UMass | Dartmouth UNIVERSITY OF MASSACHUSETTS DARTMOUTH

Maximum Likelihood Estimation of a Non-homogeneous Poisson Process Software Reliability Model with the Expectation Conditional Maximization Algorithm

- Vidhyashree Nagaraju, BE, *University of Massachusetts Dartmouth*
- Thierry Wandji, PhD, *Naval Air Systems Command*
- Lance Fiondella, PhD, *University of Massachusetts Dartmouth*

Motivation

- Department of Defense (DoD) increasingly depends on software intensive systems
  - Mission and life critical
  - Must preserve high reliability and availability
- Urgency to deploy new technologies and military capabilities may result in
  - Inadequate reliability testing
  - Severe economic damage and loss of life

Background

- Recent National Academies report on Enhancing Defense System Reliability recommends
  - Use of reliability growth models to direct contractor design and test activities
- Several tools to
  - Automatically apply reliability models
  - Automate reliability test and evaluation

Background (2)

- Software reliability tools over two decades old
  - Difficult to configure on modern operating systems
- Developing open source software reliability tool (SRT) for
  - Naval Air Systems Command (NAVAIR)
  - Department of Defense
  - Broader software engineering community
- **Technical challenge**: Stability of underlying model fitting algorithms

Present work

- To improve robustness of model fitting process
  - Developing algorithms to compute maximum likelihood estimates (MLE) of software reliability growth models (SRGM)
    - Expectation maximization (EM)
    - Expectation conditional maximization (ECM)
- Our contribution: Implicit ECM algorithm
  - Eliminates computationally intensive integration from update rules of ECM
  - Guarantees dimensionality reduction
  - Speedup 200-400 times explicit EM and ECM algorithms

Nonhomogeneous Poisson process (NHPP) SRGM

- Stochastic process counts number of events observed as function of time
  - In context of software reliability, NHPP counts number of faults detected by time t
- Counting process characterized by mean value function (MVF)
  - Form of MVF of several SRGM: $m(t) = a \times F(t)$
    - a - number software that would be detected with indefinite testing
    - $F(t)$ - cumulative distribution function (CDF)

Weibull SRGM

- Substituting Weibull distribution for $F(t)$

$$m(t) = a\left(1 - e^{-bt^{c}}\right)$$

  - b - scale parameter
  - c - shape parameters
- $c = 1$ simplifies to exponential distribution
  - Also known as Goel-Okumoto SRGM

Maximum likelihood estimation

- Procedure to identify numerical values of model parameters that best fit observed failure data
- Two common types of failure data
  - **Failure times**: vector of individual failure times

    $\mathrm{T} = < t_1, t_2, \dots, t_n >$

    - n – number of faults observed
  - **Failure counts**: $\mathrm{T} = \{(t_1, k_1), (t_2, k_2), \dots, (t_n, k_n)\}$
    - t_i – time at which ith interval ends
    - k_i – number of faults detected in interval k

Maximum likelihood estimation (2)

- NHPP failure times data log-likelihood

$$LL(\mathrm{T}|\Theta) = -m(t_n) + \sum_{i=1}^{n} \log\big(\lambda(t_i)\big)$$

  - Θ - vector of model parameters
  - $\lambda(t) \coloneqq \frac{dm(t)}{dt}$ - instantaneous failure rate at time t_i
- Traditional approach solves simultaneous system of equations with Newton's Method

$$\frac{\partial}{\partial\Theta} LL(\mathrm{T}|\Theta) = \mathbf{0}$$

Expectation maximization (EM)

- EM algorithm also maximizes log-likelihood
- Unlike maximum likelihood estimation
  - EM algorithm maximizes with respect to complete data
    - Observed and unobserved data
- E-step: Function for expectation of log-likelihood evaluated using current parameter estimates
- M-step: Computes parameters maximizing expected log-likelihood found in E-step.
- Monotonically improves log-likelihood in each iteration

Expectation condition maximization (ECM)

- Preserves monotonicity property of EM algorithm
- Replaces M-step of EM algorithm with p conditional maximization (CM)-steps
  - p - number of model parameters
- Divides single p −dimensional problem into

  p 1-dimensional problems

Main steps of ECM algorithm

1. Specify log-likelihood of NHPP SRGM
2. For each parameter, differentiate log-likelihood function to obtain conditional maximum (CM)-steps
   - Solve for closed form expression if possible
     - Otherwise systems of implicit algebraic expressions results

3. Holding $(p-1)$ parameters constant, cycle through p CM-steps until improvement in log-likelihood is small
   - Must cycle through subsets of simultaneous CM-steps if close form expressions not obtainable

4. Return MLE of parameter vector $\widehat{\Theta}^{(t)}$

Explicit Weibull ECM

- $a'' = n + a' e^{-b' t_n^{c'}}$

- $b'' = \dfrac{\left(n + a' e^{-b' t_n^{c'}}\right) b'}{\sum_{i=1}^{n} t_i \left(\dfrac{\Gamma\left(b' t_i^{c'};2\right) - \Gamma\left(b' t_{i-1}^{c'};2\right)}{e^{-b' t_{i-1}^{c'}} - e^{-b' t_i^{c'}}} \right) - a' \overline{\Gamma}(\mathrm{b}' t_n^{c'};2)}$

- $c'' = \dfrac{n + a' e^{-b' t_n^{c''}}}{\sum_{i=1}^{n} t_i \dfrac{\int_{t_{i-1}}^{t_i} \left(b'' u^{c''} - 1\right) \ln(u) f(u; b', c') du}{e^{-b' t_{i-1}^{c'}} - e^{-b' t_i^{c'}}} + a' \int_{t_k}^{\infty} (b'' u^{c''} - 1) \ln(u) f(u; b', c') du}$

May require solution of system of subset of equations

Implicit ECM (IECM) algorithm

- Specify log-likelihood function of failure times NHPP SRGM by substituting $m(t)$ into and simplifying

$$LL(\mathrm{T}|\Theta) = -m(t_n) + \sum_{i=1}^{n} \log\big(\lambda(t_i)\big)$$

- Reduce log-likelihood function from p to $(p-1)$ parameters by computing $\frac{\partial LL}{\partial a} = 0$ and solve for $\hat{a}$
  - When MVF possesses form $a \times F(t)$

$$\hat{a} = \frac{n}{F(t_n)}$$

- Substitute $\hat{a}$ into LL to obtain reduced log-likelihood (RLL)

Implicit ECM algorithm (2)

- Derive conditional maximum (CM)-step for remaining $(p-1)$ parameters by
$$\frac{\partial RLL}{\partial \Theta_i} = 0, \qquad (1 \leq i \leq p-1)$$
- No closed form expression sought
  - Implicit ECM guarantees dimensionality reduction
- Algorithm cycles through $(p-1)$ CM-steps
  - Holds other $(p-2)$ parameters constant at present estimate and applies numerical root finding algorithm to update

Implicit ECM algorithm (3)

- Cycle repeats until convergence

$$\left|RLL_j - RLL_{j-1}\right| < \varepsilon$$

- ε - small positive constant

- Identifies MLEs $\widehat{\Theta}/a$
- Substitute values of $\widehat{\Theta}/a$ into equation for $\hat{a}$
  - Produces MLE $\widehat{\Theta}$ for all p model parameters

Weibull SRGM – IECM

- Instantaneous failure rate

$$\lambda(t) = abct^{c-1}e^{-bt^c}$$

- Log-likelihood function

$$LL(\mathrm{T};\Theta) = -a\left(1 - e^{-bt_n^c}\right) + \sum_{i=1}^{n} \log\left[abct_i^{c-1}e^{-bt_i^c}\right]$$

- MLE of parameter a

$$\hat{a} = \frac{n}{1 - e^{-bt_n^c}}$$

Weibull SRGM – IECM (2)

- Substituting for $\hat{a}$ in log-likelihood

$$RLL(\mathrm{T};\Theta) = -n + \sum_{i=1}^{n} \log\left[\frac{nbct_i^{c-1}e^{-bt_i^c}}{1-e^{-bt_n^c}}\right]$$

- Differentiating w.r.t b produces CM-step for b

$$b'' = \frac{\left(-nt_n^{c'}e^{-b''t_n^{c'}} + \left(\frac{n}{b''} - \sum_{i=1}^{n} t_i^{c'}\right)\left(1-e^{-b''t_n^{c'}}\right)\right)}{1-e^{-b''t_n^{c'}}}$$

- Differentiating w.r.t c produces CM-step for c

$$c'' = \frac{\left(-nt_n^{c''}b'\log[t_n]e^{-b't_n^{c''}}\right)}{1-e^{-b't_n^{c''}}} + \frac{n}{c''} + \sum_{i=1}^{n}\log[t_i] - b'\sum_{i=1}^{n} t_i^{c''}\log[t_i]$$

Initial parameter estimates selection

- By the first order optimality condition,
  - Initial estimates of parameters of the function $F(t;\Theta)$

$$a^{(0)} = n$$

and

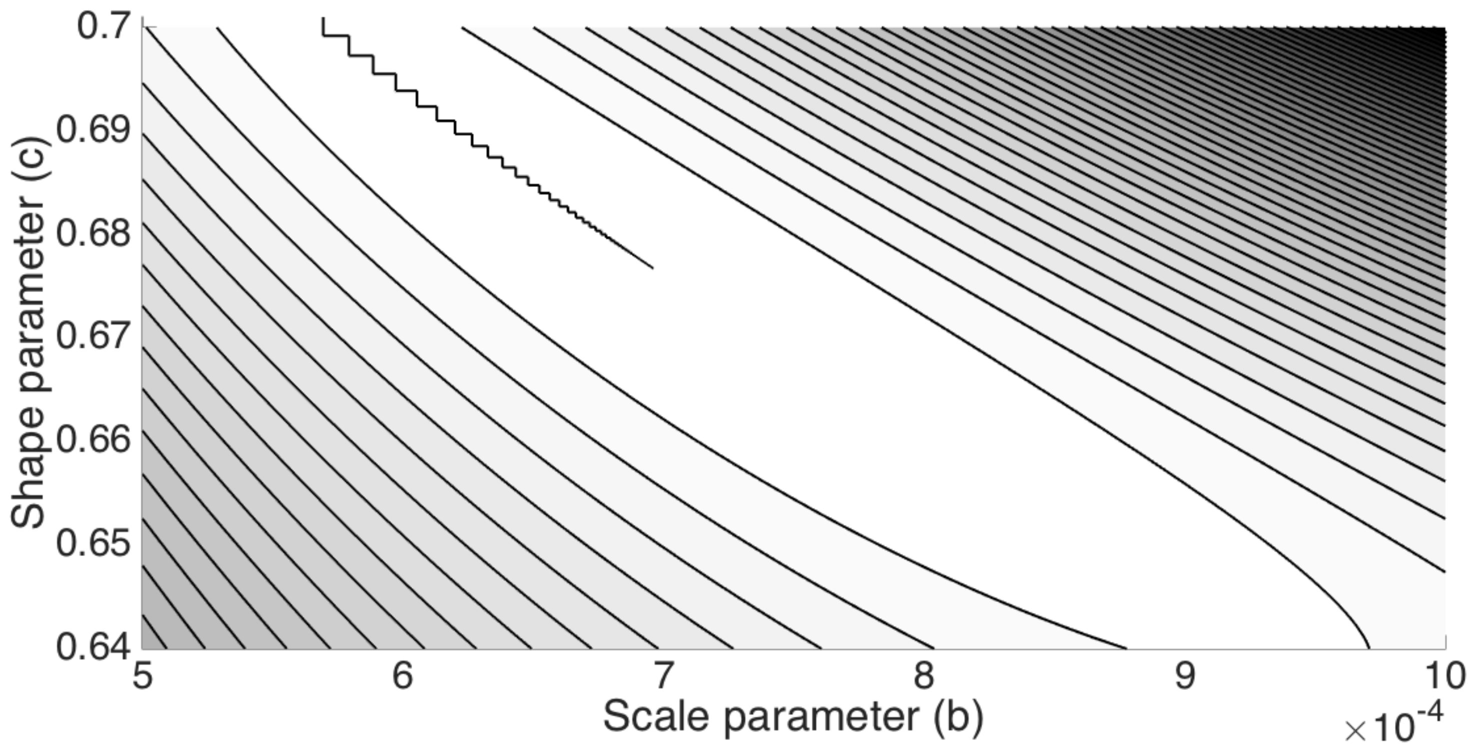
$$\Theta^{(0)} := \sum_{i=1}^{n} \frac{\partial}{\partial \Theta} \log[f(t_i;\Theta)] = \mathbf{0}$$

Illustration

- SYS1 dataset
  - $n = 136$ failure times
- Initial estimates
  - $c^{(0)} = 1$ – simplifies to exponential SRGM
    - Provide feasible initial solution
  - $b^{(0)} = \frac{n}{\sum_{i=1}^{n} t_i^c}$ (initial EM estimate when $c = 1$)

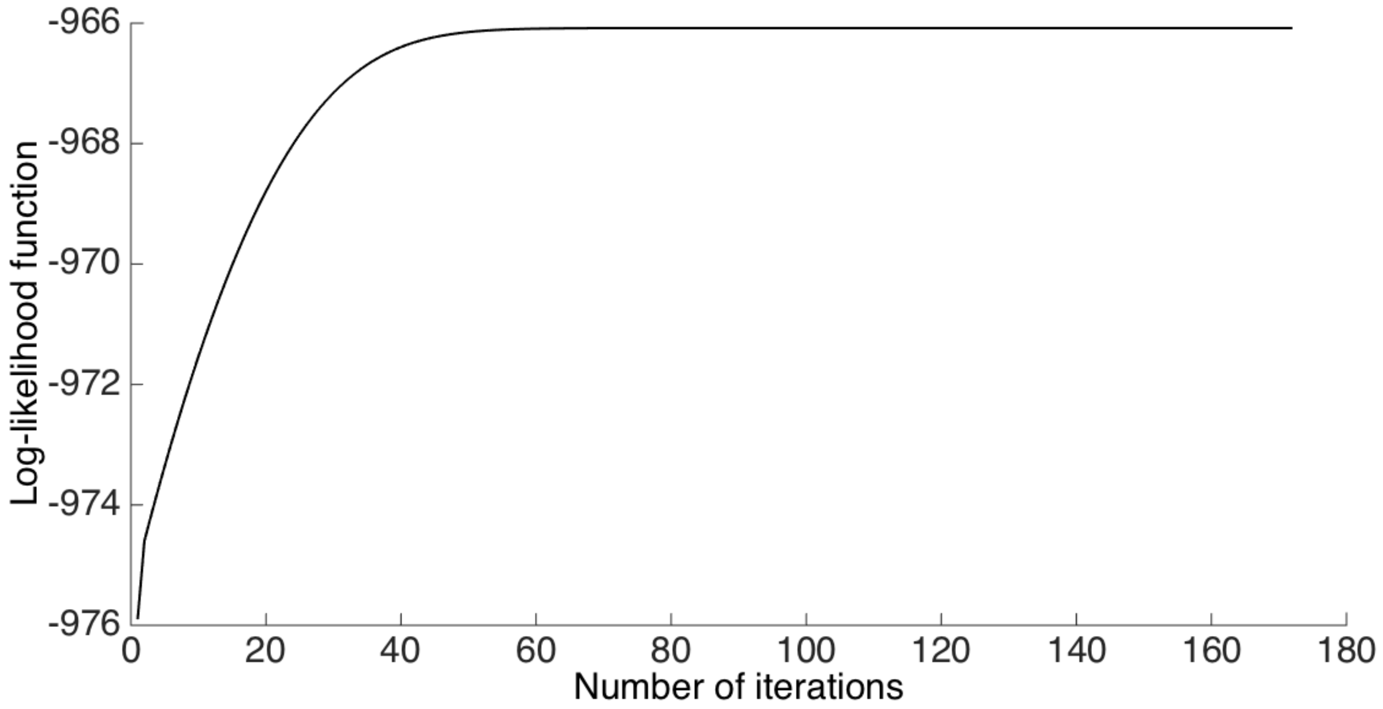
UMass | Dartmouth UNIVERSITY OF MASSACHUSETTS DARTMOUTH

Iterations of IECM of Weibull



Monotonic improvements made in each of 172 iterations

Log-likelihood function



IECM iterations improve log-likelihood function monotonically

Performance analysis

Initial value factor (ρ)	EM	ECM	IECM	$\frac{ECM}{IECM}$	$\frac{EM}{IECM}$
0.25	2.695	2.310	0.012	197.434	230.342
0.50	2.232	1.994	0.010	196.645	220.116
0.75	1.808	1.840	0.005	393.151	386.322
0.90	1.764	1.689	0.005	309.338	323.075
1.25	1.826	1.780	0.006	285.255	292.626
1.50	1.888	1.761	0.005	322.525	345.785
1.75	1.965	1.966	0.006	315.062	314.902
2.0	2.123	1.983	0.006	317.786	340.222

IECM algorithm 200 to 400 times faster than EM and ECM

With initial estimates, IECM<0.015 seconds, ECM>5 minutes

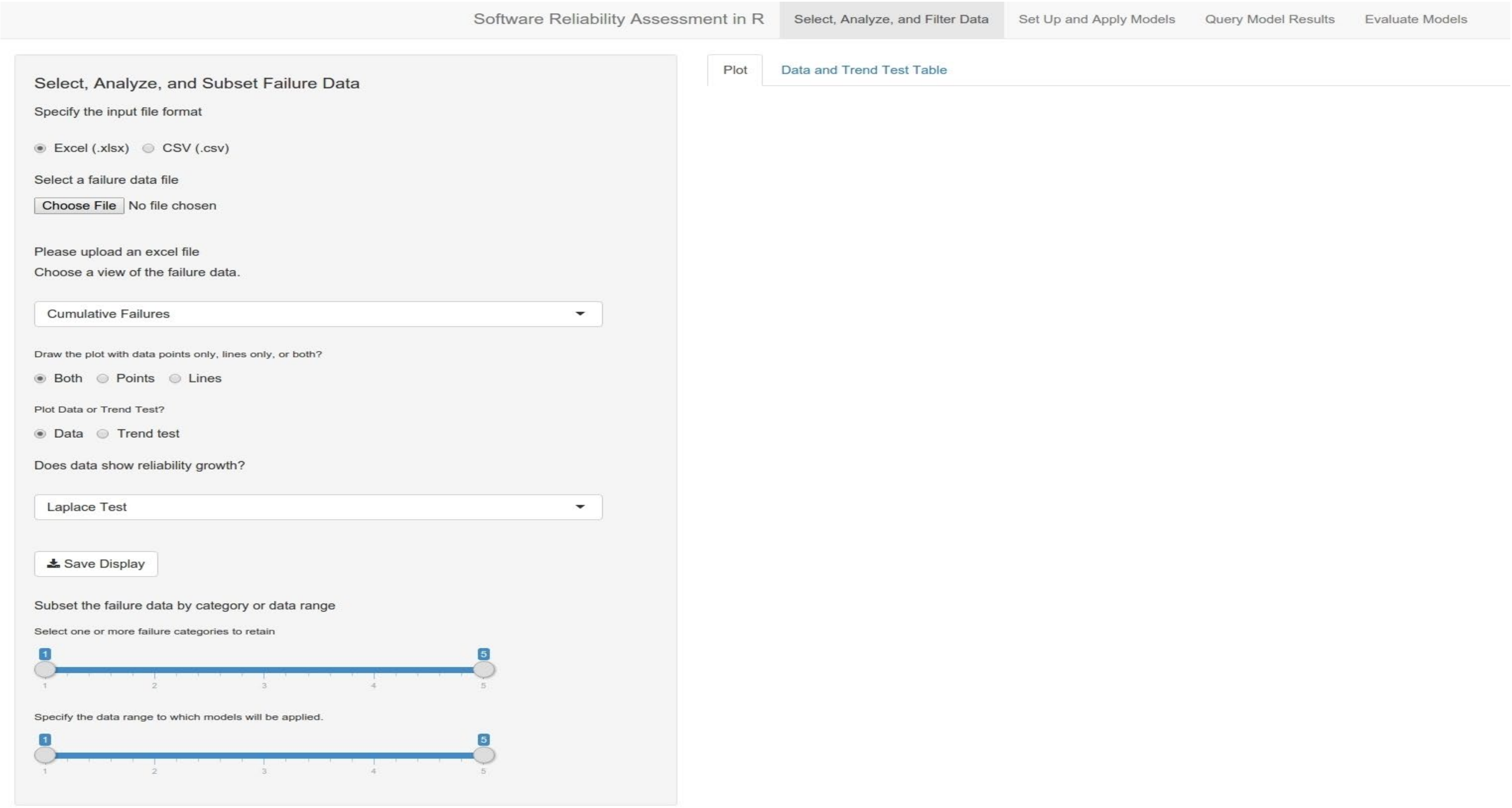
Summary and conclusions

- Proposed approach to accelerate explicit EM and ECM algorithms in context of NHPP SRGM
- Results indicate IECM
  - Avoids computationally expensive gamma function and numerical integration present in explicit CM-steps
  - Guarantees dimensionality reduction
  - Performs two orders of magnitude faster than explicit approaches

Future work

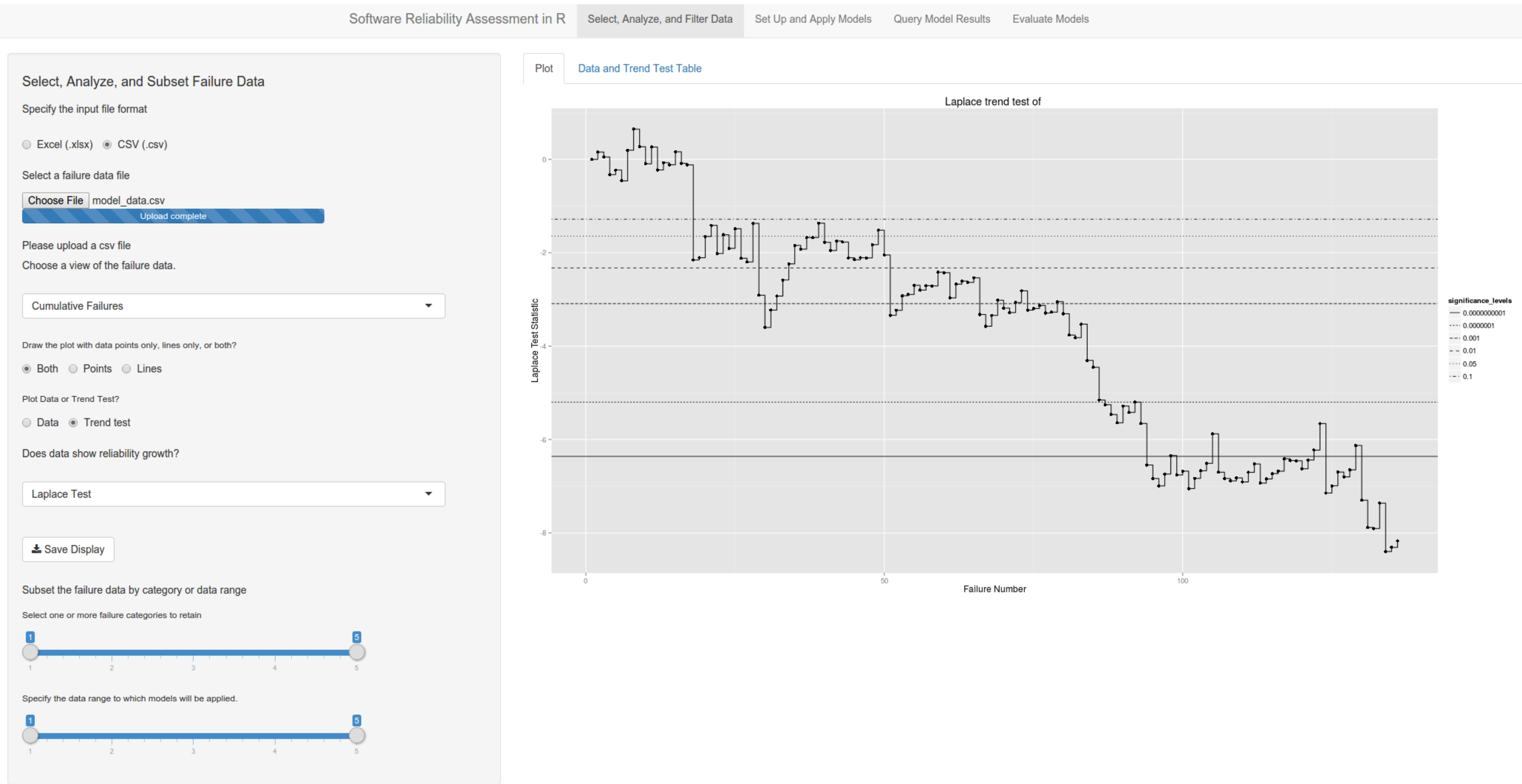
- Develop multi-phase algorithms that employ a combination of algorithms to enhance stability and performance
  - Assess tradeoff between stability and performance of algorithms
- Incorporate IECM algorithms into Software Reliability Tool (SRT)

Tab view


Software Reliability Assessment in R
Select, Analyze, and Filter Data
Set Up and Apply Models
Query Model Results
Evaluate Models
Select, Analyze, and Subset Failure Data
Specify the input file format
Excel (.xlsx)
CSV (.csv)
Select a failure data file
Choose File
No file chosen
Please upload an excel file
Choose a view of the failure data.
Cumulative Failures
Draw the plot with data points only, lines only, or both?
Both
Points
Lines
Plot Data or Trend Test?
Data
Trend test
Does data show reliability growth?
Laplace Test
Save Display
Subset the failure data by category or data range
Select one or more failure categories to retain
Specify the data range to which models will be applied.
Plot
Data and Trend Test Table

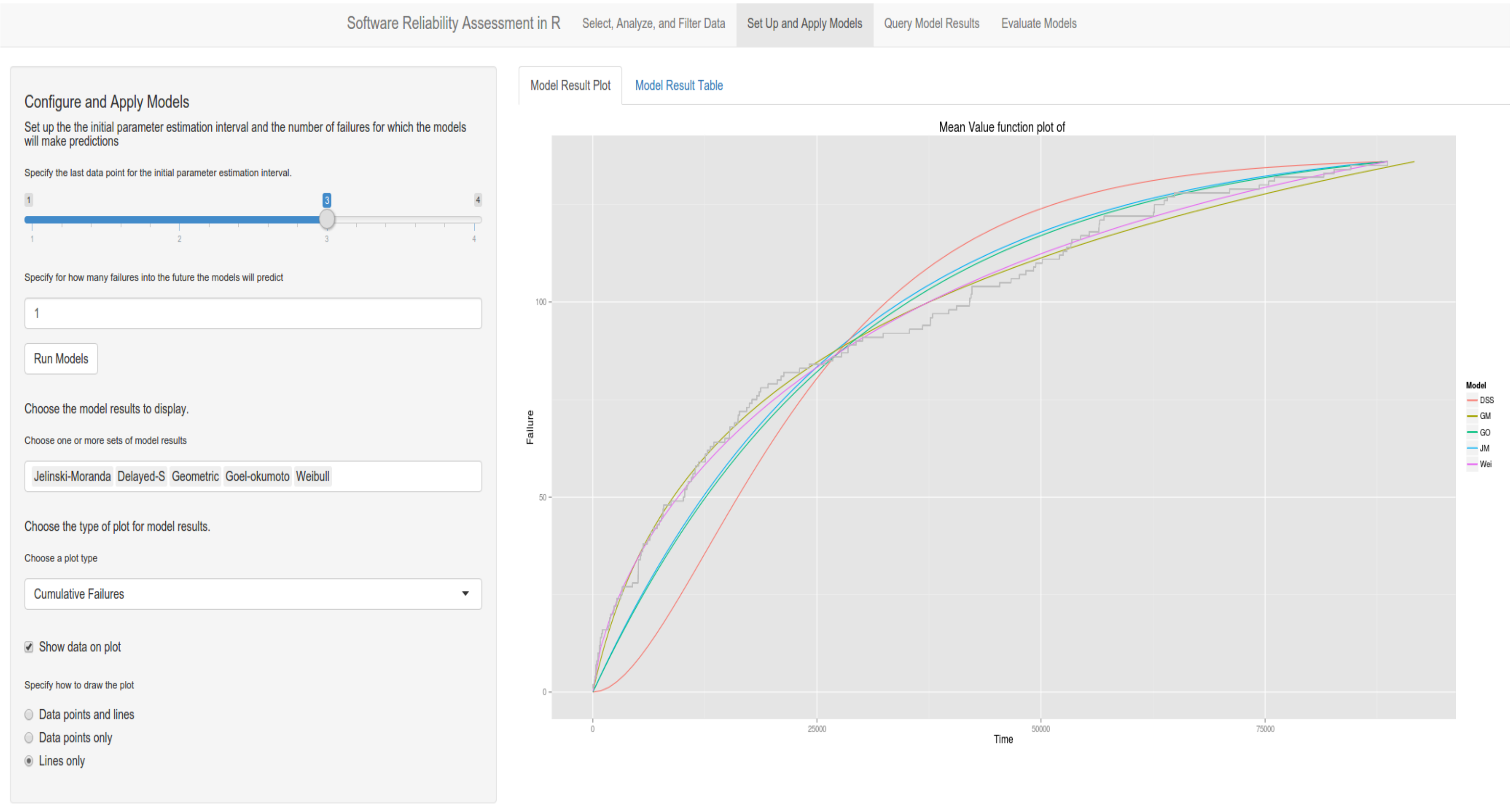
UMass | Dartmouth UNIVERSITY OF MASSACHUSETTS DARTMOUTH

Laplace trend test

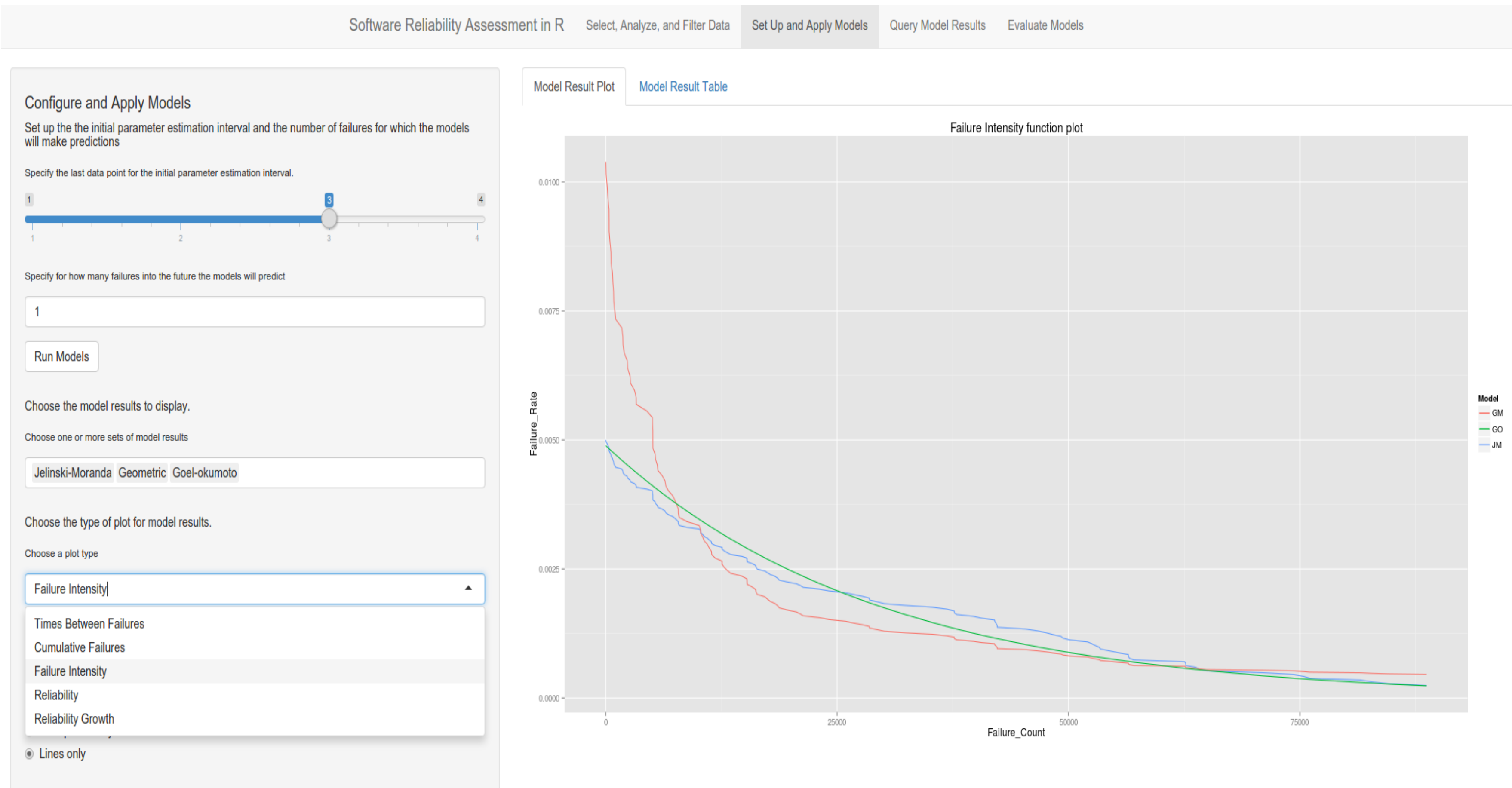


UMass Dartmouth UNIVERSITY OF MASSACHUSETTS DARTMOUTH

Cumulative failures



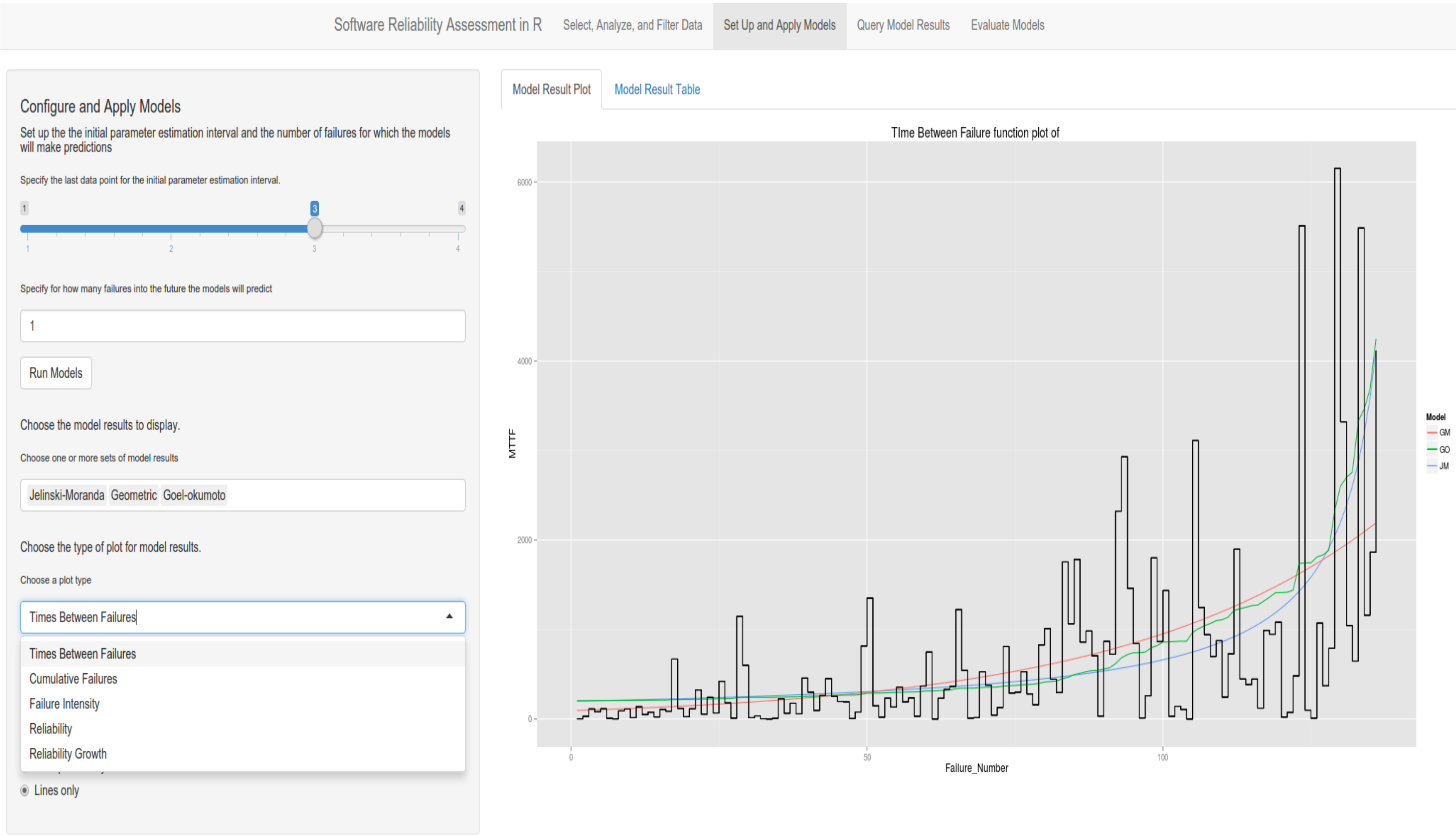
Failure intensity



UMass Dartmouth

UNIVERSITY OF MASSACHUSETTS DARTMOUTH

Time between failures



Reliability growth curve