

## Software Defect Prediction via Operating Characteristic Curves

Nazanin Hazrati\*, A. Ben Hamza\*, Yucel Karabulut•, Nolwen Mahe†

\*Concordia University, Montreal, Canada

•SAP Labs, LLC, Palo Alto, CA, USA

†SAP Labs Canada, Montreal, Canada

# Introduction

## Motivation:

- ❑ Software defects cost businesses and software industry billions of dollars in lost productivity.
- ❑ Software quality impacts development costs, delivery schedules, and user satisfaction.
- ❑ Defects can cause serious damage and even physical harm.

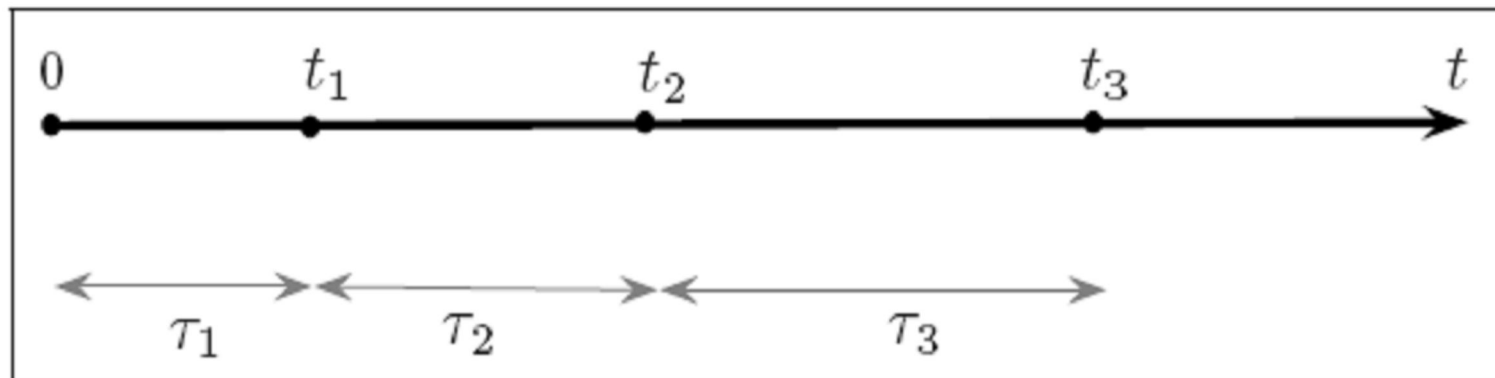
## Therefore,

- ❑ Early defect prevention and removal techniques can substantially enhance the profit realized on software products.
- ❑ Knowledge about how many defects to expect in a software product at any given stage during its development process is a very valuable asset.
- ❑ Being able to estimate the defects related cost driver characteristics will substantially improve the decision processes relevant for e.g. releasing a software product .
- ❑ How to devise efficient and reliable defect prediction models?

# Introduction

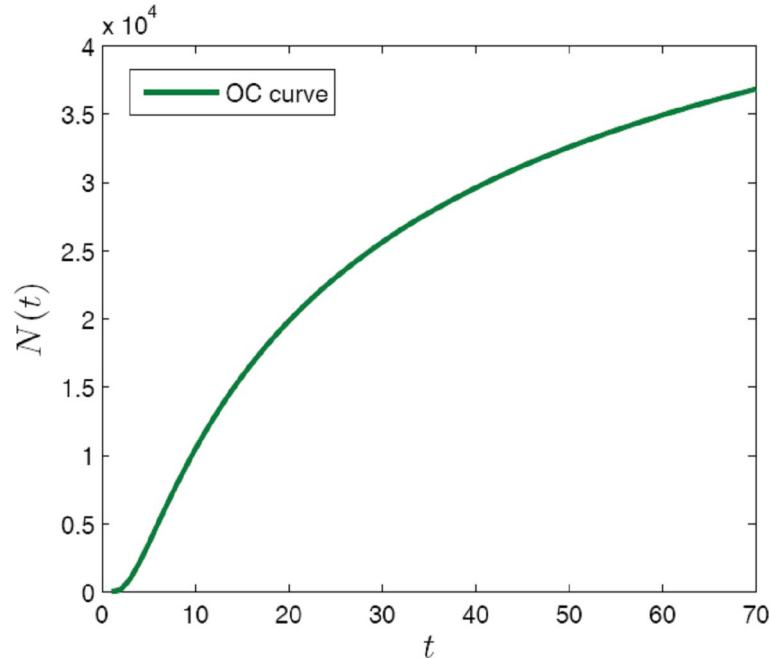
Three basic forms of software defect data:

- ❑ Ordered defect times:  $0 < t_1 < t_2 < \dots < t_n$
- ❑ Inter-defect times:  $\tau_i$  where  $\tau_i = t_i - t_{i-1}$  for  $i = 1, \dots, n$
- ❑ The cumulative number of defects.



# Defect Prediction using POC Curves

- Operating Characteristic (OC) Curves are motivated by the geometric concept of the cumulative number of defects plot.



$$N(t) = \left( \frac{\sqrt{2p(t)}}{\delta} \right)^2 \left( \chi_{\alpha, \delta}^2 + \chi_{\beta, \delta}^2 \right)^2$$

$N(t)$  : the predicted cumulative number of defects

$p$  : the cumulative number of defects

$\delta$  :  $t - t_k$  (where :  $k$  is a index)

$\alpha$  : the error of the type I ( significance level )

$\beta$  : the error of the type II

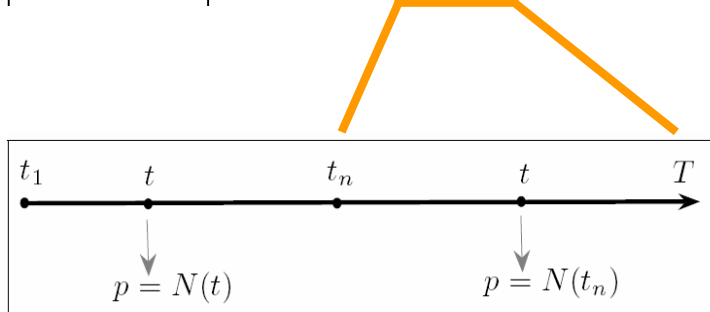
$\chi^2$  : chi-square statistic

- OC curves do NOT take into account historical data

# Defect Prediction using POC Curves

- Predictive Operating Characteristic (POC) Curves take into account the historical data

Month	Cumulative number of defects
1	17
2	39
3	53
4	87
5	106
6	140



$$p(t) = \begin{cases} N(t) & \text{if } t \leq t_n \\ N(t_n) & \text{if } t_n \leq t \leq T \end{cases}$$

$$N(t) = \left( \frac{\sqrt{2p(t)}}{\delta} \right)^2 \left( \chi_{\alpha, \delta}^2 + \chi_{\beta, \delta}^2 \right)^2$$

# Experimental Results

## Data description

- ❑ Monthly software failures.
- ❑ A period of 60 months.

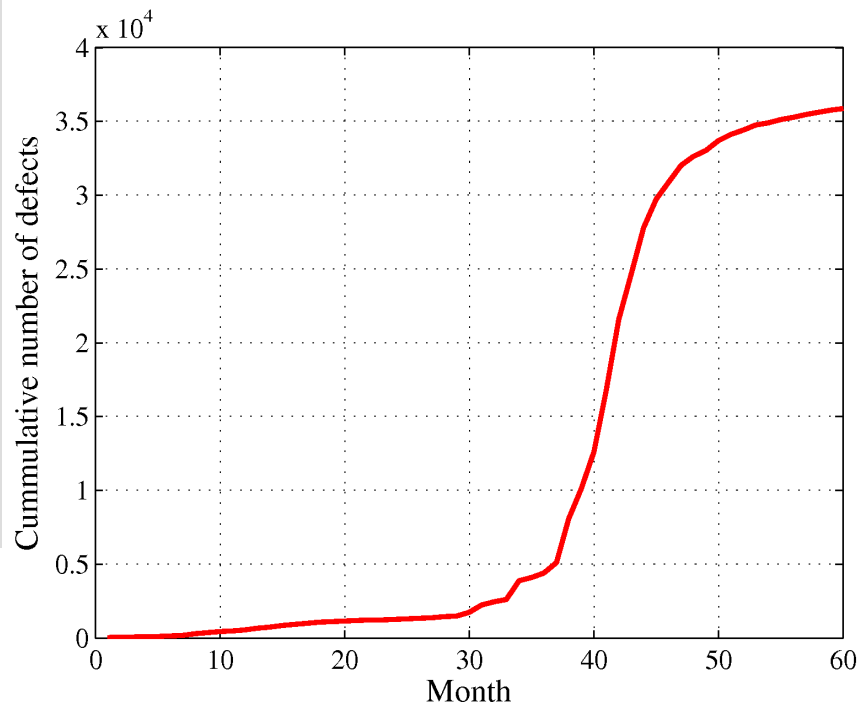
## Data

Month	Cumulative number of defects	Month	Cumulative number defects
1	17	31	2,217
2	39	32	2,430
3	53	33	2,586
4	87	34	3,884
5	106	35	4,099
6	140	36	4,385
7	165	37	5,104
8	286	38	8,074
9	359	39	10,120
10	412	40	12,618
11	461	41	16,715
12	555	42	21,606
13	654	43	24,592
14	747	44	27,789
15	836	45	29,739
16	926	46	30,843
17	989	47	32,011
18	1,049	48	32,599
19	1,103	49	33,010
20	1,152	50	33,707
21	1,182	51	34,103
22	1,213	52	34,426
23	1,225	53	34,736
24	1,266	54	34,903
25	1,306	55	35,110
26	1,331	56	35,261
27	1,363	57	35,440
28	1,443	58	35,614
29	1,495	59	35,763
30	1,737	60	35,876

# CRM data

## Data description

- ❑ Monthly software defects
- ❑ Over a period of 60 months.



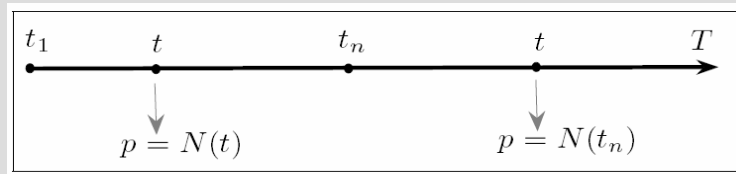
## Data

Month	Cumulative number of defects	Month	Cumulative number of defects
1	17	31	2,217
2	39	32	2,430
3	53	33	2,586
4	87	34	3,884
5	106	35	4,099
6	140	36	4,385
7	165	37	5,104
8	286	38	8,074
9	359	39	10,120
10	412	40	12,618
11	461	41	16,715
12	555	42	21,606
13	654	43	24,592
14	747	44	27,789
15	836	45	29,739
16	926	46	30,843
17	989	47	32,011
18	1,049	48	32,599
19	1,103	49	33,010
20	1,152	50	33,707
21	1,182	51	34,103
22	1,213	52	34,426
23	1,225	53	34,736
24	1,266	54	34,903
25	1,306	55	35,110
26	1,331	56	35,261
27	1,363	57	35,440
28	1,443	58	35,614
29	1,495	59	35,763
30	1,737	60	35,876

# Experimental Results

## Integrate Activity

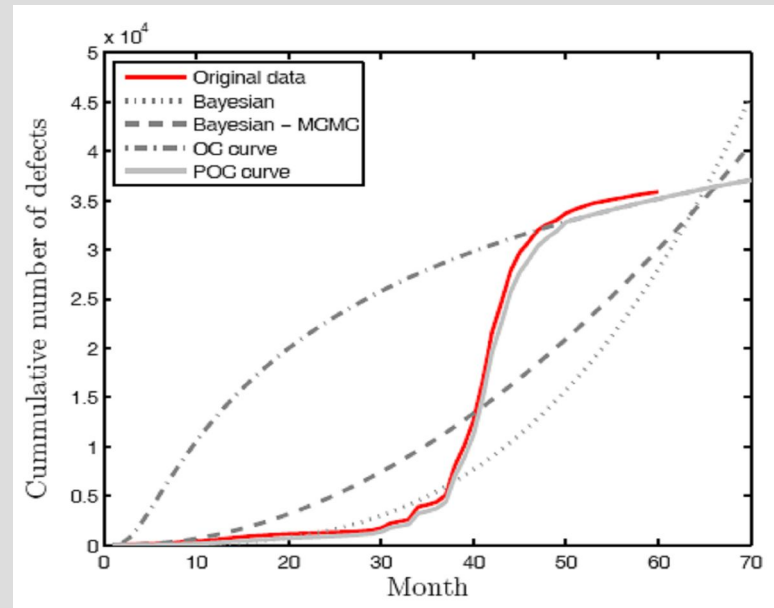
Parameter  $p$



Integrate into function

$$N(t) = \left(\frac{\sqrt{2p}}{\delta}\right)^2 (\chi_{\alpha,\delta}^2 + \chi_{\beta,\delta}^2)^2$$

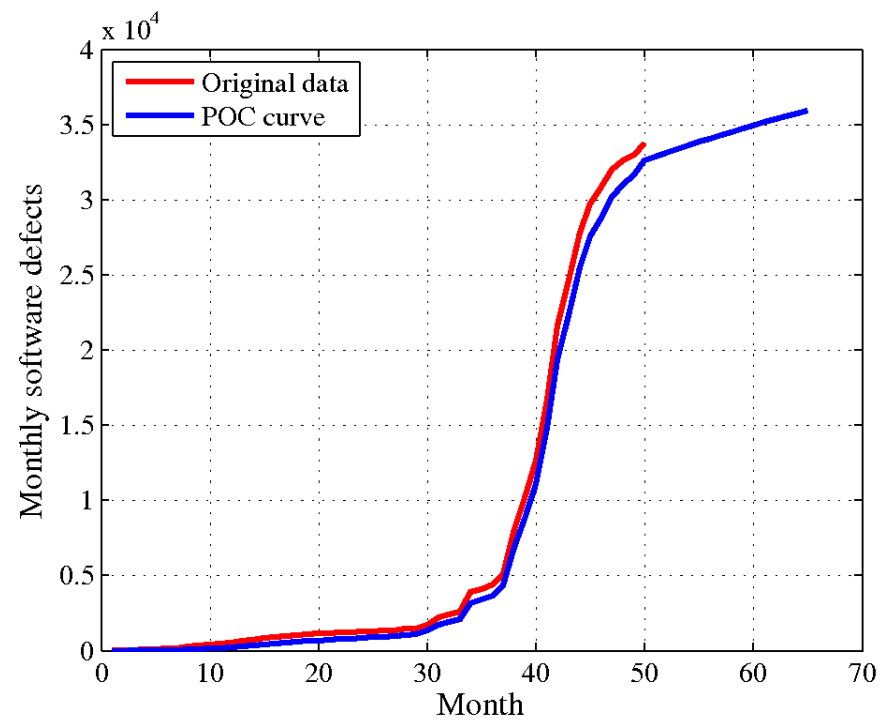
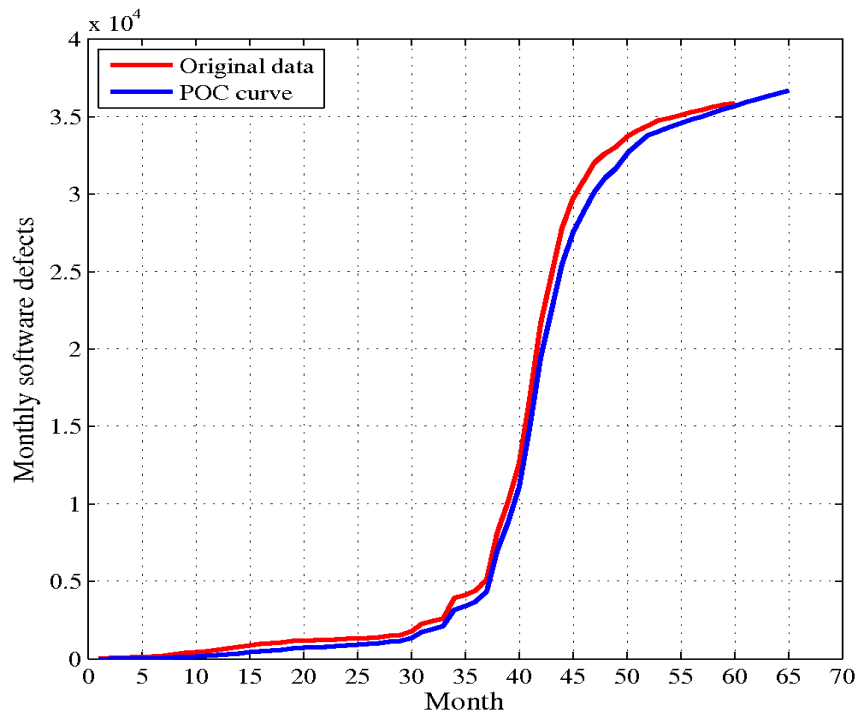
## Predictive operating characteristic (POC) curve



# CRM data: Experimental Results

Prediction at 65<sup>th</sup> month

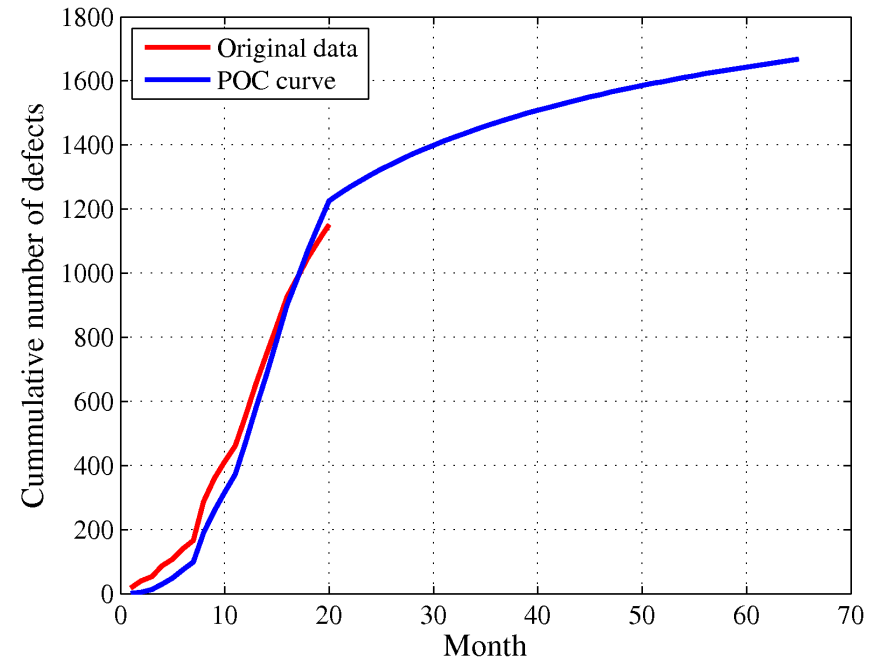
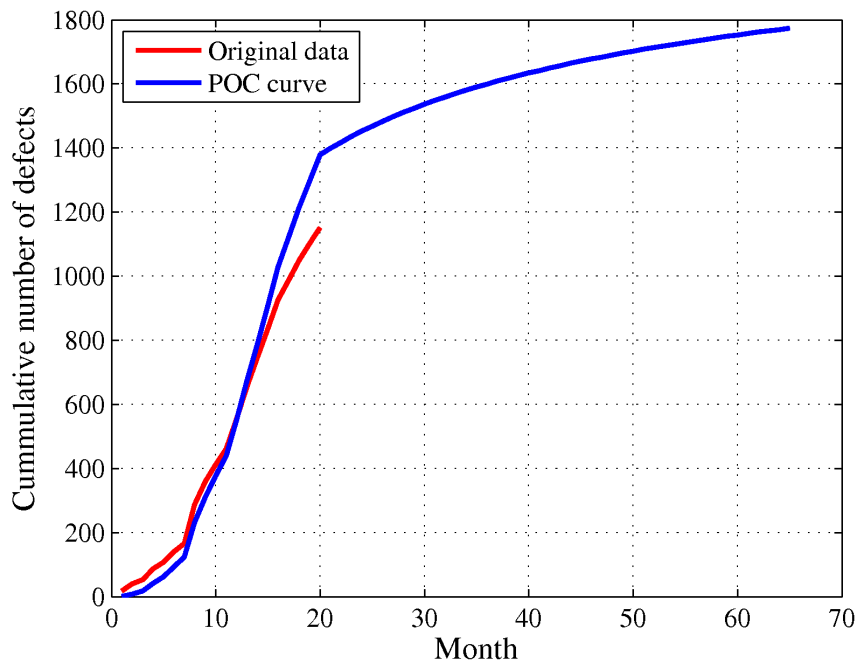
Trimming at 50<sup>th</sup> month



# CRM data: Experimental Results

Trimming at 20<sup>th</sup> month ( $\alpha=0.5$ )

Trimming at 20<sup>th</sup> month ( $\alpha=0.6$ )



# Experimental Results

## Skill Score

An error statistic used to quantify the accuracy of prediction models:

$$SS = 1 - \frac{\sqrt{\frac{1}{T_m} \sum_{t=1}^{T_m} (N_o(t) - N_p(t))^2}}{\sqrt{\frac{1}{T_{m-1}} \sum_{t=1}^{T_m} (N_o(t) - \bar{N}_o)^2}}$$

(  $N_p$  : predicted data;  $N_o$  : observed data )

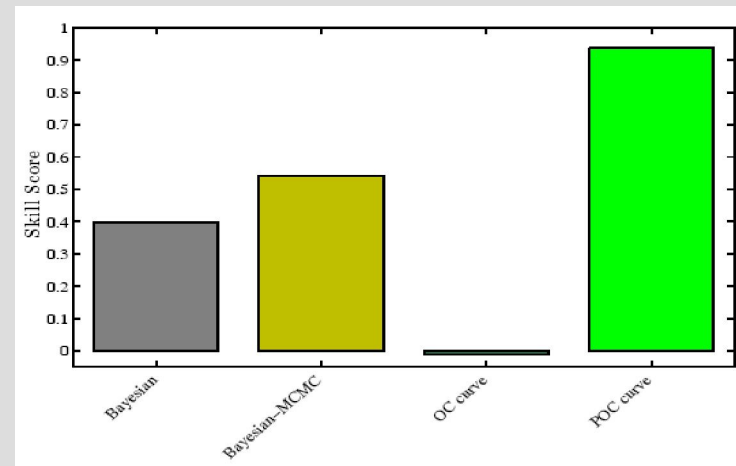
Prediction is better:

SS is closer to one

Prediction is poor:

SS is closer to zero.

## Skill score results



# Experimental Results

## Nash-Sutcliffe model efficiency coefficient

Shows the model's ability to predict

$$E = 1 - \frac{\sum_{t=1}^{T_m} (N_o(t) - N_p(t))^2}{\sum_{t=1}^{T_m} (N_o(t) - \bar{N}_o)^2}$$

(  $N_p$  : predicted data;  $N_o$  : observed data )

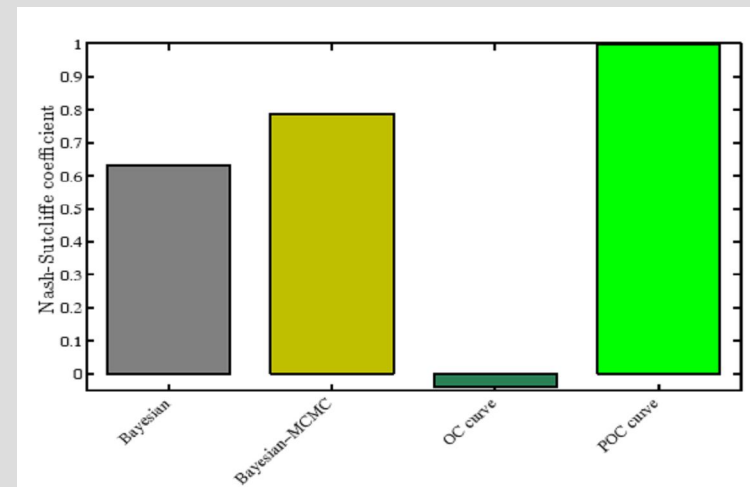
Prediction is perfect:

E is equal to one

Prediction is poor:

E is equal to or less than zero.

## Nash-Sutcliffe model efficiency coefficient results



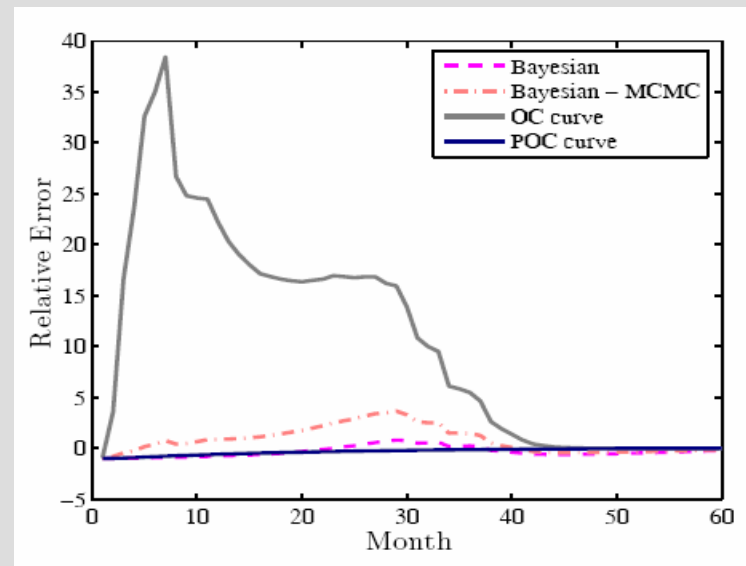
# Experimental Results

## Relative error

It measures how close the predicted data  $N_p$  and the observed data  $N_o$  are:

$$RE = \frac{N_p(t) - N_o(t)}{N_o(t)}$$

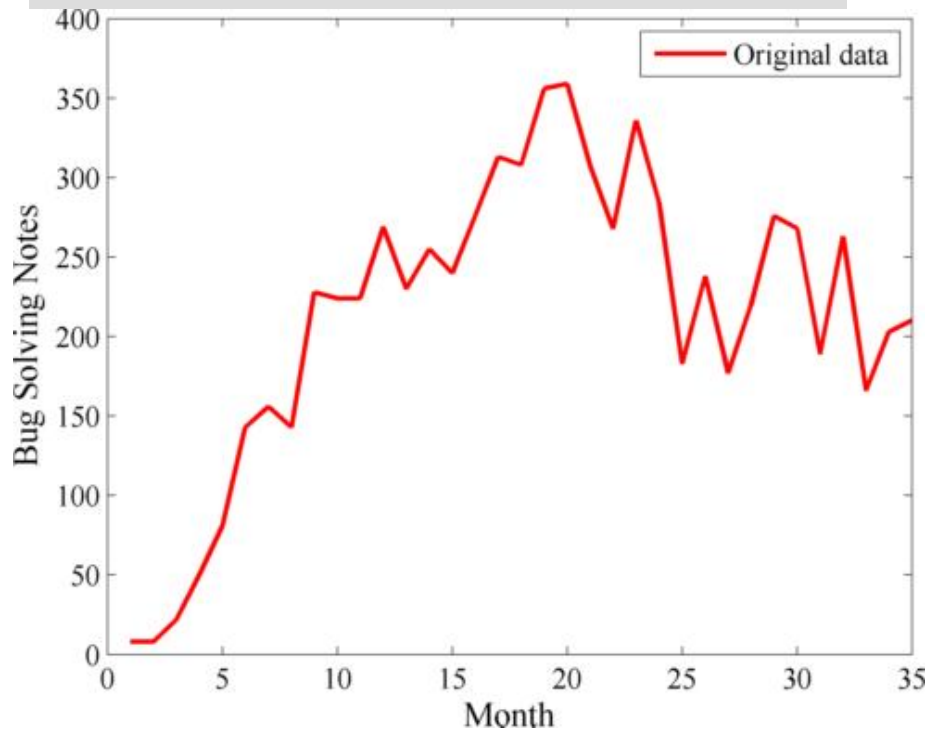
## Relative error results



# Bug Solving Notes

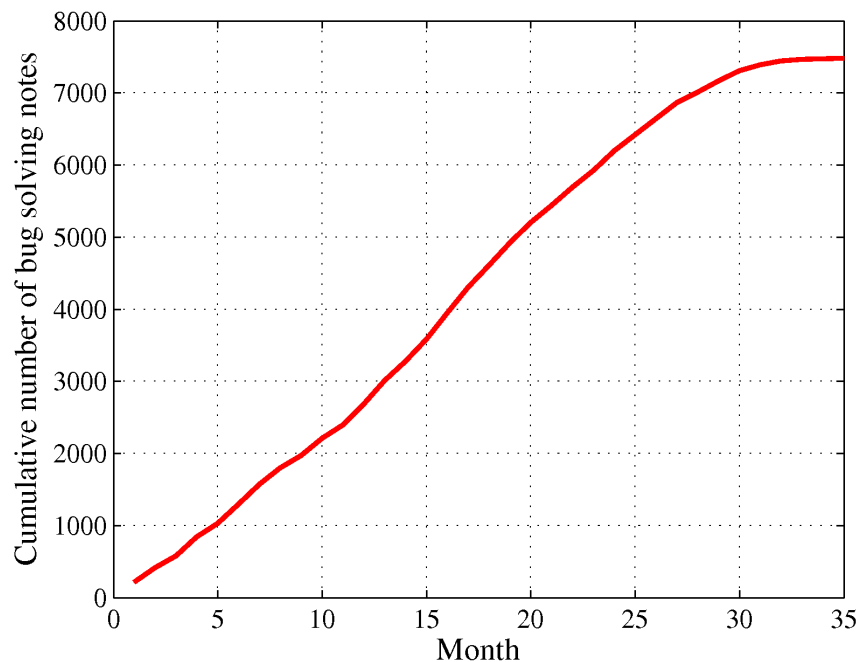
## Release Testing Data description

- ❑ Monthly bug solving notes
- ❑ Over a period of 35 months.



Reporting Month	Bug Solving Notes
10.2008	210
09.2008	203
08.2008	166
07.2008	263
06.2008	189
05.2008	268
04.2008	276
03.2008	220
02.2008	177
01.2008	238
12.2007	183
11.2007	284
10.2007	336
09.2007	268
08.2007	307
07.2007	359
06.2007	356
05.2007	308
04.2007	313
03.2007	276
02.2007	240
01.2007	255
12.2006	230
11.2006	269
10.2006	224
09.2006	224
08.2006	228
07.2006	143
06.2006	156
05.2006	143
04.2006	81
03.2006	50
02.2006	22
01.2006	8
12.2005	8

# Defect Prediction using POC Curves



$$N(t) = \left( \frac{\sqrt{2p(t)}}{\delta} \right)^2 \left( \chi_{\alpha,\delta}^2 + \chi_{\beta,\delta}^2 \right)^2$$

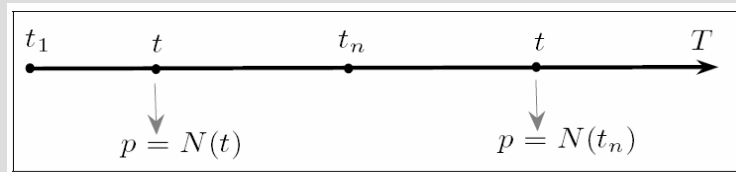
$N(t)$  : predicted number of defects  
 $p$  : number of defects  
 $\delta$  :  $t - t_k$  (predicted\_time - start\_time)  
 $\alpha$  : type I error ( significance level )  
 $\beta$  : type II error  
 $\chi_{\alpha,\delta}^2$  : chi-square statistic

- ❑ This dataset contains monthly bug solving notes that were reported over a period of 35 months

# Bug data: Experimental Results

Integrate OC concept

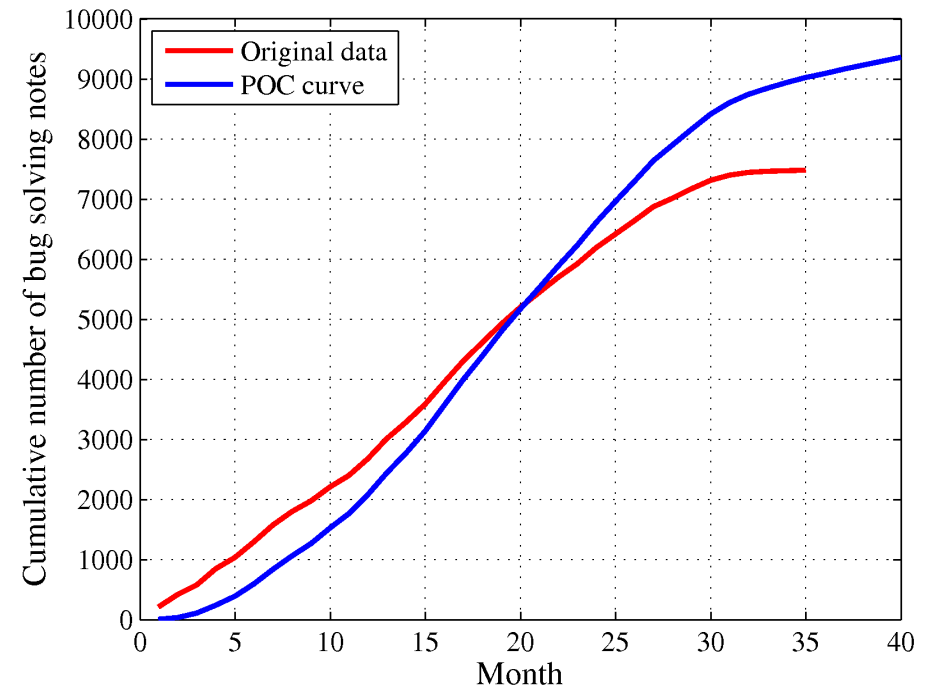
Parameter  $p$



Integrate into function

$$N(t) = \left( \frac{\sqrt{2p(t)}}{\delta} \right)^2 \left( \chi_{\alpha,\delta}^2 + \chi_{\beta,\delta}^2 \right)^2$$

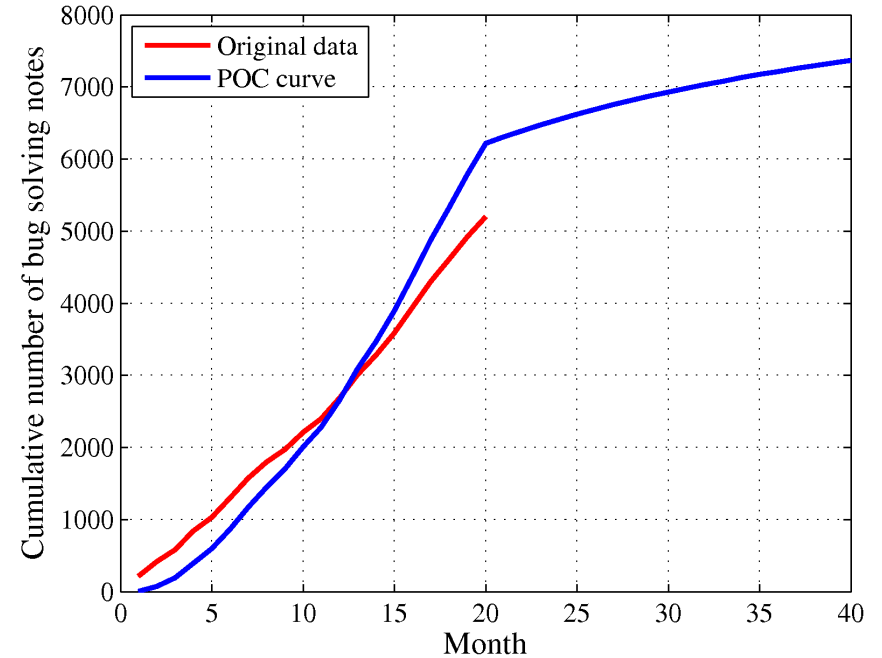
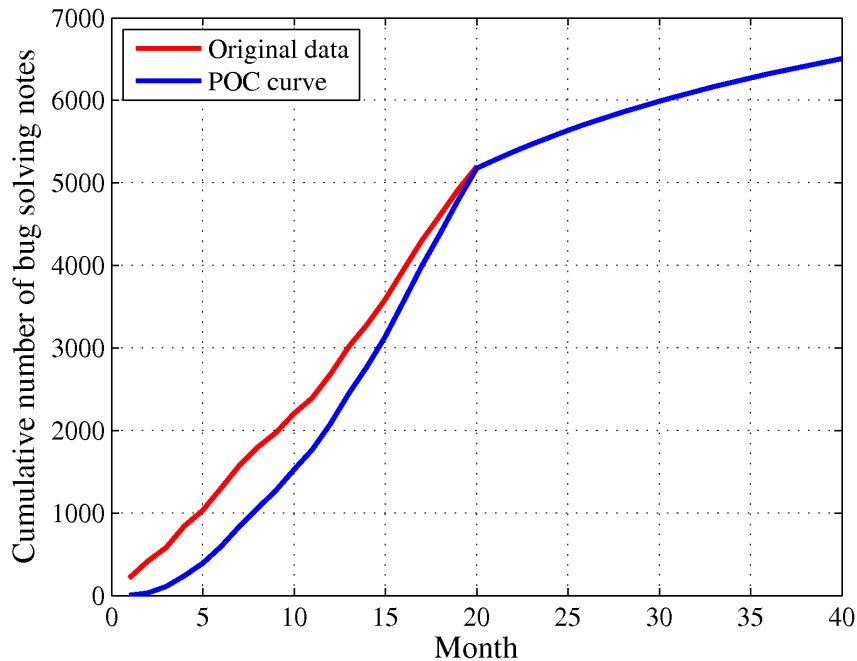
Predictive operating characteristic (POC) curve (alpha=0.65)



# Bug data: Experimental Results (cont.)

Trimming at 20<sup>th</sup> month (alpha=0.65)

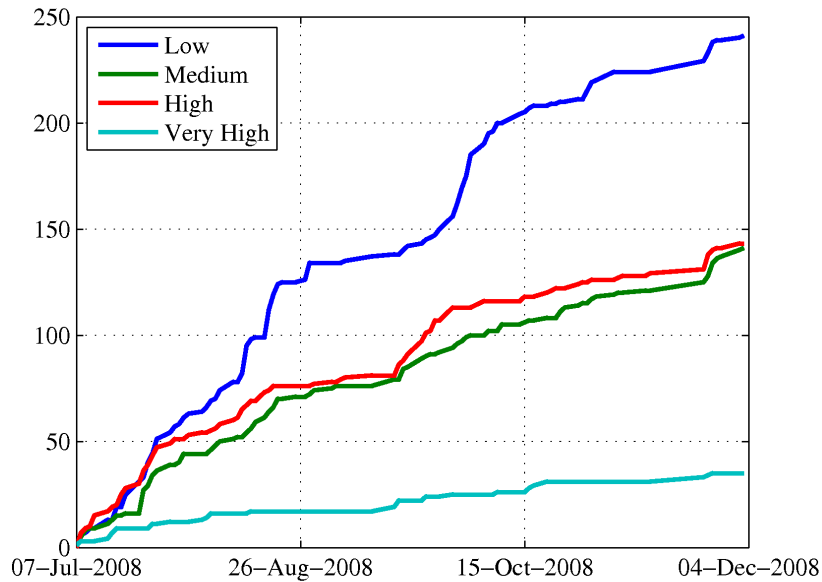
Trimming at 20<sup>th</sup> month (alpha=0.5)



# Bugs Severity data

## Data description

- ❑ Bugs severity
- ❑ Over a period of 90 days.

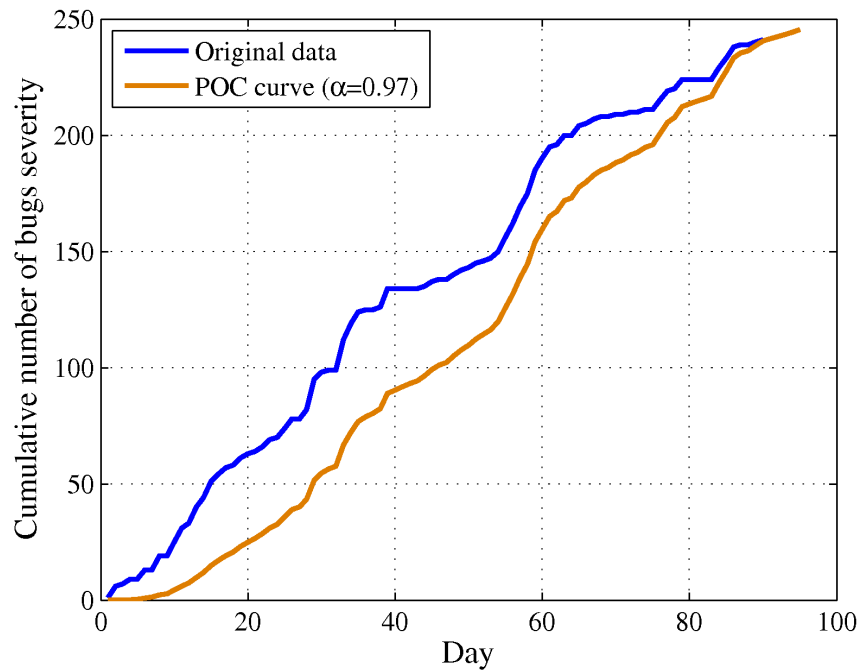


## Data

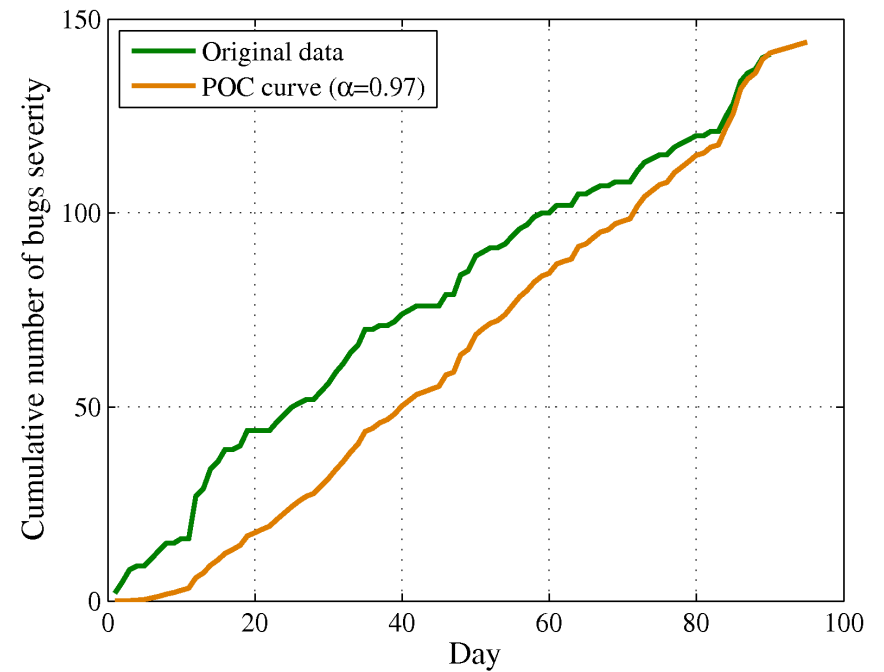
Created on	Low	Medium	High	Very High
28/06/2008	8	1	0	0
03/12/2008	1	1	0	0
02/12/2008	1	3	2	0
28/11/2008	0	1	0	0
27/11/2008	1	2	1	0
26/11/2008	5	6	2	1
25/11/2008	4	3	7	1
24/11/2008	5	4	2	2
12/11/2008	0	0	1	0
11/11/2008	0	1	0	0
06/11/2008	0	0	1	0
05/11/2008	0	1	1	0
04/11/2008	4	1	0	0
31/10/2008	1	1	0	0
30/10/2008	4	2	1	0
29/10/2008	4	0	0	0
28/10/2008	0	1	1	0
27/10/2008	1	1	2	0
24/10/2008	0	2	0	0
23/10/2008	1	3	0	0
22/10/2008	0	0	1	0
21/10/2008	1	0	1	0
20/10/2008	0	1	2	2
17/10/2008	1	0	0	1
16/10/2008	2	1	0	2
15/10/2008	1	1	2	0
14/10/2008	4	0	0	0
10/10/2008	0	3	0	0
09/10/2008	4	0	0	1
08/10/2008	1	0	0	0
07/10/2008	5	2	0	0
06/10/2008	5	0	3	0
03/10/2008	10	1	0	0
02/10/2008	6	2	0	0
01/10/2008	7	1	0	0
30/09/2008	6	2	0	0
29/09/2008	6	2	6	1
26/09/2008	3	1	0	0
25/09/2008	1	0	5	0
24/09/2008	1	1	1	0
23/09/2008	2	1	4	2
22/09/2008	1	4	6	0
19/09/2008	2	1	3	0
18/09/2008	2	5	2	0
17/09/2008	0	0	5	3
16/09/2008	1	3	0	2
11/09/2008	2	0	1	0
05/09/2008	1	0	1	0
04/09/2008	0	0	1	0
03/09/2008	0	1	0	0
02/09/2008	0	1	1	0
29/08/2008	0	2	1	0
27/08/2008	1	0	0	0
25/08/2008	0	1	0	0
22/08/2008	1	0	0	0
21/08/2008	5	4	0	0
20/08/2008	7	2	2	0
19/08/2008	13	3	1	0
18/08/2008	0	2	4	0
16/08/2008	1	3	0	0
15/08/2008	3	2	2	1
14/08/2008	13	2	2	0
13/08/2008	4	0	4	0
12/08/2008	0	1	1	0
11/08/2008	4	1	2	0
08/08/2008	4	2	2	0
07/08/2008	1	2	1	0
06/08/2008	3	2	1	2
05/08/2008	2	0	0	1
04/08/2008	1	0	1	1
01/08/2008	2	0	2	0
31/07/2008	3	4	0	0
30/07/2008	1	1	0	0
29/07/2008	3	0	2	0
28/07/2008	3	3	2	1
25/07/2008	7	2	4	0
24/07/2008	4	5	4	2
23/07/2008	7	2	3	0
22/07/2008	2	11	6	0
21/07/2008	6	0	2	0
18/07/2008	6	1	3	0
17/07/2008	0	0	5	0
16/07/2008	6	2	1	2
15/07/2008	0	2	2	3
14/07/2008	4	2	2	1
11/07/2008	0	0	5	0
10/07/2008	2	1	1	0
09/07/2008	1	3	2	0
08/07/2008	5	3	7	2
07/07/2008	1	2	0	1

# Bugs Severity data: Experimental Results

Prediction at 95<sup>th</sup> day (Low)



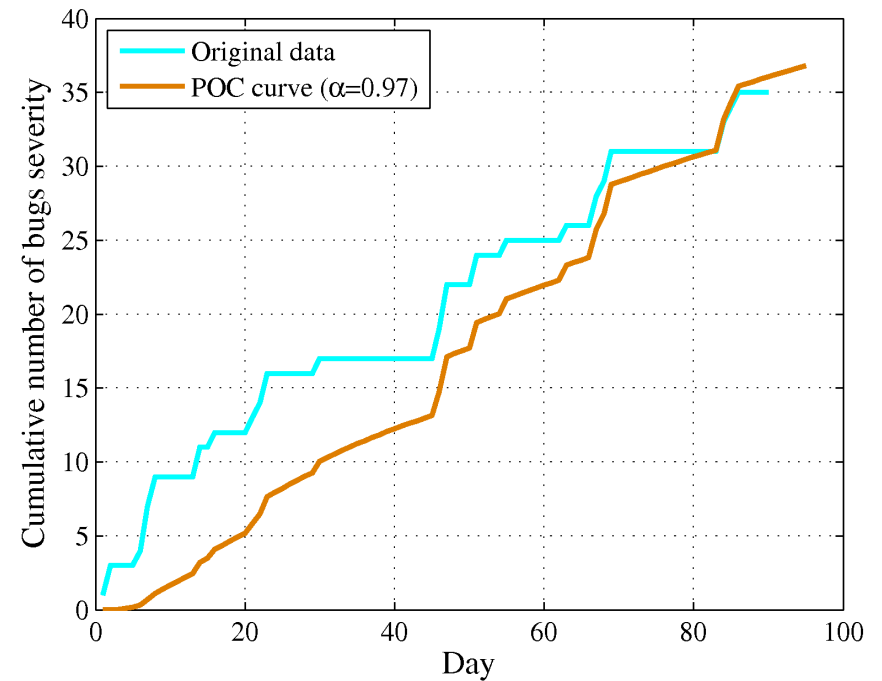
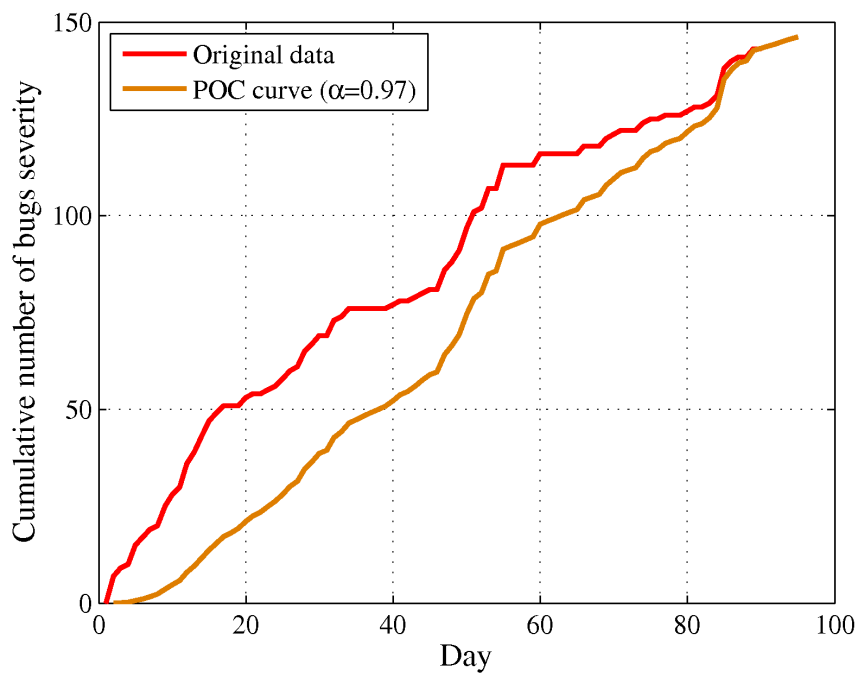
Prediction at 95<sup>th</sup> day (Medium)



# Bugs Severity data: Experimental Results

Prediction at 95<sup>th</sup> day (High)

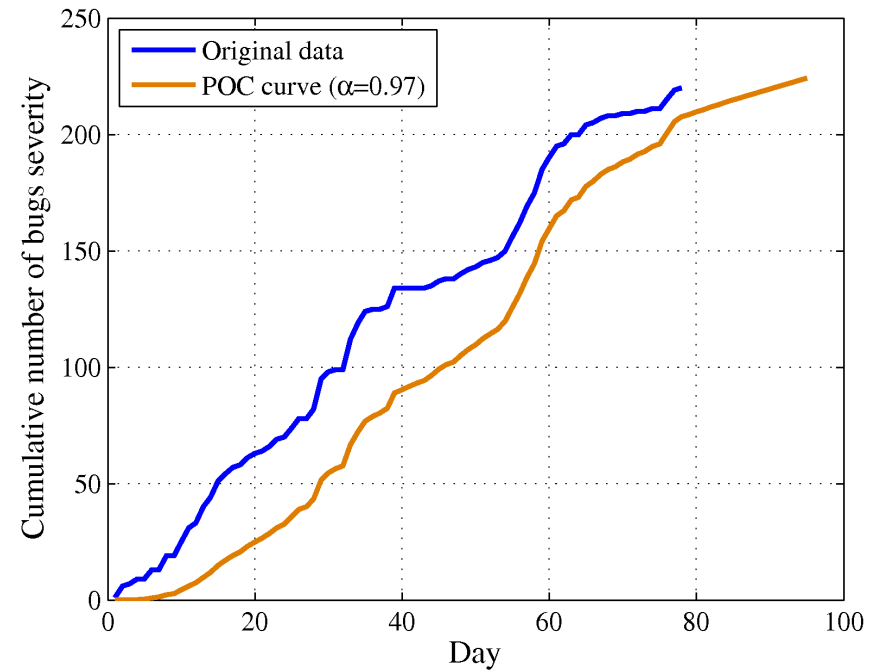
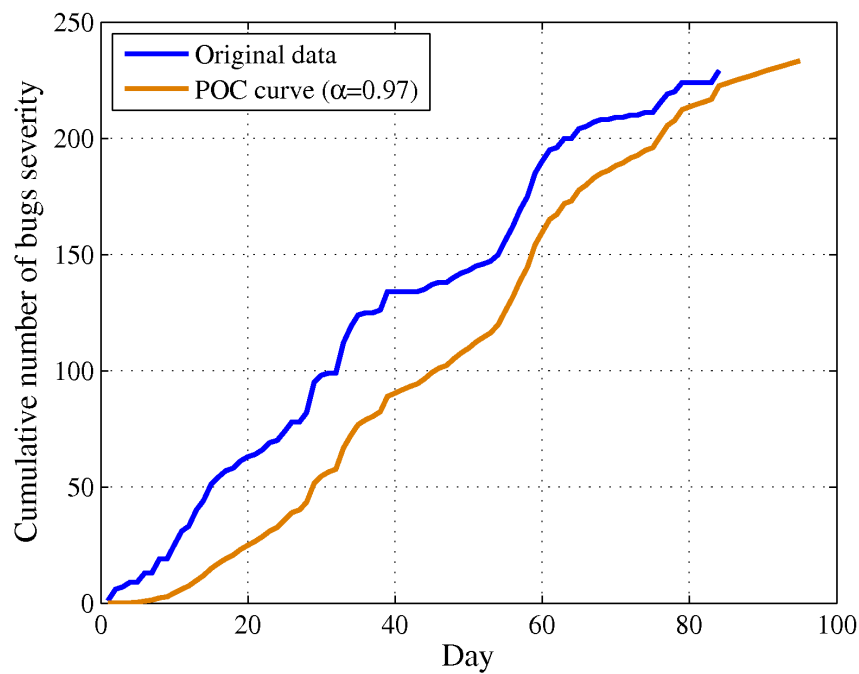
Prediction at 95<sup>th</sup> day (Very High)



# Bugs Severity data: Experimental Results

Trimming on 24/11/2008 (Low)

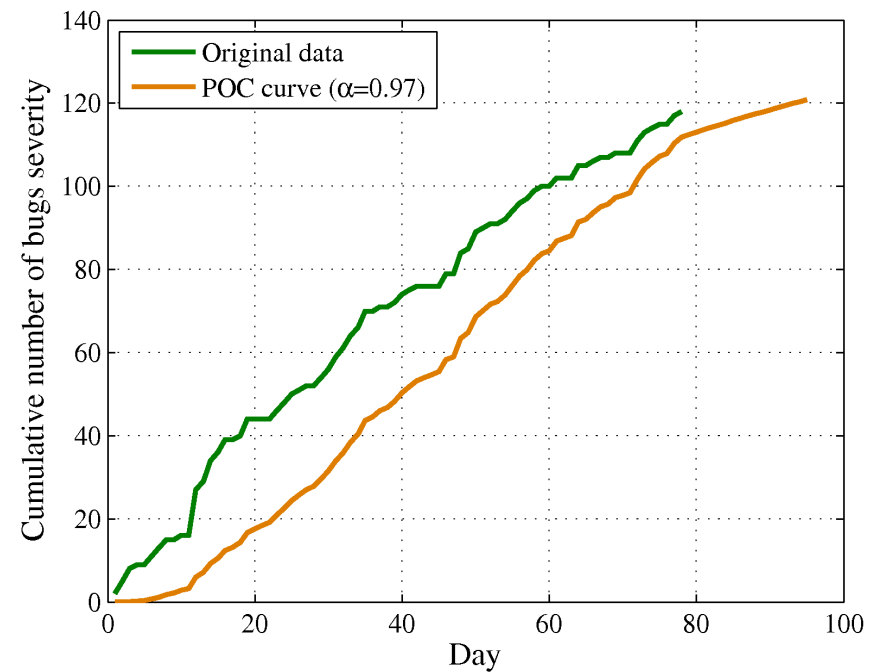
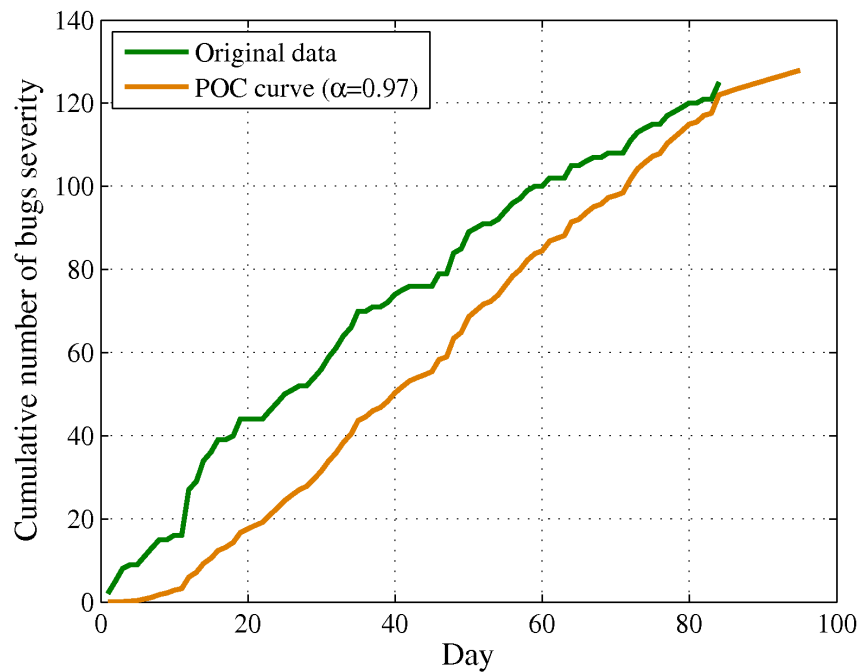
Trimming on 31/10/2008 (Low)



# Bugs Severity data: Experimental Results

Trimming on 24/11/2008 (Medium)

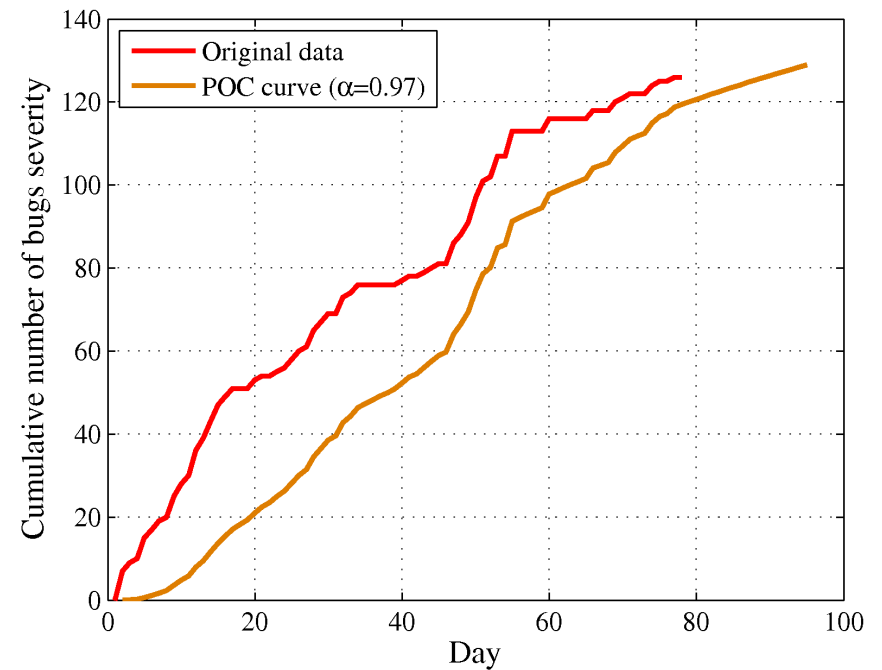
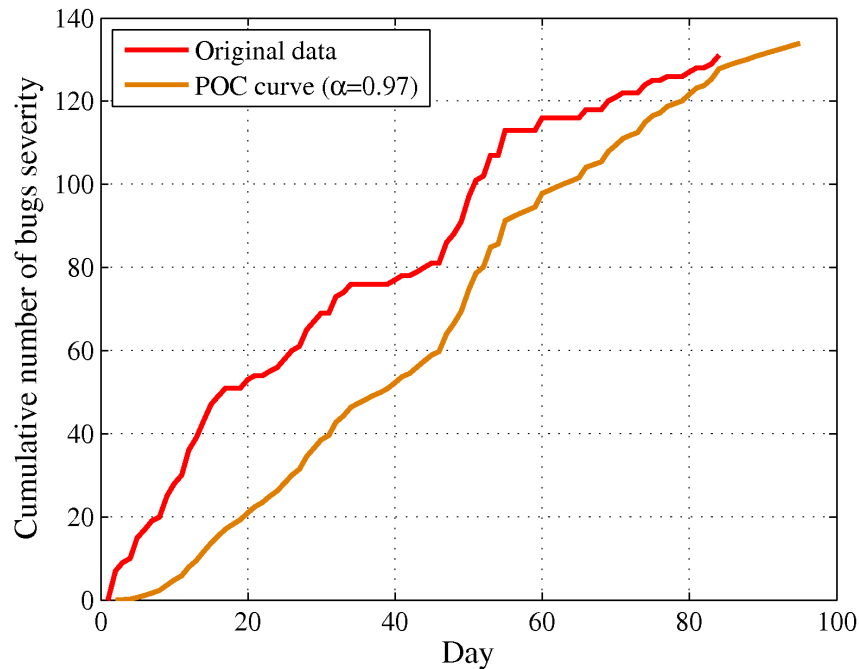
Trimming on 31/10/2008 (Medium)



# Bugs Severity data: Experimental Results

Trimming on 24/11/2008 (High)

Trimming on 31/10/2008 (High)



# Bugs Severity data: Experimental Results

Trimming on 24/11/2008 (Very High)

Trimming on 31/10/2008 (Very High)

