SoRel: A tool for reliability growth analysis and prediction from statistical failure data

Karama Kanoun, Mohamed Kaâniche, Jean-Claude Laprie, Sylvain Metge

SoRel: A Tool For Reliability Growth Analysis and Prediction From Statistical Failure Data*

K. Kanoun M. Kaâniche J.-C. Laprie S. Metge**
LAAS - CNRS
7, Avenue du Colonel Roche - 31077 Toulouse Cedex (France)

Abstract*

*his paper presents a tool for **So**ftware (and lware) **Rel**iability analysis and evaluation: SoRel. tool implements a global method for reliability follow nd evaluation in presence of reliability growth due to gn fault removal. SoRel is composed of two parts wing respectively application of trend tests and bility growth models. The paper presents the method, rend tests and reliability growth models implemented oRel, shows how they can help during the validation ess and describes some functionalities of the tool. main features of the demonstration are outlined.*

oduction**

he objectives of software reliability evaluation in ence of reliability growth are numerous and are ely related to the point of view adopted (the supplier ie customer), and the life-cycle phase concerned. The lier is interested in the management of the validation maintenance activities whereas the customer is more erned by the reliability of the resulting product in ational life. **SoRel** — which is a tool for **So**ftware hardware) **Rel**iability analysis and prediction — s in achieving these objectives thanks to the bined use of trend tests and of reliability growth els. It provides qualitative and quantitative elements erning, for instance, a) the evolution of the reliability sponse to the debugging effort, b) the estimation of number of failures for the following periods of time so plan the test effort and the numerical importance of est and/or maintenance team and c) the prediction of reliability measures such as the mean time to failure, t failure rate or the failure intensity.

SoRel is based on a global method for softw reliability follow up and evaluation that has be developed at LAAS and applied to several real-l systems (see e.g., [7] or [8]). This method relies qualitative and quantitative analyses, it is intended better define the users' real needs in the field of softw reliability. It is briefly reviewed in the paper.

The paper is divided into three sections. Section outlines the method implemented by SoRel, gives general overview of the tool and shows the type of resu that can be obtained from SoRel. Section 2 is devoted the description of the tool. Section 3 describes t demonstration.

1. SoRel general presentation

SoRel is composed of two modules allowi respectively the application of **trend tests** and **reliabil growth models**. It is able to operate on two types failure data a) inter-failure times and b) number failures per unit of time (i.e., failure intensity), allowi application of two types of reliability growth mode respectively time domain and interval domain as called [15]. Figure 1 gives the organization of SoRel as it is se by the user. The reliability evolution is analysed throu trend test application. Selection of the model to be appli is based on the result of trend tests and the objectives the analysis. The remaining part of this section prese the trend tests and the reliability growth mod implemented in SoRel.

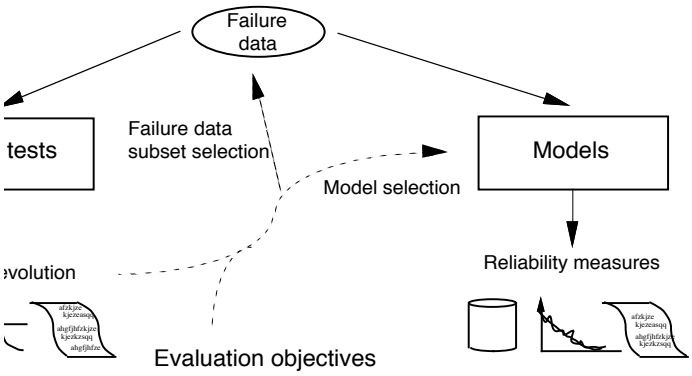


Figure 1: SoRel organization

Trend tests

'wo reliability trend tests are available: the arithmetic n and the Laplace test—for both inter-failure times failure intensity data.

l. Arithmetical mean test

'his test consists of calculating τ_k, the arithmetic n of the first k inter-failure times (resp. number of res per unit of time). When τ_k's form an increasing s (resp. decreasing), reliability growth can be ıced. This test is very simple and its interpretation is as it is directly related to the collected measure.

2. Laplace test

'he Laplace test, which is a statistical test, consists of ılating the Laplace factor, u, whose expressions are n in Figure 2. In practice, in the context of reliability vth, negative values of the Laplace factor suggest bility growth whereas positive values suggest bility decrease; values oscillating between -2 and +2 cate stable reliability.

'hese practical considerations are deduced from the ificance levels associated with the statistics, for ınce, for a significance level of 5% the null othesis "no trend against trend" is rejected for >1.96.

Inter-failure time data

$$u(k) = \frac{c - m}{s_k}\sqrt{12\,(k-1)} \qquad k = 2,\ldots, n$$

$$c = \frac{1}{k-1}\sum_{i=1}^{k-1} s_i$$

$$m = \frac{s_k}{2}$$

$$s_i = \sum_{j=1}^{i} t_j$$

t_j : time interval between failures (j-1) and j
s_i : instant of failure i occurrence
n : number of failures observed

$$u(k) = (c - m)\Big/\sqrt{\frac{k^2 - 1}{12\,y_k}} \qquad k = 2,\ldots, p$$

$$c = \sum_{i=1}^{k} (i-1)\, n_i \,/ y_k$$

$$m = \frac{k-1}{2}$$

$$y_k = \sum_{i=1}^{k} n_i$$

n_i : number of failures during time unit i
y_k : cumulative number of failures until time unit k
p : total number of units of time

Figure 2: Laplace factor expressions

The users of SoRel can utilize the Laplace test as conventional statistical test. However, in our approac we extended it to identify global and local trends [9]. the Laplace factor is evaluated using all the data collect up to the unit of time considered, it reflects the *glob* variation of reliability. *Local* fluctuations can be detect by studying the variation of u(k): for example, when u is positive and tends to decrease it suggests a decrease the number of failures observed over the consider period which means that, locally, reliability tends increase although a global decrease is observed. This summarized in Figure 3. Global reliability decrease ov A and B is due to data observed during period A, if t latter are not considered for evaluation purposes, period will display local and global reliability growth.

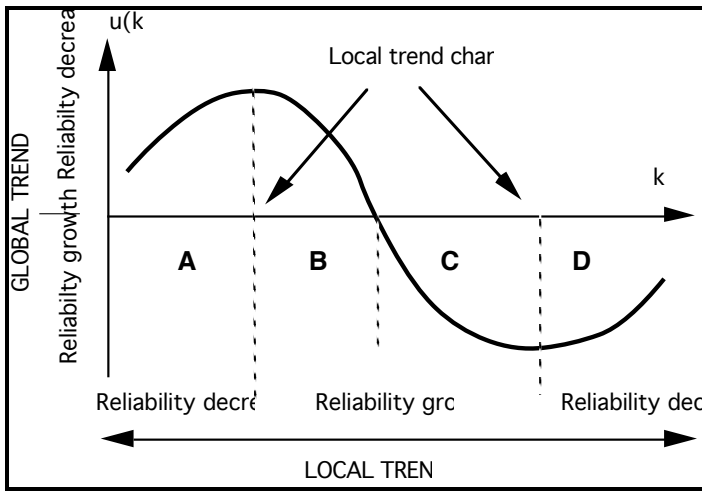


Figure 3: Laplace test interpretation

1.1.3. Practical use of trend tests

Trend analyses are of great help in appreciating t efficiency of test activities and controlling their progre They help considerably the software development follo up. Indeed, graphical tests are very often used in t industrial field [3, 14, 16], even though they are call differently, such as descriptive statistics or control cha

normal situation. Reliability decrease may also result
ı regression faults. If the duration of the period of
ease seems long, one has to pay attention and, in
e situations, if it keeps decreasing this can point out
e problems within the software: the analysis of the
ons of this decrease as well as the nature of the
ated faults is of prime importance in such situations.
ability growth after reliability decrease is usually
omed since it indicates that, after first faults removal,
corresponding validation activity reveals less and less
s. **Stable reliability** with almost no failures indicates
the related validation activity has reached a
uration": application of the associated test sets does
eveal new faults, or the corrective actions performed
of no perceptible effect on reliability. One has either
top testing or to introduce new sets of tests or to
eed to the next phase.
urthermore, trend analyses may be of great help for
bility growth models to give better predictions as
vn in the next section.

Reliability growth models

our reliability growth models are implemented: the
rexponential model (Kanoun-Laprie) [11], the
nential model (Goel-Okumoto) [4], the S-Shaped
el (Yamada et al.) [17] and the doubly stochastic
el (Littlewood-Verrall) [12]. The S-Shaped model
is an interval domain model, the doubly stochastic
el (DS) is time domain whereas the hyperexponential
el (HE) and the exponential model (EXP) are both
and interval domain.
hese models allow different kinds of behavior to be
eled: HE, EXP and DS model a decreasing failure
they yield better results when applied to data
laying reliability growth, that is, interval C of Figure
ıdeed, these models can be applied also to interval B-
Figure 3 discarding failure data pertaining to A. The
nodel is characterized by an increasing failure rate
wed by a decreasing failure rate, it produces good
lts when applied to failure data belonging to an
val such as A-B-C. For D, prediction is delayed until
rving an interval of reliability growth.
igure 4 summarizes the characteristics of these
els. Depending on the model, the main quantities that
be evaluated are: the mean time to next failure (or
TF), the failure intensity, the cumulative number of
res and the residual failure rate of the software. It is
h noting that HE is the only model allowing
uation of the residual failure rate in operation.
Model execution is carried out into two steps:
maximum likelihood or least square. Both of them nee
numerical optimization procedure to estimate
parameters of the models. For two-parameter mod
(EXP and SS), the numerical values are obtained via
Newton-Raphson iterative method whereas for thr
parameter models (HE and DS) they are evaluated via
Powell numerical method [13].

SoRel enables the user a) to determine how well
selected model fits the data and b) to compare estimatic
issued from several models. The goodness-of-fit crite
are: a) the Kolmogorov-Smirnov statistics [1], b)
prequential likelihood [2] and c) the residue [6]. The fi
two criteria are evaluated only for inter-failure time da
The residue is evaluated for both failure data types, it
based on the difference between the observed measu
and its expectation from the model.

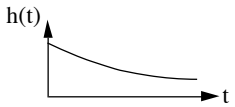
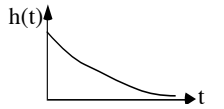
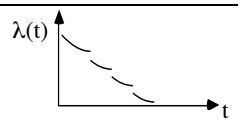
Model	h(t) or λ(t) shape
Hyperexponential $h(t) = \frac{\omega\zeta_{sup}e^{-\zeta_{sup}t}+\varpi\zeta_{inf}e^{-\zeta_{inf}t}}{\omega e^{-\zeta_{sup}t}+\varpi e^{-\zeta_{inf}t}}$	
Exponential $h(t) = N\,\phi\,\exp^{-\phi t}$	
S-Shaped $h(t) = N\,\phi^2\,t\,\exp^{-\phi t}$	
Double Stochastic $\lambda_i(t) = \frac{\alpha}{t+\psi(i)}$ $\psi(i) = \beta_1+\beta_2 i$	

Figure 4: Reliability growth models implemented in SoRel

A model can be analysed according to its *retrodict capability* and *predictive capability*. The retrodictive pability expresses model ability in reproducing the served behavior of the software. The predictive capabil reflects the model ability in predicting future behavior the software, from the observed failure data. Retrodicti and predictive capabilities are measured throu goodness-of-fit criteria.

1.3. SoRel within the software validation process and reliability evaluation

SoRel has been used to follow up and evaluate

or the validation phase, the main results concern the ution of reliability in response to debugging activities the prediction of the number of faults that will be ated over the next periods of time [5]. During ation, the objectives of reliability analysis are more ous, we give hereafter some examples illustrated ugh the results obtained for the first three systems ch are electronic switching systems (ESS). For the -B, the evaluation of the residual failure rate in ation carried out from failure data collected on the ware in operation [6] allowed the dependability of the le ESS to be evaluated (accounting for hardware and ware). For the TROPICO-R ESSs we have followed complementary approaches [8, 10]:

- from the supplier point of view, estimation of the maintenance effort to provide in operation in order to satisfy the correction reports issued from the various customers,
- from the customer point of view, estimation of the residual failure rate in operation in order to evaluate the impact of software reliability on the whole ESS reliability.

System	Languages	Volume	Observation	Phases	# Systems	# FR and/or CR
E10-B	Assembler	100 k-bytes	3 years	Val. / Op.	1400	58 FR / 136 CR
TROPICO-R 1500	Assembler	300 k-bytes	27 months	Val. / Op.	15	461 CR
TROPICO-R 4096	Assembler	350 k-bytes	32 months	Val. / Op.	42	227 CR
Telecommunication Equipment	PLM-86	5 10^5 inst.	16 months	Val.	4	2150 FR
Work station	various	--	4 years	Op.	1	414 FR

FR: Failure Report CR : Correcrion Report Val.: Validation Op.: Operation

Table 1: Characteristics of some real-life software systems studied by SoRel

oRel Description

oRel runs on Macintosh II-xx computer equipment an arithmetical co-processor. The human / machine face has been denoted special attention. It is active: it is menu-driven and uses the multiple low management facilities of the Macintosh. The gram is modular and new reliability growth tests and els can easily be added. It is written in Pascal (5.000 of code) and requires about 200 K bytes of memory. ser guide and a tutorial are available. The user guide ains how to use the tool and provides example ions as well as samples of input and output files. The rial presents the method, the trend tests and the

SoRel is composed of two modules "**TREND**" a "**MODELS**" corresponding respectively to trend analy and model application The two modules accept the sa input data files which can be created and changed by word processing or graphic editor. Numerical results a displayed immediately on the screen during the executi process. Additionally, the corresponding curves can plotted upon user's request. The results are also record in the form of ASCII files that can serve as input to otl Macintosh applications (such as Excel) allowing instance comparison of results issued from differ model applications.

The main menu commands allow cancellation of t last changes (Resume) and exit from the program (Qu The features specific to each module are described in t rest of the section.

2.1. "TREND" module

Selection of the trend test to be applied is achieved indicated in Figure 5. SoRel prompts the user for the da set type as shown in Figure 6 and then for the input da file name (Figure 7). The user can either a) choose the f name among the available data set names in the sa folder (if not, s/he enters directly the name of the f input. (Figure 8).

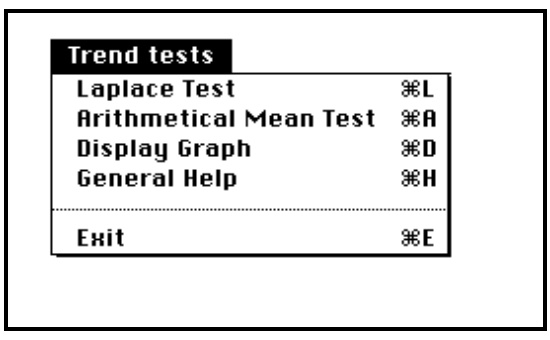


Figure 5: Trend test selection

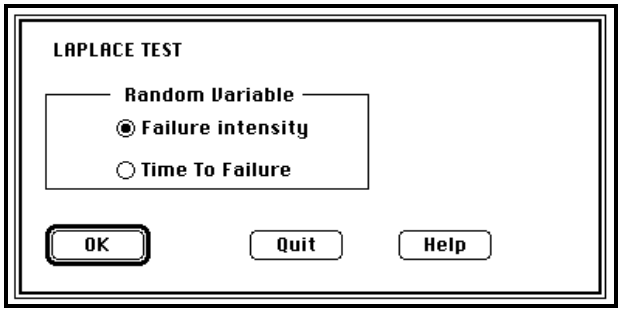


Figure 6: Selection of data type

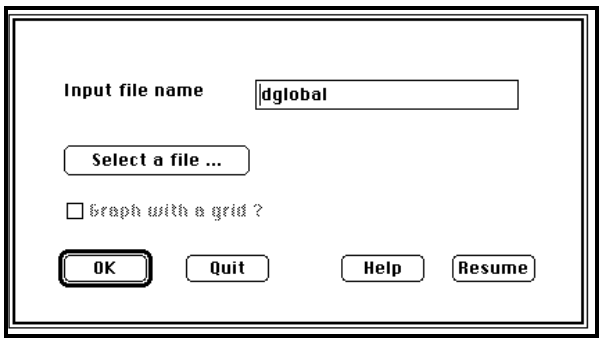


Figure 7: Input file name entry

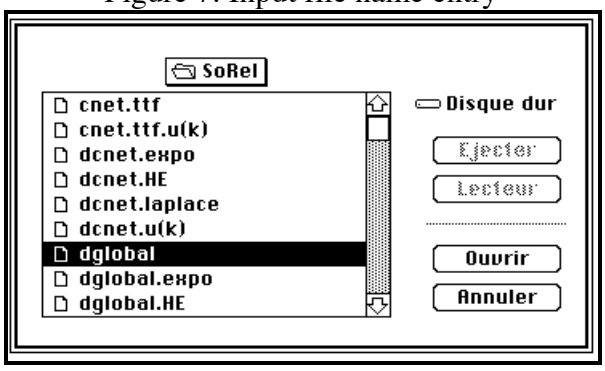


Figure 8: Selection of the input file name

he trend test may be applied to sub-sets of the data rded in the selected file in order to highlight the local l: the user indicates the rank of the first data item to onsidered (Figure 9). Figure 10 gives an example of hical results.

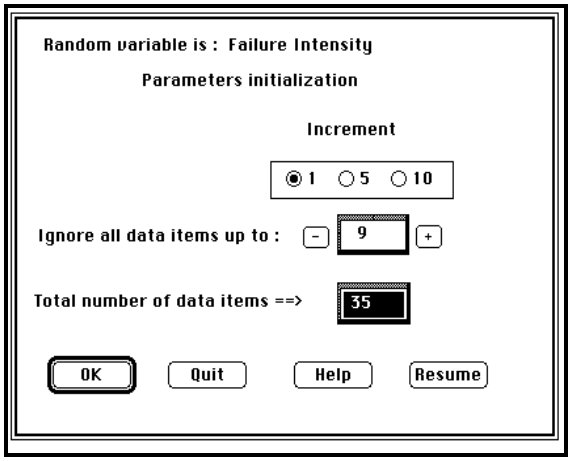


Figure 9: Definition of the failure data sub-set

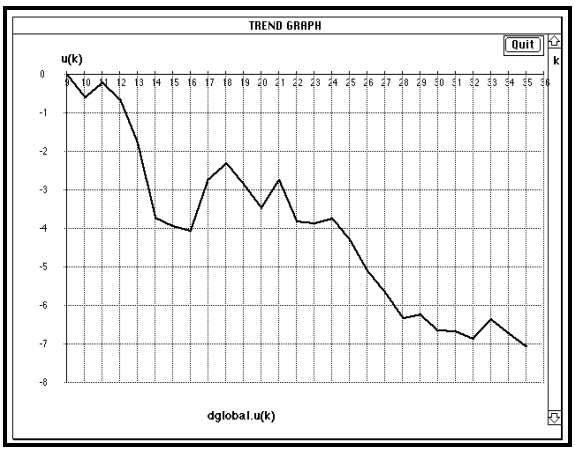


Figure 10: Graphical results for trend tests

2.2. "MODELS" module

The user selects the model (Figure 11) according to t trend displayed by the data set and to the evaluati objectives. The user has to indicate the input data type well as the measure to be evaluated (Figure 12), then he prompted for all the input needed to apply the select model as indicated in Figure 13.

The dialogue areas are defined as follows. Area **I** for initial Parameter Setting, is needed when the mo parameters are evaluated via the Powell method [13]. T user has to supply initial approximations of t parameters at the optimum. Area **DP**, for Data Partiti allows the user to define a) the data sub-set from whi the parameters will be evaluated and b) the predicti interval. Another required input is the use of a window not for model Calibration, area **C**. Finally area **V** poi out the interval over which the Validation criteria will evaluated (this interval may or may not correspond to t prediction interval). The **Set Options** command allo the user to adjust some parameters of the optimizati procedure, such as the maximum number of iterations a the convergence criteria which have been given fix values by default.

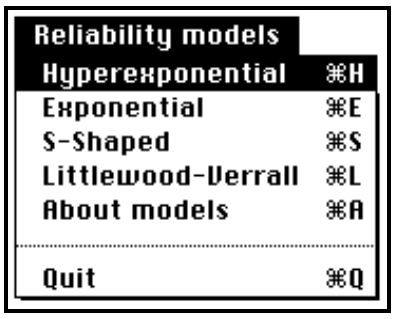
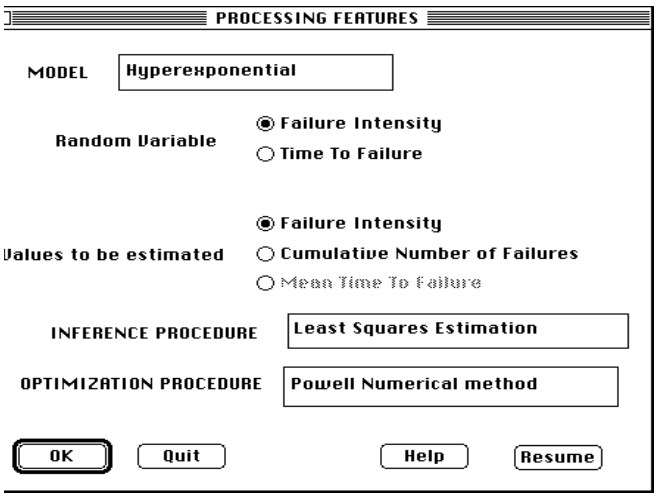


Figure 11: Model selection



ure 12: Input data type and output measure selection

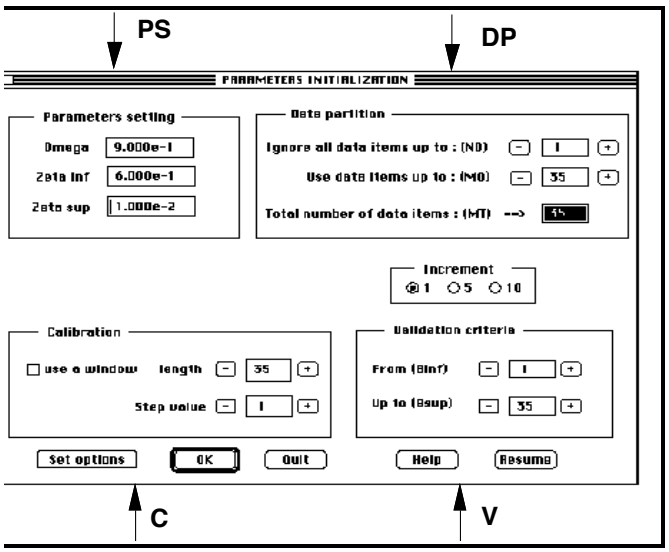


Figure 13: Model application initialization

igure 14 gives an example of numerical results reas graphical results are displayed on user's request ure 15). The goodness-of-fit criteria are given with numerical results.

he demonstration

oRel will be demonstrated using failure data ected on a real-life software system: the TROPICO-R 5 ESS.Data have been collected over about two years uding the end of validation and the beginning of ational life. The two types of input data files (inter-re times and failure intensity) will be addressed. e the results are displayed immediately, it is possible ake several executions to show the main features of ool.

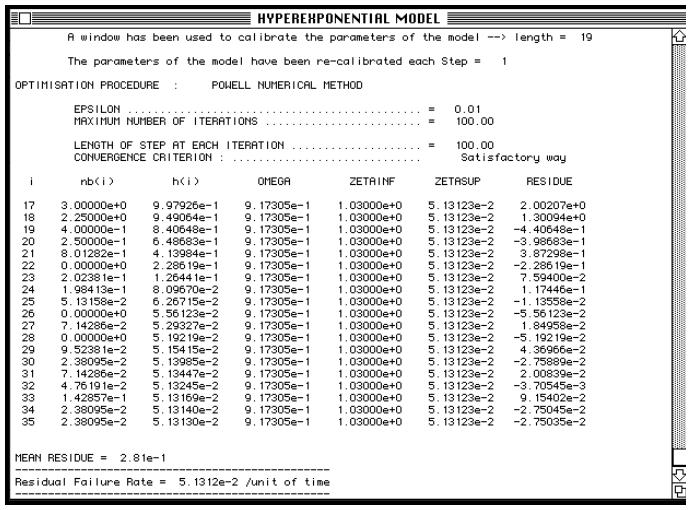


Figure 14: Numerical results

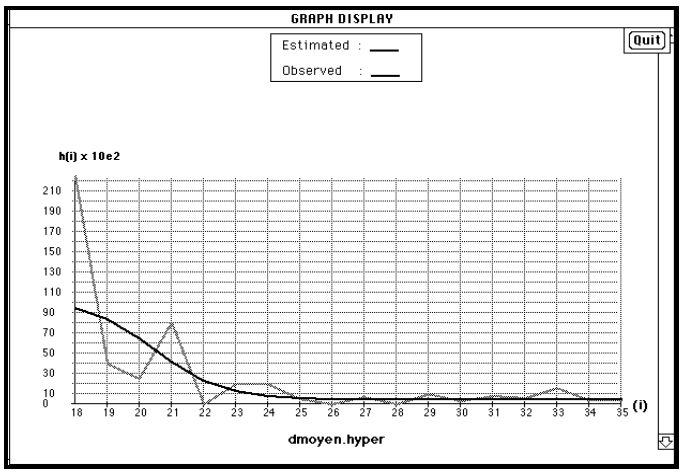


Figure 15: Graphical results

The input data files will be displayed and comment Trend tests will be applied to the considered data se Emphasis will be put on the conclusions that can drawn from trend test application and the results will commented showing the link between the reliability tre change indicated by SoRel and the different phases of t software life-cycle. We will also show how to use tre results before reliability growth model application.

Reliability growth models will then be applied to su sets of these input data files with and without accounti for trend results in order to show the improvement of t estimations. We will first carry out retrodicti evaluations and show the capability of the models reproduce the observed behavior of the softwa Application of the models in a predictive way will then carried out. The two types of failure data will be us during the demonstration allowing various reliabil measures to be evaluated, i.e., MTTF, cumulative numb of failure, failure intensity and failure rate. We will a compare results issued from application of vario models to the same data sets.

eagues, in particular Yves Crouzet and the late and ly-missed Christian Béounes.

erences

D.R.Cox, P.A.W.Lewis, *The statistical analysis of series of events*, London, Chapman & Hall, 1966.

A.P.Dawid, "Statistical theory: The prequential approach", *J. Roy. Statist. Soc. A*, vol 147, 1984, pp. 274-292.

R.B.Grady, D.R.Caswell, *Software metrics: establishing a company-wide program*, Hewlett Packard Company, Prentice Hall, Inc., 1987.

A.L.Goel, K.Okumoto, "Time Dependent Error-Detection Rate Model for Software and other Performance Measures", *IEEE Trans. on Reliability*, vol R-28, no. 3, 1979, pp. 206-211.

M.Kaâniche, K.Kanoun, S.Metge, "Failure analysis and validation monitoring of a telecommunication equipment software system", *Annales des Télécommunications*, Vol. 45, no. 11-12, 1990, pp. 657-670, *in French*.

K.Kanoun, T.Sabourin, "Software dependability of a telephone switching system", Proc. 17th *IEEE Int. Symp. on Fault-Tolerant Comp. (FTCS-17),* Pittsburgh, Pennsylvania, 1987, pp. 236-241.

K.Kanoun, J.-C.Laprie, T.Sabourin, "A method for software reliability growth analysis and assessment", Proc. *Software Engineering & its Applications*, Toulouse, France, 1988, pp. 859-878.

K.Kanoun, M.Bastos Martini, J.Moreira De Souza, "A method for software reliability analysis and prediction — application to the TROPICO-R switching System", *IEEE Transactions on Software Engineering,* Vol. 17, no. 4, 1991, pp. 334-344.

K.Kanoun, J.-C.Laprie, "The role of trend analysis in software development and validation", Proc. *SAFECOMP'91*, Trondheim, Norway, 1991, pp. 169-174.

K.Kanoun, M.Kaâniche, J.-C.Laprie, S.Metge, "SoRel: a tool for software reliability analysis and evaluation from statistical failure data", Proc. *8th Symposium on Reliability and Maintainability*, Grenoble, France, 1992, pp. 422-431, *in French*.

J.-C.Laprie, K.Kanoun, C.Béounes, M.Kaâniche, "The KAT — (Knowledge-Action-Transformation) — Approach to the modeling and evaluation of reliability and availability growth", *IEEE Trans. on Software Engineering*, Vol. 17, no. 4, 1991 pp. 370-382.

B.Littlewood, J.L.Verrall, "A Bayesian Reliability Growth Model for Computer Software" J. Royal Stat. Soc., C(App. stat.), 22, 1973, pp. 332-336.

S.S.Rao, *Optimization Theory and Applications*, Wiley Eastern Limited, 1978..

N.Ross, "The collection and use of data for monitoring software projects", *Measurement for software control and assurance*, Edited by B.A.Kitchenham and B.Littlewood, Elsevier Applied Science, London and

[16] V.Valette, "An environment for software reliabil evaluation", Proc. *Software Engineering & Applications*, Toulouse, France, 1988, pp. 879-897.

[17] S.Yamada, M.Ohba, S.Osaki, "S-Shaped Reliabil Growth Modeling for Software error Detection", IE *Trans. on Rel.* vol. R-32, no. 5, 1983, pp. 475-478