Cox survivial statistical analysis and its use in complex systems

Jay Benesch

2 August 2017

Outline

- Definitions
- Survival analysis - a C30 example
- Survival analysis – a C100 example
- Cox proportional hazard analysis
- Conclusions

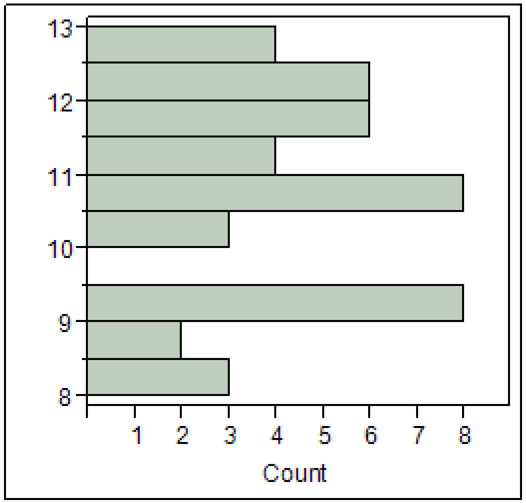
Definitions

- Survival analysis examines and models the time it takes for events to occur, termed survival time.
- The Cox proportional-hazards regression model is the most common tool for studying the dependency of survival time on predictor variables.
- Another representation of the distribution of survival times is the hazard function, which assesses the instantaneous risk of demise at time t, conditional on survival to that time
- Cox model leaves baseline hazard function unspecified yet a partial likelihood method can estimate it because taking ratios cancels it.
- $h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$ where β_k are parameters and x_i observations. With many observations one take ratios and get adequate approximations of the β_k .
- Spoiler: I was unable to find useful parameters among the C100 fault PVs. Values in covariate matrices were high only for variables which summed others or outliers which occurred only a few times. Sum variables are not orthogonal to components, of course.
- Everything except last bullet from http://socserv.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Cox-Regression.pdf

1L187 histogram and survival

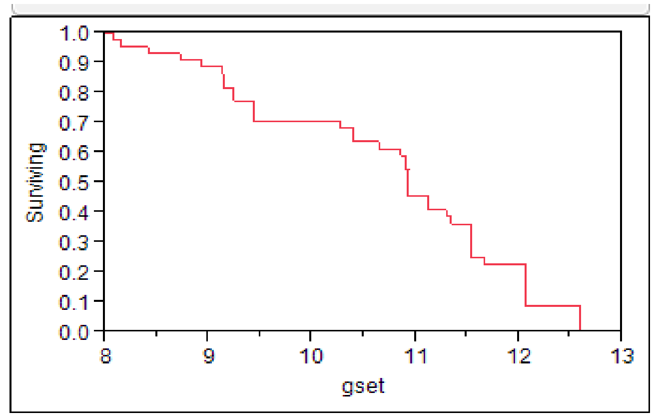
Distributions

gset



Quantiles

100.0%	maximum	12.59
99.5%		12.59
97.5%		12.59
90.0%		12.325
75.0%	quartile	11.63
50.0%	median	10.91
25.0%	quartile	9.44
10.0%		8.825
2.5%		8.08
0.5%		8.07
0.0%	minimum	8.07



The area above the survival curve is the integral of the histogram of number of faults versus gradient, normalized to one

Time to event:gset

Quantiles

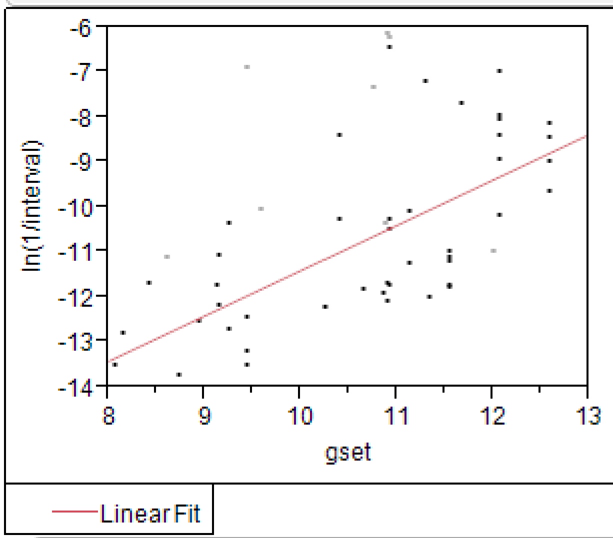
Group	Median Time	Lower 95%	Upper 95%	25% Failures	75% Failures
Combined	10.91	10.39	11.54	9.44	11.6

Combined

gset	Survival	Failure	SurvStdErr	Number failed	Number censored	At Risk
0.0000	1.0000	0.0000	0.0000	0	0	44
8.0700	0.9773	0.0227	0.0225	1	0	44
8.1500	0.9545	0.0455	0.0314	1	0	43
8.4100	0.9318	0.0682	0.0380	1	0	42
8.7200	0.9091	0.0909	0.0433	1	0	41
8.9300	0.8864	0.1136	0.0478	1	0	40
9.1300	0.8636	0.1364	0.0517	1	0	39
9.1400	0.8182	0.1818	0.0581	2	0	38
9.2400	0.7727	0.2273	0.0632	2	0	36
9.4400	0.7045	0.2955	0.0688	3	0	34
10.2600	0.6818	0.3182	0.0702	1	0	31
10.3900	0.6364	0.3636	0.0725	2	0	30
10.6400	0.6136	0.3864	0.0734	1	0	28
10.8500	0.5909	0.4091	0.0741	1	0	27
10.9000	0.5455	0.4545	0.0751	2	0	26
10.9100	0.4545	0.5455	0.0751	4	0	24
11.1200	0.4091	0.5909	0.0741	2	0	20
11.2900	0.3864	0.6136	0.0734	1	0	18
11.3400	0.3636	0.6364	0.0725	1	0	17
11.5400	0.2500	0.7500	0.0653	5	0	16
11.6600	0.2273	0.7727	0.0632	1	0	11
12.0600	0.0909	0.9091	0.0433	6	0	10
12.5900	0.0000	1.0000	0.0000	4	0	4

Fowler-Nordheim analysis 1L187

Bivariate Fit of ln(1/interval) By gset



Linear Fit

ln(1/interval) = -21.47284 + 1.0082026*gset

Summary of Fit

RSquare	0.461478
RSquare Adj	0.448656
Root Mean Square Error	1.442536
Mean of Response	-10.6711
Observations (or Sum Wgts)	44

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-21.47284	1.813595	-11.84	<.0001*
gset	1.0082026	0.168054	6.00	<.0001*

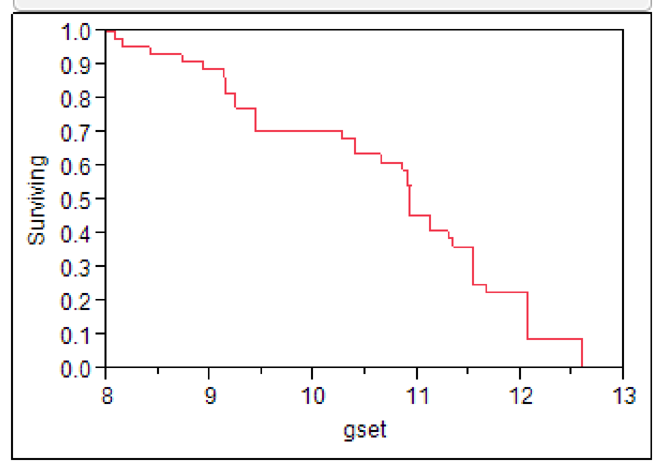
T ratio = Estimate/Std_Error

Gset predicted by linear fit for various intervals

8 hours (22800 s)	11.12 MV/m
1 day (86400 s)	10.03 MV/m
2 days (172800 s)	9.34 MV/m

First step in survival curve occurs ~9.5 MV/m because lem puts cavities around gset for two day interval

Survival Plot



Time to event: gset

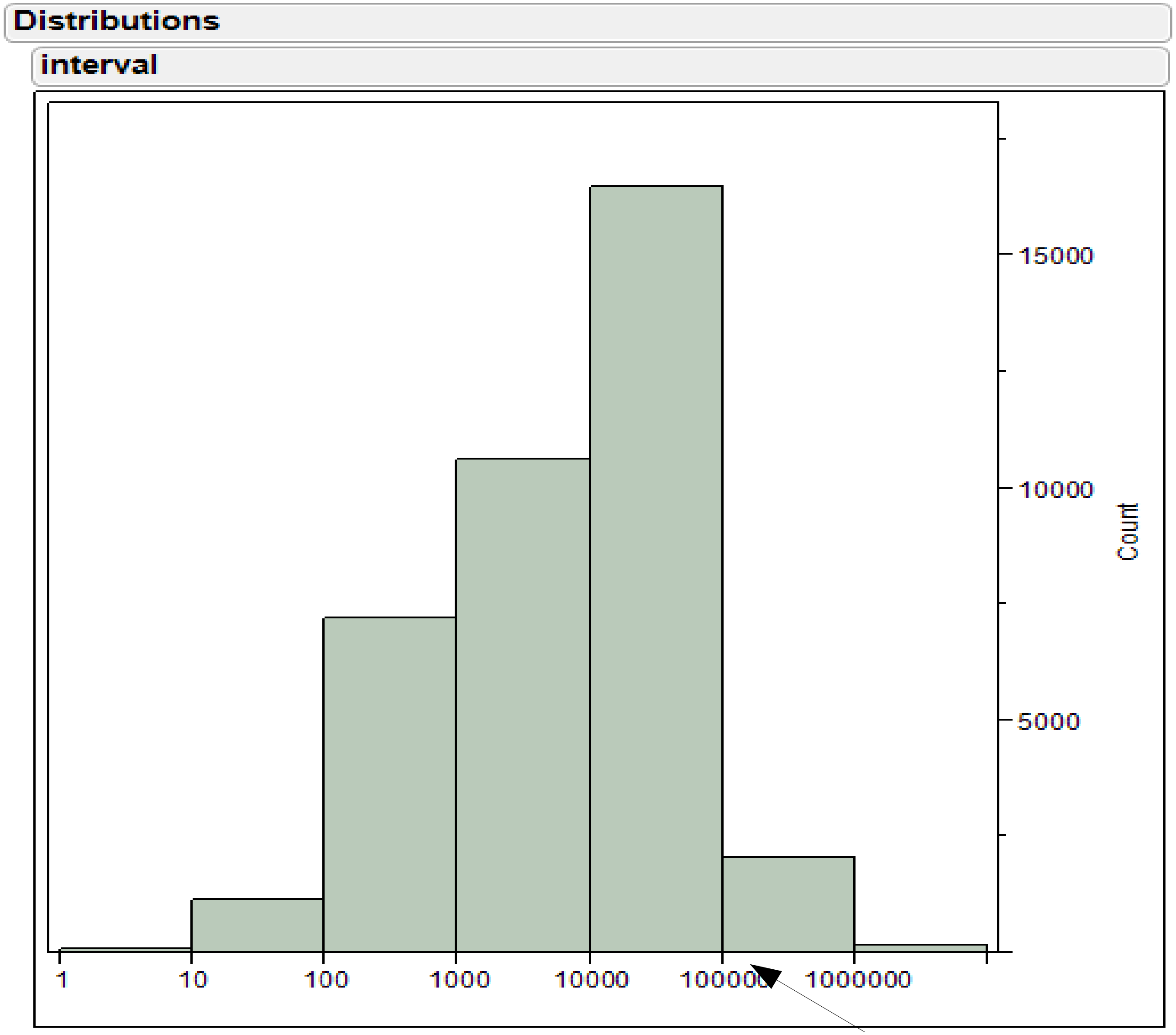
Quantiles

Group	Median Time	Lower 95%	Upper 95%	25% Failures	75% Failures
Combined	10.91	10.39	11.54	9.44	11.6

C100 data source

- Michele Joyce released the C100 Fault Logger Sept 25, 2016
- 109853 FCC and INTERLOCK faults recorded in linacs, 5021 in 0L04 through end of spring 2017 run. 67.6% simultaneous fault bit trips.
- JMP exploratory data analysis software (SAS) was used to view the raw data. Someone competent can move this work into R.
- Survival plots of intervals between faults can be generated by cavity for each GSET value. Exclusive of the energy lock cavities about 750 such plots could be produced. Energy lock cavities double the total.
- Proportional hazard analysis was used instead with GSET as the "time" variable. These plots show the fraction of faults which have occurred prior to/at GSET. Where there are many faults at the same GSET an estimator is used. JMP uses Breslow's https://en.wikipedia.org/wiki/Proportional_hazards_model

Distribution of event intervals

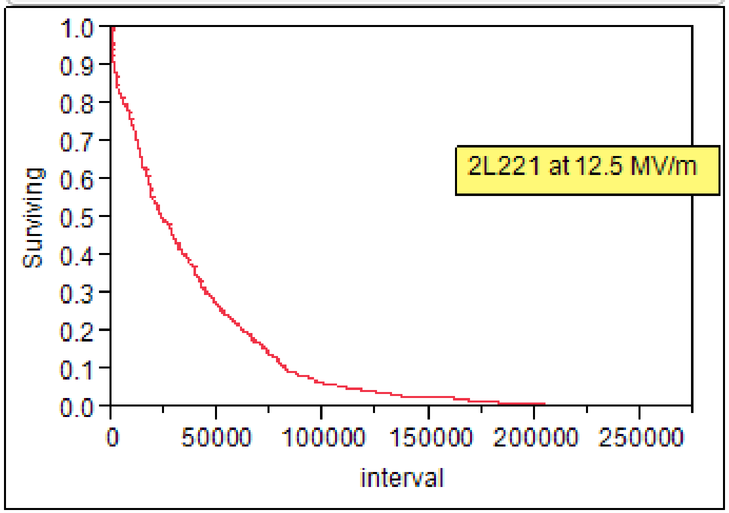


Two days = 172800 seconds

Survival plots at fixed GSET, 2L22

Product-Limit Survival Fit

Survival Plot



Time to event:interval

Summary

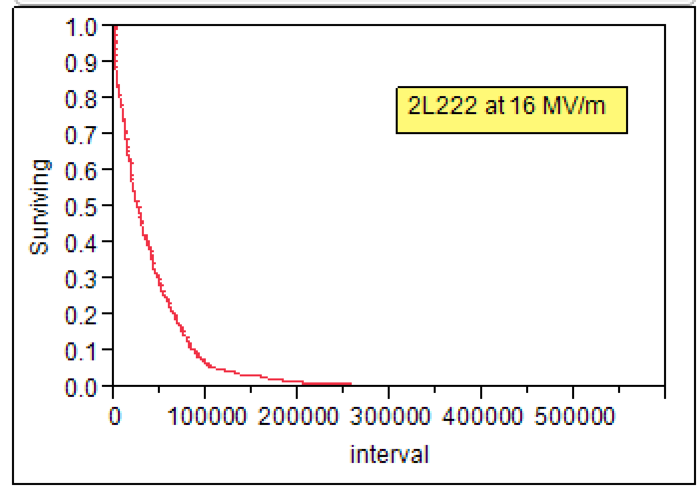
Group	Number failed	Number censored	Mean	Std Error
Combined	306	0	36854.5	2316.13

Quantiles

Group	Median Time	Lower 95%	Upper 95%	25% Failures
Combined	23027	18425	29094	9078

Product-Limit Survival Fit

Survival Plot



Time to event:interval

Summary

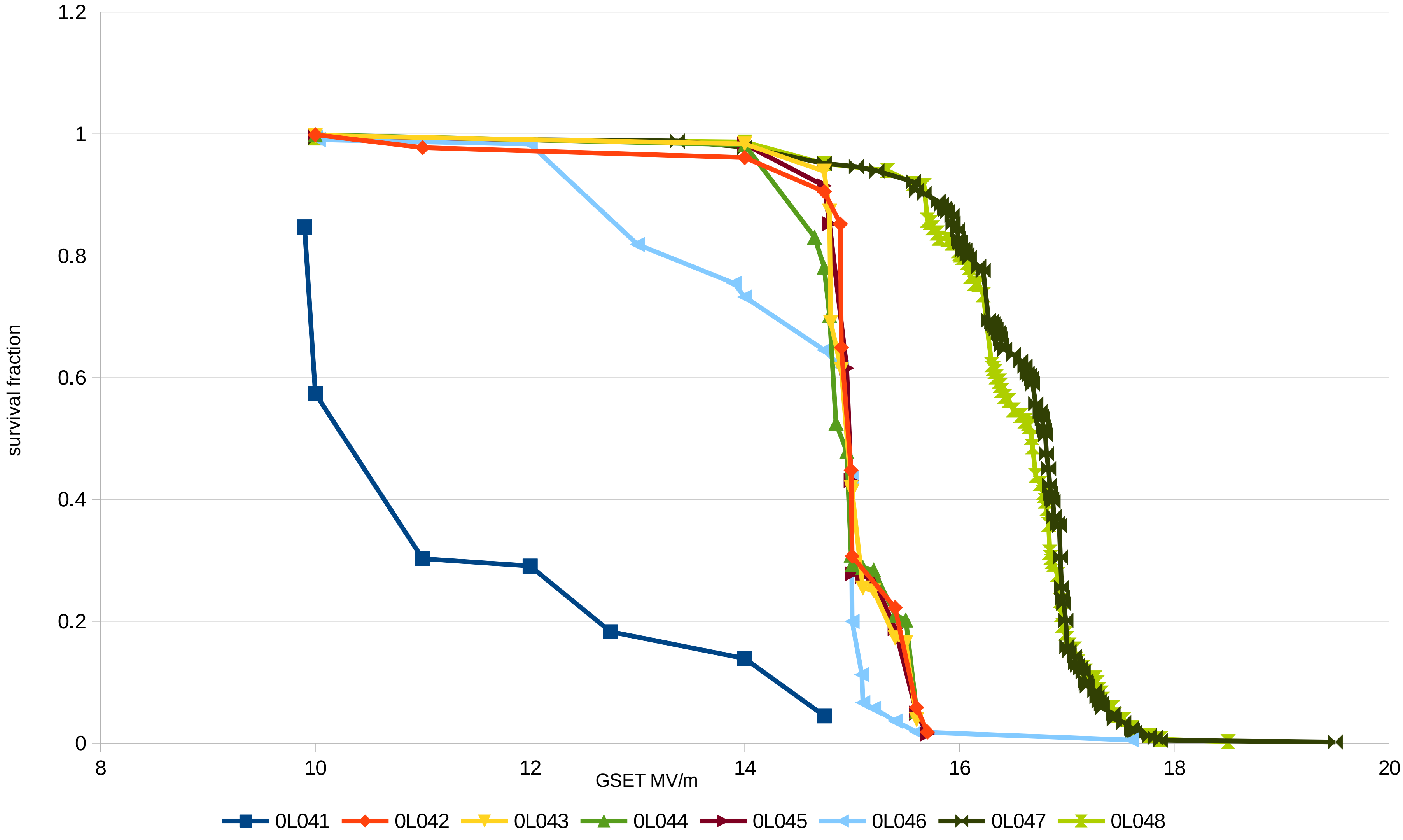
Group	Number failed	Number censored	Mean	Std Error
Combined	296	0	39967.2	3252.98

Quantiles

Group	Median Time	Lower 95%	Upper 95%	25% Failures
Combined	24787	18636	29663	8643.5

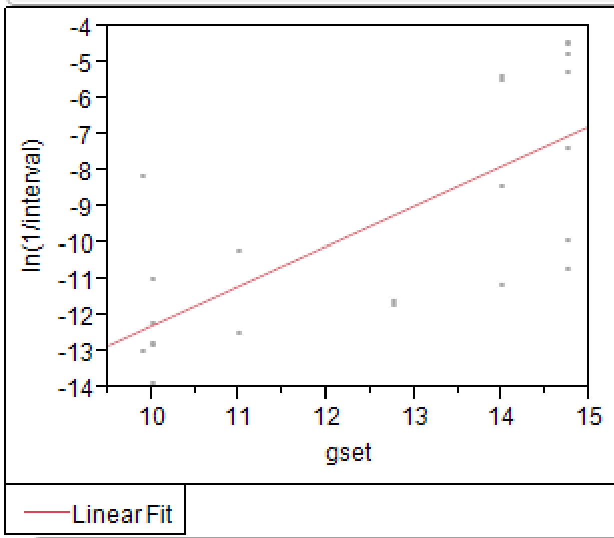
300,000 second interval results in one C100 trip per hour. All the 2L22 cavities show 25% of faults recorded with less than 3 hour interval.

0L04 survival plot – all faults recorded, many simultaneous



0L041 quench behavior

Bivariate Fit of ln(1/interval) By gset



Linear Fit

ln(1/interval) = -23.28434 + 1.1024248*gset

Summary of Fit

RSquare	0.520936
RSquare Adj	0.496983
Root Mean Square Error	2.284838
Mean of Response	-9.42986
Observations (or Sum Wgts)	22

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-23.28434	3.010515	-7.73	<.0001*
gset	1.1024248	0.236395	4.66	0.0001*

Plotting ln(1/interval) vs gset for the 22 quench faults recorded for 0L041 produces a fit which is similar to those found for C20 cavities where field emission charges the cold window. Here field emission may quench an end group or iris. Geng Rongli suggested field emission may liberate H2 until minimum in Paschen curve is reached, allowing discharge which FCC registers as quench. The gset predicted to produce a 300,000 second fault interval by the fit at left is 9.7 MV/m, comparable to the gradient at which the 0L041 curve on the previous slide drops off a cliff.

Similarly significant fits may be obtained from several linac C100 cavities but no others in 0L04. See TN-17-021, which covers all C100 zones.

Proportional Hazard Analysis - 0L041

Covariance of Estimates

Cov

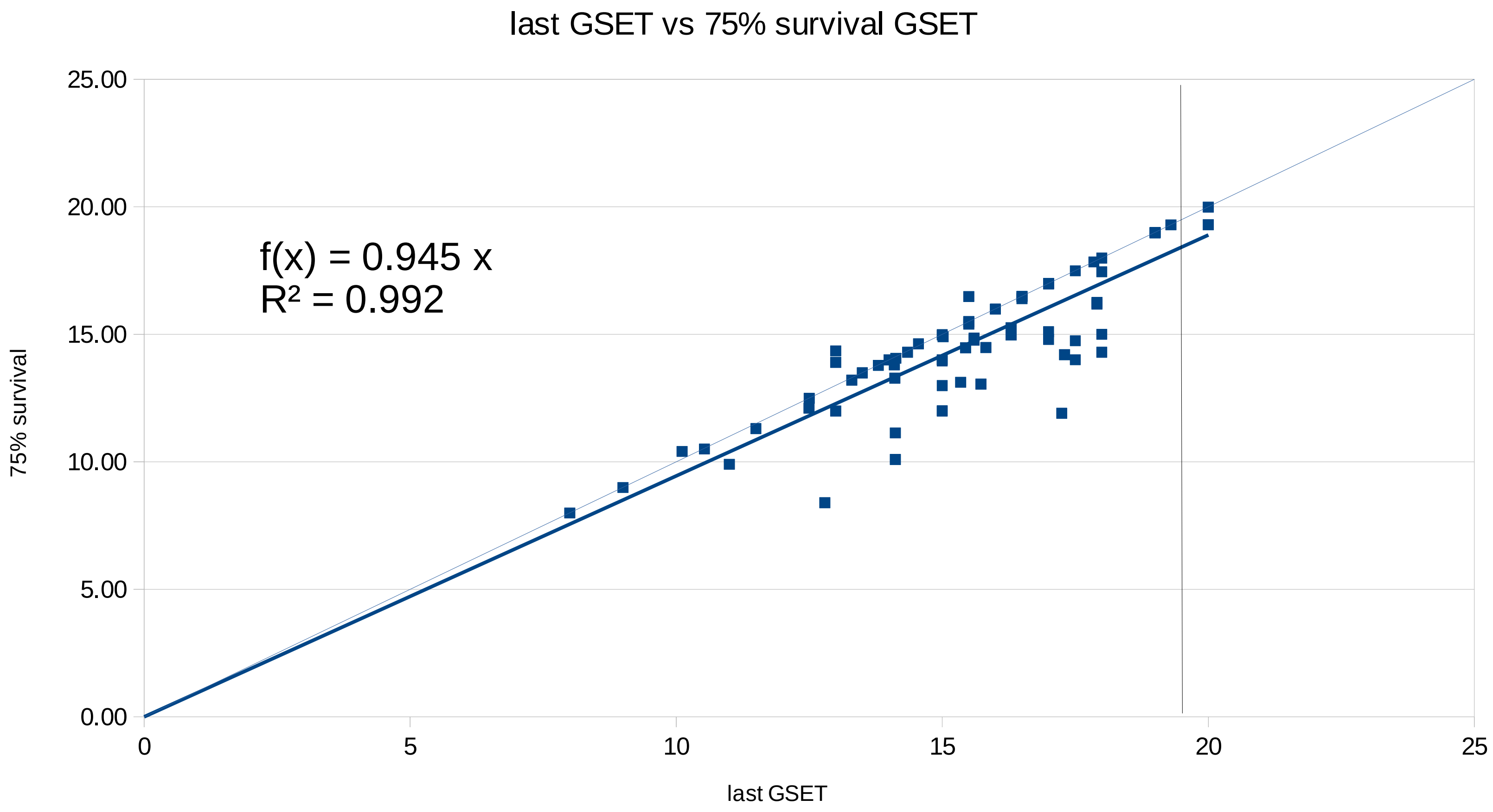
	fault[FCC.FSD	fault[FCC.GD(	fault[FCC.GLD	fault[FCC.GMI	fault[FCC.HPA	fault[FCC.HPA	fault[FCC.INTI	fault[FCC.PLD	fault[FCC.QN(	fault[FCC.RF	fault[FCC.SQN	fault[INTERL	fault[INTERLOCK.FSDI]
fault[FCC.FSD	0.0394	0.0128	-0.004	-0.054	0.0184	0.0166	0.0048	0.0139	0.0159	0.0183	-0.053	-0.053	0.0184
fault[FCC.GD(	0.0128	0.0995	-0.01	-0.059	0.014	0.0118	-0.001	0.0085	0.0118	0.0137	-0.056	-0.058	0.0129
fault[FCC.GLD	-0.004	-0.01	0.307	-0.076	-0.004	-0.006	-0.017	-0.008	-0.007	-0.004	-0.076	-0.075	-0.004
fault[FCC.GMI	-0.054	-0.059	-0.076	0.8878	-0.052	-0.053	-0.065	-0.056	-0.056	-0.053	-0.125	-0.122	-0.053
fault[FCC.HPA	0.0184	0.014	-0.004	-0.052	0.028	0.0178	0.0053	0.0141	0.0178	0.0195	-0.05	-0.052	0.0185
fault[FCC.HPA	0.0166	0.0118	-0.006	-0.053	0.0178	0.0468	0.0043	0.0127	0.0152	0.0175	-0.053	-0.052	0.017
fault[FCC.INTI	0.0048	-0.001	-0.017	-0.065	0.0053	0.0043	0.1927	0.001	0.0023	0.0052	-0.066	-0.064	0.0054
fault[FCC.PLD	0.0139	0.0085	-0.008	-0.056	0.0141	0.0127	0.001	0.087	0.0113	0.0141	-0.057	-0.056	0.0143
fault[FCC.QN(	0.0159	0.0118	-0.007	-0.056	0.0178	0.0152	0.0023	0.0113	0.0595	0.0172	-0.051	-0.055	0.0157
fault[FCC.RF	0.0183	0.0137	-0.004	-0.053	0.0195	0.0175	0.0052	0.0141	0.0172	0.0304	-0.051	-0.052	0.0185
fault[FCC.SQN	-0.053	-0.056	-0.076	-0.125	-0.05	-0.053	-0.066	-0.057	-0.051	-0.051	0.8822	-0.124	-0.053
fault[INTERL	-0.053	-0.058	-0.075	-0.122	-0.052	-0.052	-0.064	-0.056	-0.055	-0.052	-0.124	0.8796	-0.052
fault[INTERLC	0.0184	0.0129	-0.004	-0.053	0.0185	0.017	0.0054	0.0143	0.0157	0.0185	-0.053	-0.052	0.036

(Above) Covariance estimates of variables (betas of page 2) generated by JMP for 0L041 fault data. Nothing I found useful, as spoiler indicated. Parameter estimates themselves are below. *Even the diagonal elements above are non-unity, perhaps due to the $h_0(t)$ term being ignored or summed variables included. Analysis of just the quench faults, 7/15 of total time stamps, proved more useful as discussed above and in TN-17-021.*

Parameter Estimates

Term	Estimate	Std Error	Lower 95%	Upper 95%
fault[FCC.FSDO]	-0.274661	0.1984566	-0.655542	0.1351601
fault[FCC.GDCL]	-0.6260659	0.3155144	-1.288139	-0.033354
fault[FCC.GLDE]	-0.1086491	0.5540427	-1.423423	0.8294858
fault[FCC.GMES]	1.6723716	0.9422213	-0.999551	3.0962824
fault[FCC.HPAEN]	-0.1324424	0.1671923	-0.440085	0.2293554
fault[FCC.HPAO]	0.33449083	0.2163872	-0.088786	0.7721187
fault[FCC.INTLKI]	0.25364783	0.4390107	-0.731316	1.0305967
fault[FCC.PLDE]	0.34877125	0.2949539	-0.263067	0.9092766
fault[FCC.QNCH]	-0.6734251	0.243982	-1.160713	-0.191737
fault[FCC.RFSWOP]	-0.1655762	0.174488	-0.490467	0.2072391
fault[FCC.SQNCH]	-1.2601846	0.9392814	-3.929331	0.1538105
fault[INTERLOCK.CWWTW]	0.46082596	0.9378515	-2.206953	1.8702061
fault[INTERLOCK.FSDI]	-0.0827511	0.1897322	-0.443097	0.3136063

Last GSET (3/9/17) vs 75% survival GSET



Dark line and equation are fit forced through origin. Lighter line is x=y, showing that very few cavities are likely to be capable of more than Clyde's GSET values. Average of 80 linac C100s: **15.1** MV/m (includes three zeroes as cavities were off) C100 spec 19.3 MV/m including 10% headroom for cavities off, see TN-05-044. 17.5 MV/m with all cavities on. Three were off in FY17.

Conclusions

- Something like the charge accumulation on the cold windows of original modules is occurring in the C100s.
- Rongli's hypothesis that H2 is being liberated by field emission heating and a discharge eventually occurs seems reasonable to me. Temperature data needed.
- There are temperature diodes on only two FPCs, cavities 2 and 4. Cavities 4 and 5 have diodes on tube between FPCs. These four should be connected to DAQ and archived so we can see if there is a correlation to quench faults. One diode on each HOM pair should also be archived in case FE heating at that end directly causes quench faults, not H2 liberation.
- Where a field emission effect exists, either cold window charging or C100 "quench", logarithmic fit should be used in lem to minimize fault rate
- Where field emission effects are not obvious, as in most C50 and C100 cavities, GSET values at which significant drops in survival occur are a good starting point for Ops drvh (ODVH). 75% survival is a good place to start.

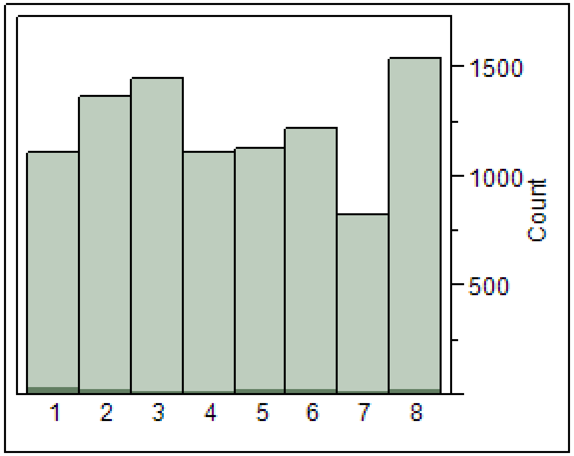
Backup

TN-17-021 Conclusions

- At one C100 fault per hour total
  - NL can deliver ~400 MeV vs 540 MeV specification
  - SL can deliver ~350 MeV vs 540 MeV specification
  - 0L04 can deliver ~76 MeV, 10 MeV less than needed for 123 MeV injector (aka 12 GeV to Hall D)

Quenches by location

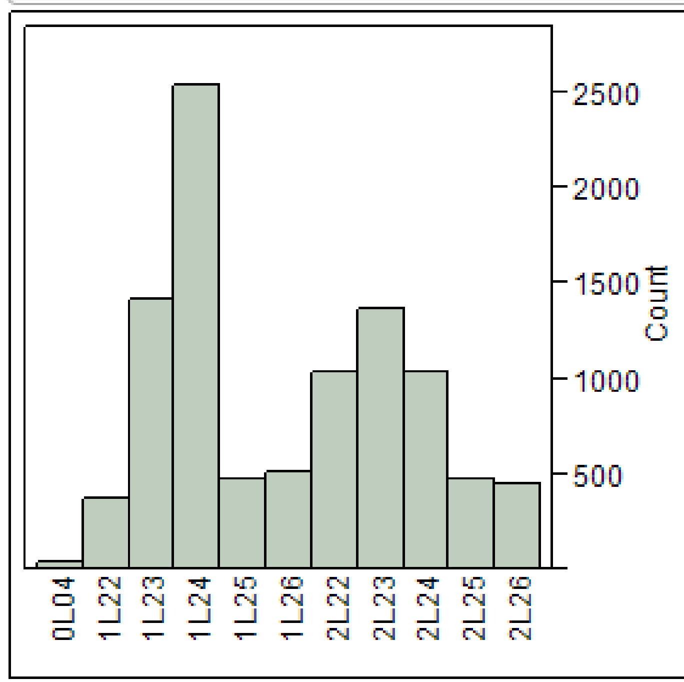
cavity_number



Frequencies

Level	Count	Prob
1	1102	0.11391
2	1355	0.14007
3	1441	0.14896
4	1104	0.11412
5	1115	0.11526
6	1207	0.12477
7	817	0.08445
8	1533	0.15847

module



Frequencies

Level	Count	Prob
0L04	22	0.00230
1L22	354	0.03707
1L23	1399	0.14651
1L24	2529	0.26484
1L25	461	0.04828
1L26	495	0.05184
2L22	1018	0.10661
2L23	1351	0.14148
2L24	1019	0.10671
2L25	461	0.04828
2L26	440	0.04608
Total	9549	1.00000

N Missing 125

11 Levels