm{P}} K_0^3 = 7.823 \times 10^{-5}\,\mathrm{m}^{-1/2} \text{ per cycle},$$

and with $1 - m/2 = -\tfrac{1}{2}$,

$$N_{\mathrm{rem}}(a_0 = 1\,\mathrm{mm}) = \frac{(0.010)^{-1/2} - (0.001)^{-1/2}}{7.823 \times 10^{-5} \times (-0.5)} = \frac{10.00 - 31.62}{-3.912 \times 10^{-5}} = 5.53 \times 10^5 \text{ cycles},$$

the value of Chapter 4.

21.3.5 Crack length as a predictor of remaining life

The crack-growth model can outperform a model based on age alone. Two identical components of equal calendar age may have different crack lengths, for instance $a_0 = 1\,\mathrm{mm}$ and $a_0 = 4\,\mathrm{mm}$. For the second, $(0.004)^{-1/2} = 15.81\,\mathrm{m}^{-1/2}$ and

$$N_{\mathrm{rem}}(a_0 = 4\,\mathrm{mm}) = \frac{15.81 - 10.00}{0.5 \times 7.823 \times 10^{-5}} = 1.49 \times 10^5 \text{ cycles},$$

3.72 times fewer than for the first, although an age-only model gives both the same reliability. A degradation model recognizes the distinction because its state carries the history that the age discards.

Figure 21.4 shows the remaining life as a function of the present crack length. The horizontal axis is the crack length a_0 found at an inspection, in millimeters, and the vertical axis, on a logarithmic scale, the remaining number of cycles to the critical length of 10 mm, in units of 10^5, from equation (21.20). The three curves are three stress ranges, and the two marked points are the components of the example. Each curve falls by more than two orders of magnitude between 0.5 and 9.5 mm, and in absolute terms it falls fastest at small crack lengths: a crack of 2 mm at 100 MPa leaves 3.16×10^5 cycles, 57 % of the life of a 1 mm crack. The curves for different stress ranges are parallel because the remaining life scales as $\Delta\sigma^{-m} = \Delta\sigma^{-3}$ at every

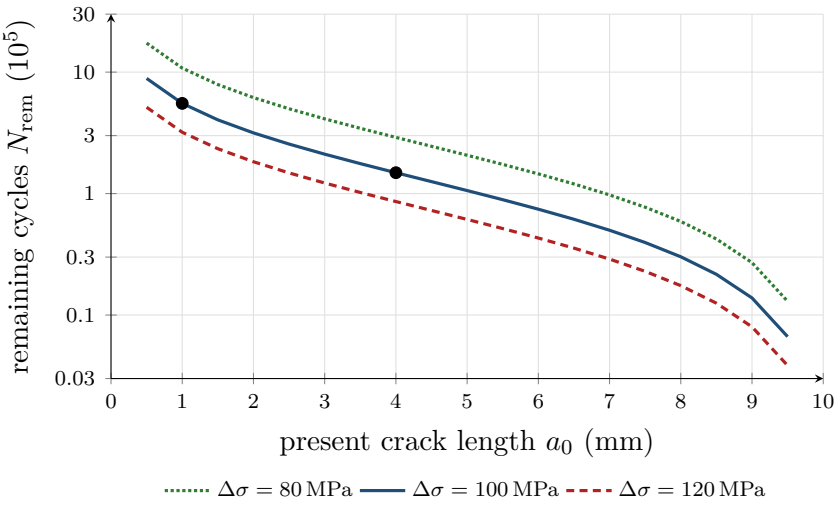


Figure 21.4: Remaining fatigue life against the present crack length, from the Paris law, equation (21.20), with illustrative steel parameters $C_{\mathrm{P}} = 10^{-11}$, $m = 3$, $Y = 1.12$ and a critical length $a_c = 10\,\mathrm{mm}$. Marked: 5.53×10^5 cycles for $a_0 = 1\,\mathrm{mm}$ and 1.49×10^5 cycles for $a_0 = 4\,\mathrm{mm}$ at 100 MPa.

crack length. A measurement of the crack length therefore moves the prediction more than any refinement of the age distribution could.

The Paris law applies only within its crack-growth regime. Near the threshold stress-intensity range below which cracks do not propagate, and near unstable fracture as the crack approaches its critical size, other fracture-mechanics relations are needed, and the critical length itself depends on the fracture toughness of the material and the peak load.

21.3.6 Thermal aging and the Arrhenius acceleration factor

Many insulation materials and polymers age faster at higher temperatures. Section 4.3.4 wrote the rate of a thermally activated degradation process as $k(T) = k_0 \exp[-E_a/(k_B T)]$, with the activation energy E_a in electronvolts per molecule and the Boltzmann constant k_B. The same phenomenological law is often written per mole,

$$k(T) = k_0 \exp\!\left(-\frac{E_a}{R_{\mathrm{g}} T}\right), \tag{21.22}$$

where $k(T)$ is the degradation reaction rate, k_0 the pre-exponential factor, E_a the activation energy in J/mol, $R_{\mathrm{g}} = 8.314\,\mathrm{J/(mol\,K)}$ the molar gas constant (neither the reliability function R nor a resistance), and T the absolute temperature. The temperature must be in kelvin. The two forms are the same law, since E_a/R_{g} with E_a per mole equals E_a/k_B with E_a per molecule.

For two temperatures T_1 and T_2 the pre-exponential factor cancels in the ratio of the rates,

$$\frac{k(T_2)}{k(T_1)} = \frac{k_0\, \mathrm{e}^{-E_a/(R_{\mathrm{g}} T_2)}}{k_0\, \mathrm{e}^{-E_a/(R_{\mathrm{g}} T_1)}},$$

so the acceleration factor of T_2 relative to T_1 is

$$\mathrm{AF} = \exp\!\left[\frac{E_a}{R_{\mathrm{g}}}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)\right]. \tag{21.23}$$

If $T_2 > T_1$, then $1/T_1 > 1/T_2$ and the acceleration factor exceeds one for a positive activation energy.

END OF THE PREVIEW

Reliability Engineering Volume II

These are the opening pages. The complete book runs 479 pages.

The full PDF opens with a token. Write to **rmarquez123@gmail.com**, name the book, and I will send you one.

One token opens every book on the site.

Ricardo Marquez, Ph.D.

ricardo-marquez.com