UMass | Dartmouth UNIVERSITY OF MASSACHUSETTS DARTMOUTH

Software Reliability Modeling: Heterogeneous Fault Detection Processes

- Vidhyashree Nagaraju, MS, University of Massachusetts Dartmouth
- Lance Fiondella, PhD, University of Massachusetts Dartmouth

Outline

- Motivation
- Background
- Contribution
- NHPP SRGM with single changepoint
  - Homogeneous and heterogeneous changepoint models
  - Illustrations
- Conclusions and future work

Motivation

- Software reliability growth models (SRGM)
  - Well-established methodology
  - Many based on non-homogeneous Poisson process (NHPP)
    - Expected number of faults if debugging performed indefinitely
    - Failure intensity
    - Mean time to failure (MTTF)
    - Reliability
- SRGMs assume
  - Fault detection rate dependent on software testing time

Background

- During software testing
  - Failure detection affected by many factors
    - Change in testing environment
    - Testing strategy
    - Integration testing
    - Resource allocation
- Impact of factors on the fault detection process – Changepoints
  - Failing to model changepoints may adversely affect system assessment
- Several NHPP SRGM consider changepoint since 1992
  - Consider only homogeneous combinations of failure distribution before and after changepoints

Contribution

- Heterogeneous single changepoint models
  - Applies ECM algorithm to maximize log-likelihood
- Compared with existing homogeneous models
  - Demonstrate heterogeneous changepoint models characterize some data sets better than homogeneous models

NHPP SRGM

- Stochastic process
  - Counts number of events observed as function of time
  - In software reliability,
    - 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 – expected number of faults detected with indefinite testing
    - $F(t)$ - cumulative distribution function (CDF)

NHPP SRGM

- Substituting exponential distribution for $F(t)$

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

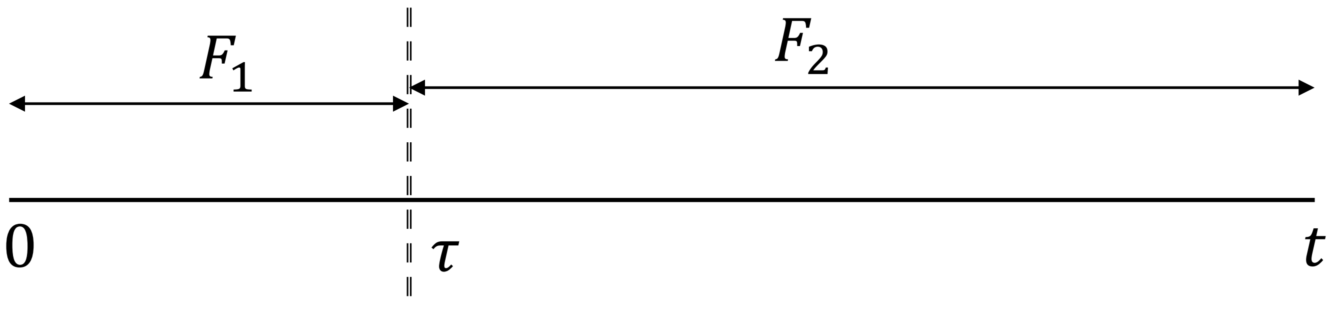
  - b - fault detection rate
  - Also known as Goel-Okumoto (GO) SRGM
- Substituting delayed S-shaped (DSS) distribution for $F(t)$

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

  - bte^{-bt} – can characterize delay between time of failure observation and reporting or fault masking

NHPP SRGM with single Changepoint

- $\mathbf{T} = \langle t_1, t_2, \dots, t_n \rangle$ - vector of failure time data with single changepoint
- $F_1(t)$ and $F_2(t)$ - failure distributions before and after τ



- MVF of NHPP SRGM with single changepoint

$$m(t) = \begin{cases} a \times F_1(t), & 0 \leq t \leq \tau \\ a\big(F_1(\tau) + F_2(t-\tau)\big), & t > \tau \end{cases}$$

NHPP SRGM w/single Changepoint (2)

- NHPP SRGM with single changepoint can be
  - Homogeneous
    - If $F_1(t)$ and $F_2(t)$ follow similar distribution
  - Heterogeneous
    - If $F_1(t)$ and $F_2(t)$ follow different distribution
- Changepoint τ identified by maximizing likelihood for each value $\tau \in \left(2, (n-1)\right)$

Homogenous changepoint models

1. GO-GO SRGM
   - If $F_1(t)$ and $F_2(t)$ follow **exponential** distribution

$$m(t) = \begin{cases} a\left(1 - e^{-bt}\right), & 0 \le t \le \tau \\ a\left(\left(1 - e^{-b\tau}\right) + \left(1 - e^{-b(t-\tau)}\right)\right), & t > \tau \end{cases}$$

2. DSS-DSS SRGM
   - If $F_1(t)$ and $F_2(t)$ follow **S-shaped** distribution

$$m(t) = \begin{cases} a\left(1 - (1 + bt)e^{-bt}\right), & 0 \le t \le \tau \\ a\left(\left(1 - (1 + b\tau)e^{-b\tau} + \left(1 + b(t - \tau)\right)e^{-b(t-\tau)}\right)\right), & t > \tau \end{cases}$$

Heterogenous changepoint models (2)

1. GO-DSS SRGM
   - If $F_1(t)$ and $F_2(t)$ follows **exponential** and **S-shaped** distribution respectively

$$m(t) = \begin{cases} a\left(1 - e^{-bt}\right), & 0 \le t \le \tau \\ a\left(\left(1 - e^{-b\tau}\right) + \left(\left(1 + b(t-\tau)\right)e^{-b(t-\tau)}\right)\right), & t > \tau \end{cases}$$

2. DSS-GO SRGM
   - If $F_1(t)$ and $F_2(t)$ follows **S-shaped** and **exponential** distribution respectively

$$m(t) = \begin{cases} a\left(1 - (1+bt)e^{-bt}\right), & 0 \le t \le \tau \\ a\left(\left(1 - (1+b\tau)e^{-b\tau}\right) + \left(1 - e^{-b(t-\tau)}\right)\right) & t > \tau \end{cases}$$

Parameter Estimation Method

- Maximum likelihood estimation (MLE) maximizes the likelihood function to identify numerical values of model parameters
- NHPP failure times data log-likelihood

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

  - $\mathbf{\Theta}$ – vector of model parameters
  - $\lambda(t) \coloneqq \frac{dm(t)}{dt}$ - instantaneous failure rate at time t
- ECM algorithm applied to identify MLEs

Initial parameter estimates selection

- Function of $f(t_i;\Theta)$

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

and

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

Initial parameter estimates selection (2)

- GO SRGM parameters

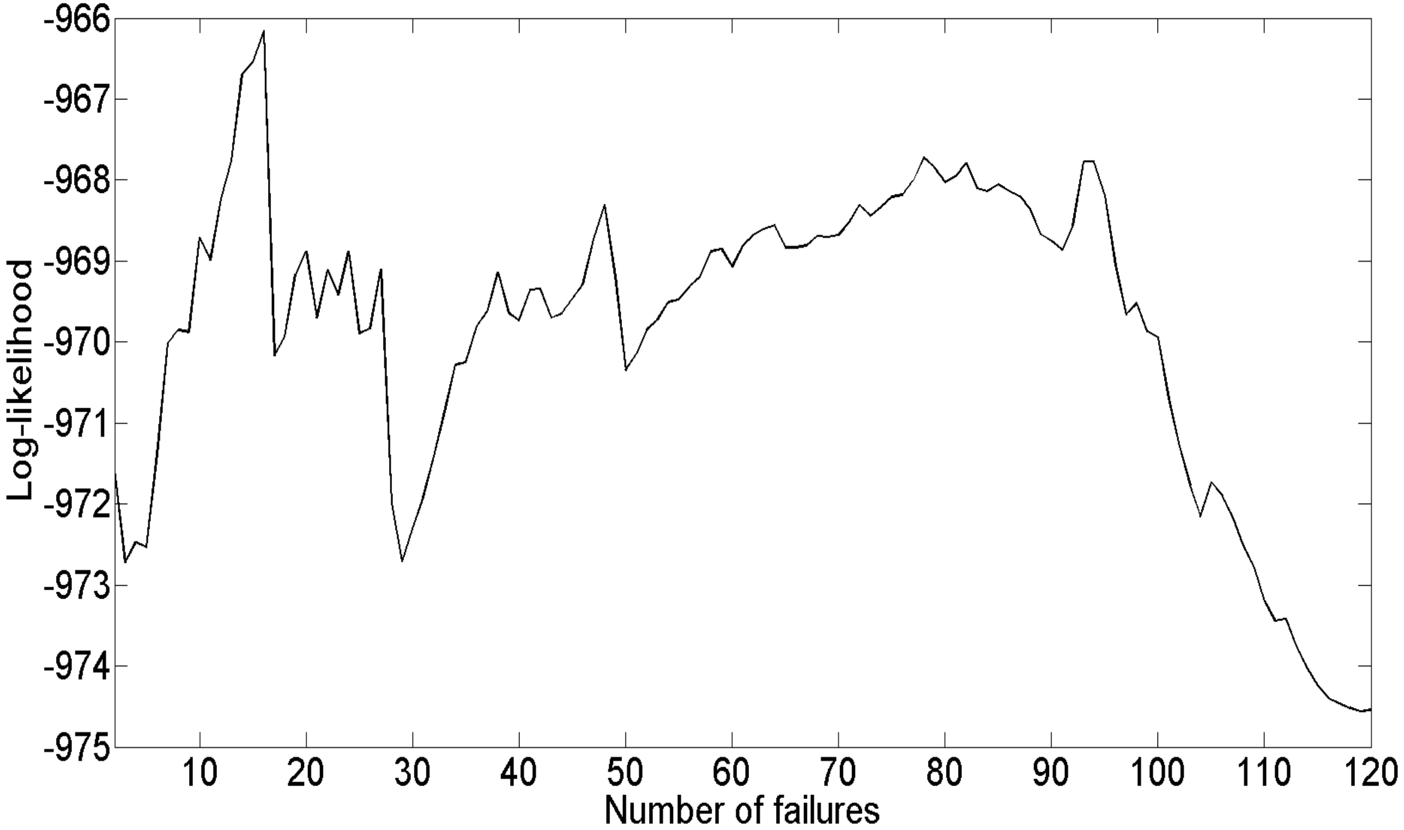
$$a^{(0)} = n, \qquad b^{(0)} = \frac{n}{\sum_{i=1}^{n} t_i}$$

- Similarly, for DSS SRGM

$$b^{(0)} = \frac{2n}{\sum_{i=1}^{n} t_i}$$

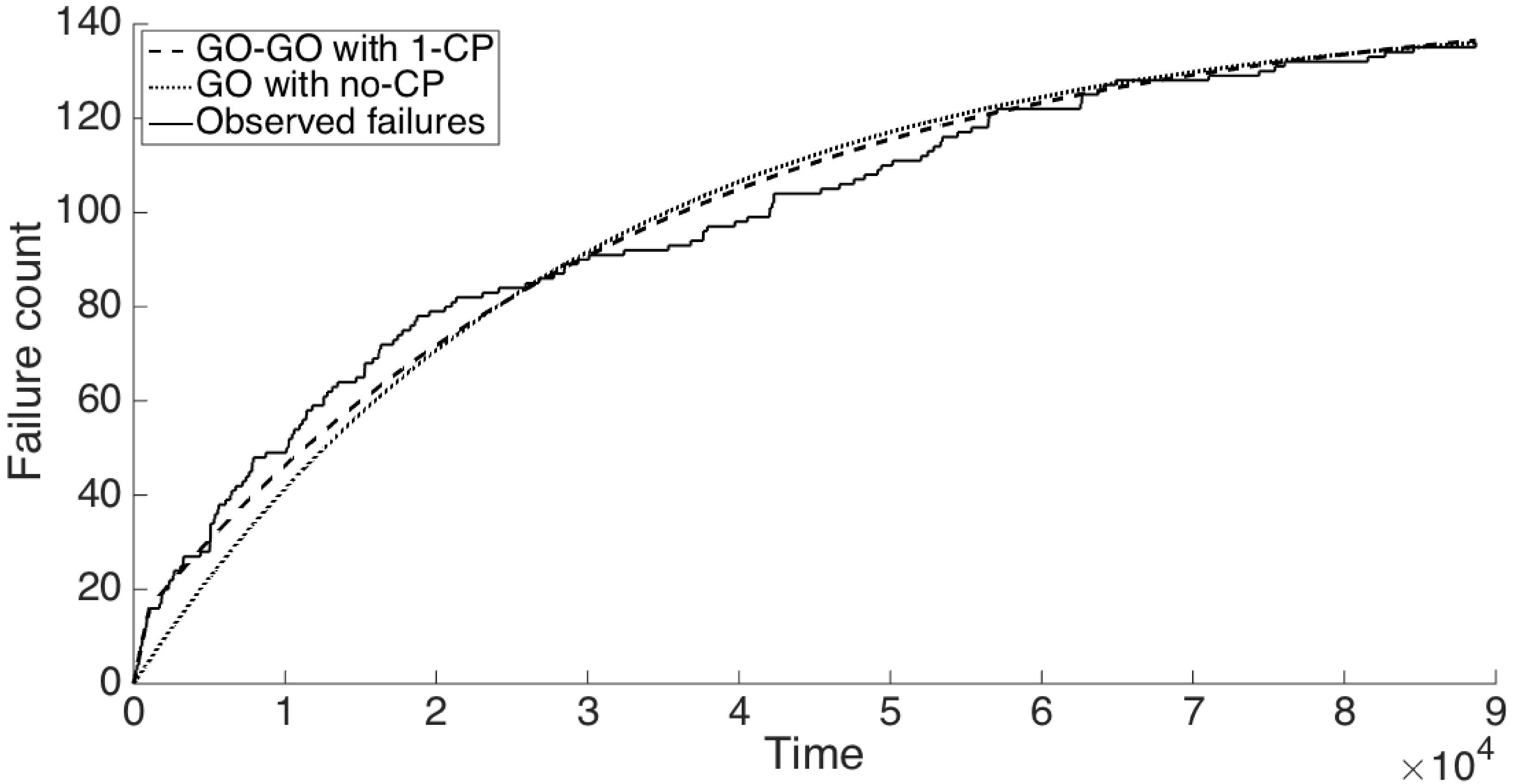
Illustrations

GO-GO SRGM applied to SYS1 dataset



Changepoint at $\tau = 16$, ($t_{\tau=16} = 1{,}056$) maximizes likelihood

GO SRGM with and without CP



GO with changepoint fits data better

GO SRGM with and without CP (2)

Datasets	0-CP GO AIC	1-CP GO AIC	Difference
SYS1	1953.61	1938.31	**15.3**
SYS2	1377.08	1377.01	0.07
SYS3	2190.91	2187.34	**3.57**
S2	902.187	894.113	**8.074**
S27	673.603	660.634	**12.969**
SS3	3468.19	3447.77	**20.42**
SS4	2576.69	2569.72	**6.97**
CSR1	4793.71	4690.27	**103.44**
CSR2	1848.57	1818.25	**30.32**
CSR3	1216.16	1193.56	**22.6**

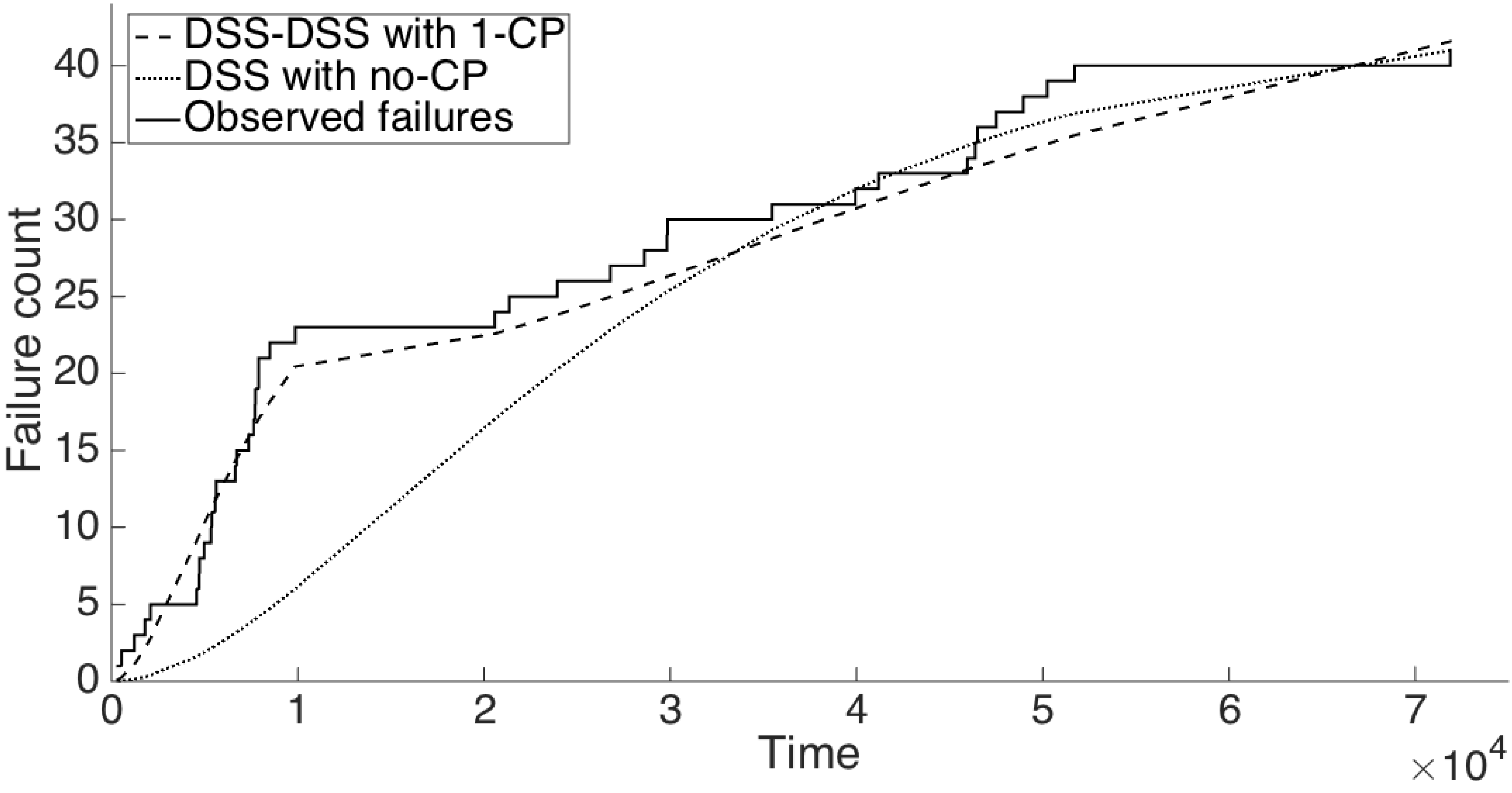
GO with changepoint significantly better on 9 of 10 datasets

DSS SRGM with and without CP

Datasets	0-CP DSS AIC	1-CP DSS AIC	Difference
SYS1	2075.15	1995.15	**80**
SYS2	1410.55	1387.59	**22.96**
SYS3	2260.81	2203.73	**57.08**
S2	962.427	902.025	**60.402**
S27	702.401	657.854	**44.547**
SS3	3561.85	3446.40	**115.45**
SS4	2599.31	2575.37	**23.94**
CSR1	5078.03	4713.92	**364.11**
CSR2	1913.69	1800.17	**113.52**
CSR3	1290.59	1211.28	**79.31**

DSS with changepoint significantly better than model without

DSS SRGM with and without CP (2)



DSS with changepoint fits S27 data better

Homogeneous vs. Heterogeneous models

Datasets	GO-GO	DSS-DSS	GO-DSS	DSS-GO
SYS1	**1938.31**	1995.15	1944.65	1950.96
SYS2	1377.01	1387.59	1377.43	**1376.3**
SYS3	2187.34	2203.73	2185.31	**2183.22**
S2	**894.11**	902.025	894.18	900.616
S27	660.63	**657.854**	660.35	659.217
SS3	3447.77	3446.40	**3440.58**	3444.45
SS4	2569.72	2575.37	**2567.81**	2574.15
CSR1	4690.27	4713.92	4743.98	**4678.29**
CSR2	1818.25	1800.17	1819.41	**1799.69**
CSR3	**1193.56**	1211.28	1200.85	1209.33

Heterogeneous models preferred in six out of ten data sets

Summary, Conclusion, and Future work

Summary and Conclusion

- Developed heterogeneous single changepoint model
- Assesses models with and without changepoint
  - Models with changepoint often characterizes data better
- Compared homogeneous vs. heterogeneous models
  - Heterogeneous models outperformed homogeneous models on 60% of datasets considered

Future work

- Theoretical
  - Model selection procedure considering additional combinations of possible models
  - Performance optimization of algorithmic approach
- Empirical
  - Methods to assess the effectiveness of milestone decisions and testing between milestones
  - Combine with metrics-based models to identify effective activities

Acknowledgement

- This work was partially supported by (i) the Naval Air Warfare Center (NAVAIR) under Award Number N00421-16-P-0521 and (ii) the National Science Foundation (NSF) (#1526128).