More Pitfalls of Accelerated Tests

William Q. Meeker, Iowa State University, Georgios Sarakakis, Tesla Motors, Athanasios Gerokostopoulos, ReliaSoft Corporation

Abstract—As product development cycles become shorter, and companies demand more rapid achievement of reliability goals, it is becoming more and more important to use quantitative Accelerated Life Tests (ALT) to predict and improve reliability. Today there is an abundance of methods to plan and analyze accelerated tests, but there are also many pitfalls. This paper identifies some major problems and concerns in conducting and interpreting the results of accelerated tests. We identify various pitfalls, such as: using an equal unit allocation at all levels of an accelerating variable, using unnecessarily complex testing and data analysis schemes, attempting to predict life from a HALT, using ALT at the system level, not having adequate time-to-failure information, using extreme extrapolation, ignoring the impact of idle time in use-rate acceleration tests, and not understanding interactions between accelerating variables.

Key Words—Accelerated degradation, Acceleration factor, activation energy, Arrhenius, extrapolation, life-stress relationship, quantitative accelerated life testing.

1 INTRODUCTION

The continuing push for rapid product development and greater reliability is placing new demands on the utility of accelerated tests, as described in Meeker and Escobar (1993). The material presented here is an extension of “Pitfalls in Accelerated Testing” paper by Meeker and Escobar (1998a). Since 1998, accelerated testing has seen increased popularity and there is an improved ability to set up and conduct accelerated tests, and analyze accelerated test data. In our collective experiences, however, we have identified some additional problems and concerns. The additional pitfalls outlined here provide a more complete set to help guide those who are planning, conducting and interpreting the results of accelerated tests. We also extend the discussion of some of the previously identified pitfalls.

The original paper by Meeker and Escobar (1998a) discussed the following nine pitfalls:

1) Multiple (unrecognized) failure modes
2) Failure to properly quantify statistical uncertainty
3) Multiple time-scales and multiple factors affecting degradation
4) Masked failure modes
5) Faulty comparison
6) Accelerating variables can cause deceleration!
7) Beware of untested design/production changes
8) Beware of drawing conclusions on the basis of specially built prototype test units
9) It is difficult to use accelerated life tests to predict field reliability

For more details on the above nine pitfalls, please refer to Meeker and Escobar (1998a). Here we present additional pitfalls that we have encountered through consulting projects, interactions with reliability engineering teams in various product development environments, in the literature, and in the development and support of accelerated life testing software. We present case studies from various industries to illustrate our key points.

All of the pitfalls (in Meeker and Escobar 1998a and in this paper) are either caused by statistical misconceptions or by the naïve application of accelerated test methods. The pitfalls in this paper are grouped into the following categories:

• Pitfalls that occur during the planning of an accelerated test
• Pitfalls that occur during the execution of an accelerated test
• Pitfalls that occur during the analysis and interpretation of accelerated test data

Traditionally, most accelerated tests have been accelerated life tests (ALTs) in which the response is time to failure (or survival times for those units that do not fail in the test). In many applications, however, engineers conduct accelerated degradation tests (ADTs) that result in “degradation data” where the response is some measure of chemical degradation (e.g., concentration), physical degradation (e.g., wear on a tire), or performance degradation (e.g., light output of an LED or power output of an amplifier). Accelerated repeated measures degradation tests arise when one can monitor or measure degradation of test units over time (see Meeker, Escobar, and Lu (1998) or Chapter 21 of Meeker and Escobar (1998b) for examples). Accelerated destructive degradation tests arise when test units need to be destroyed in order to obtain a degradation measurement (see Chapter 11 of Nelson 1990 and Escobar, Meeker, Kugler, and Kramer 2003 for examples). Because the differences among these types of accelerated tests arise primarily due to differences in what one can observe or measure, most of the pitfalls can arise in either ALTs or ADTs.

In the next three sections, we present the identified pitfalls in conducting and interpreting the results of accelerated tests. The sequence of the pitfalls within these three sections does not imply relative frequency or importance.

2 Pitfalls During ALT Planning

This section includes the pitfalls that are associated with the early stages of planning an accelerated test.

A. Use of equal unit allocation at all levels of the accelerating variable in an ALT

A common mistake when designing an ALT is to use a traditional test plan that allocates an equal number of units to each level of the accelerating variable. We have observed this several times, especially in reliability teams that are fairly new to accelerated testing and who have not carefully studied the statistical literature on the subject (e.g., Chapter 6 of Nelson (1990) and Nelson 2005a, b are valuable sources of information about ALT planning).

The probability of failure increases as the level of stress increases. For the same amount of test time, one would expect to observe early failures at the higher stress level when compared with the lower stress level. With an equal number of allocated units, the end result is usually many failed units at higher stress levels and few or even no failures at lower stress levels. Such an inappropriate test plan will result in an unnecessarily large amount of statistical uncertainty in the estimates when extrapolating back to use conditions. As stated in Meeker and Escobar (1993) “Typically one should allocate more units to the lower levels of stress. Intuitively, there are two reasons for this. First, because of the limited test time, there will be a smaller proportion of units failing at lower stress. Second, we are more interested in inferences at lower stresses.”

There is an important tradeoff involving the number of levels of an accelerating variable to use in an ALT. Test plans with more levels tend to be more robust than plans with fewer levels in the sense that good statistical properties predicted in test planning are less sensitive to departures from the assumed acceleration model or to misspecification of the planning information inputs.

Relatedly, using fewer test levels can be more efficient, statistically (i.e., result in narrower confidence intervals), but more risky because the assumed model is never exactly correct and because such tests are more sensitive to departures from the assumed model. The impact of the decrease in statistical efficiency due to increasing the number of levels can be evaluated and is generally not large, but adding additional levels can add cost due to the need for additional test chambers.

There are methods to evaluate and optimize accelerated life test plans, as described in Chapter 20 of Meeker and Escobar (1998b). ALT planning methods that use optimization have been implemented in ReliaSoft (2010), JMP (2012) and Meeker and Escobar (2010).

Evaluating and optimizing ALT plans requires knowledge of “planning information.” Generally this planning information consists of an assumption about the ALT model and its parameters. Such planning information is generally obtained from a combination of engineering judgment, previous experience with a similar product and knowledge of the physics of failure. Because the specification of the planning
information can have a big effect on the ALT planning, sensitivity analysis is recommended. Examples of such sensitivity analysis can be found in Meeker, Escobar and Zayac (2003) and Monroe et al. (2010).

Although equal allocation is poor practice in the planning of ALTs, equal allocation tests do not perform horribly for ADTs when, as is usually the case, no censoring is involved (because quantitative degradation readings are obtained on each test unit). See, for example, Meeker, Escobar, and Lu (1998) and Shi, Escobar, and Meeker (2009).

B. Choosing the wrong accelerating variable or having a test plan that will not provide useful information

The accelerating variable(s) used in an ALT should provide a means of accelerating the failure mechanism(s) expected in actual application. Screening experiments are used to identify the few key factors (accelerating variables) that have a significant effect on the life of the product to be tested. They use only a few experimental runs in order to filter out the important main effects without considering any interactions. Screening experiments are a good way to find out which variables influence a product’s life, and to help determine the maximum level of stress that should be used in an ALT. An example where a screening experiment was used successfully is provided in Lam, Guo, and Larson (2007). After an appropriate accelerating variable (most accelerated tests use only one accelerating variable, again to keep things simple) or variables have been identified, the next step is to choose the levels of the variable at which to test and decide how the available test units should be allocated to those levels. Testing at various constant levels of the accelerating variable should be the default approach, unless there is an engineering justification to utilize more complicated types of test plans, such as cyclical, step-stress or ramp stress as suggested in Chapter 6 of Nelson (1990), it is essential to simulate data corresponding to the assumed model (and planning values) and proposed design and to go through the exercise of analyzing the data, before actually conducting the test.

C. Testing at too severe stresses (requiring extreme extrapolation)

Testing at high levels of the accelerating variable, far away from the use level of the accelerating variable implies a large amount of extrapolation. Statistical models used in the analysis of accelerated test data, whether based on physical or chemical principles or from finding a good fit to the data, are, at best, idealized approximations to the truth and will always contain some degree of error. Extrapolation to use conditions will greatly amplify model these errors that are always present. Additionally, testing at levels of the accelerating variable(s) that are too extreme
- Can result in activating failure modes that would not be experienced at use conditions. Might make the assumed model inadequate at the high stress levels
- Will result in greater statistical uncertainty in estimates at the use conditions (compared to running at more moderate stresses that still induce failures)

We have observed tests where high levels of voltage or temperature caused failures to happen almost instantaneously or orders of magnitude faster than lower stress levels and use conditions. Apart from the fact that overly aggressive acceleration will usually cause new failure modes that would never be seen at use conditions (as further described in pitfall N), it is also risky to extrapolate back to use conditions. Generally, ALTs should be planned to keep the amount of extrapolation as small as possible while assuring that a sufficient number failures will be observed at several levels of the accelerating variable.

As is widely recognized today, it is not necessary to force all units to fail. Indeed upper-tail failures can distort the fit in the lower tail of the failure-time distribution that is generally of primary interest.

Computer software packages will readily (and indiscriminately) compute estimates of life at use conditions, extrapolating as much as the user requests. Obtaining credible, accurate estimates of life at use conditions requires discipline in terms of planning the test and examining the adequacy of model and the testing methods that generated the data set.

D. Using unnecessarily complicated testing and data analysis

Today there are numerous capabilities in terms of test infrastructure and software tools that can generate and analyze accelerated test data using complicated models. Two examples are simultaneous application of several different accelerating variables and test plans where the level of the accelerating variable(s) changes
(typically increased in a stepwise manner) during the test (most commonly used in HALTs). Use of such complicated approaches to reliability testing will not necessarily lead to better results from ALTs. Simpler test designs (e.g., the commonly-used constant-stress tests) provide cleaner data and require fewer assumptions. There needs to be an engineering or physics-based reason to use complex stress profiles. Complicated test plans and models require more complicated model assumptions (e.g., a cumulative damage model) and thus higher risk of incorrect conclusions because of incorrect assumptions. Working with an aerospace company, we have witnessed tests that used cyclical stresses, when cycling was not a factor that would affect the lifetime of the product being tested. This test was expensive and the resulting data were not useful for predicting life.

E. Testing at high levels of accelerating variables that cause new failure modes

It is well known among reliability engineers who use accelerated testing that one has to be careful not to increase temperature or stress so much that new failure modes, which would never be seen at use conditions, are caused. Nevertheless, this is a common pitfall. Example 19.10 of Meeker and Escobar (1998b) presents the results of an ALT of a mylar–polyurethane insulating structure that had been evaluated at five levels of voltage stress. Results at the highest voltage stress showed a serious departure from the rest of the ALT data and suggest that a different mechanism was present at that level (insulation engineers have suggested that the root cause was an internal substantial increase in temperature that could not be dissipated adequately, leading to a different failure mode). The effect of including the suspect data was to cause an overly optimistic estimate of the life distribution at the use conditions. Dropping the highest level of voltage stress from the analysis gave more credible estimates of life at use conditions.

Example 19.11 of Meeker and Escobar (1998b) presents the results of an ALT on a new IC device that had been tested at five different levels of temperature. Failures were observed only at the highest two levels of temperature. The data provided strong evidence for a different failure mode at the highest level of temperature. Again the data at the highest level of temperature had to be dropped. With failures remaining at only one level of temperature, it was necessary to use prior information about possible values for the effective activation energy and combine this information with the limited data using a Bayesian analysis, as described in Section 22.2 of Meeker and Escobar (1998b).

The example in Section 19.5.3 of Pascual, Meeker, and Escobar (2006) used both increased temperature and increased current to evaluate the lifetime of LEDs. Initially, it was impossible to find any standard ALT model to fit the data. It was discovered, however, that the test point using the highest level of junction temperature and current did not fit with the rest of the data and, when included in the analysis, caused the model to give nonsense extrapolations. When the extreme test point was dropped from the analysis, the model fit the data well and the fitted model gave credible estimates of life at the use conditions. This discovery strongly suggested that the combination of high junction temperature and high current caused a new failure mode.

When HALTs are performed before an ALT, the information from the HALT program is sometimes useful for establishing safe limits for accelerating variables to be used in the ALT.

F. Focusing only on the obvious failure mode(s)

When there is an obvious, well-known failure mode engineers can, with careful planning and execution, usually conduct an accelerated life test that will provide useful information about how that failure mode will affect product life in the field. Often, however, the most serious reliability disasters result from unanticipated failure modes that are not explicitly studied and thus not properly excited in the accelerated test. Plate growth is the well-known life-limiting failure mode of a lead-acid battery cell. An accelerated test described in Cannone, Cantor, Feder, and Stevens (2004) used elevated temperature to increase the rate of plate growth and predicted extremely long life times (hundreds of years) for that part of the system. Another accelerated test for the same cell, described in Sharpe, Shroff, and Vaccaro (1970), used elevated voltage to accelerate a known corrosion mechanism associated with the post seal and predicted that the post seal would last at least 40 years in service. Within a few years after the product was introduced into the field, however, another serious “blister corrosion” failure mode arose and caused a large number of the cells that had been installed to fail. As described in Cannone, Feder, and Biagetti (1970), this corrosion failure mode was not observed in the accelerated test because the elevated voltage actually inhibited the
blister corrosion failure mode (this is also another example of pitfall 6 in Meeker and Escobar 1998a).

3 Pitfalls during the execution of an ALT

This section includes pitfalls that are associated with the execution stage of an accelerated test.

G. Improper use of system level ALTs

The purpose of most ALTs is to focus on a particular failure mode, usually at the material or component level, and then accelerate failures by testing at higher-than-usual levels of one or more accelerating variables, such as temperature, voltage, load etc. Increasing temperature or stress on a full system (or even a subsystem) is generally less useful. This is because the maximum allowable temperature at higher levels of assembly (system or subsystem) will generally be much lower than the maximum allowable level at the material/component level. Thus one can achieve only limited amounts of acceleration. Also, if an ALT at the system level generates more than one failure mode, the different failure modes will need to be modeled separately (as described in Chapter 7 of Nelson 1990) because different failure modes will generally have different acceleration relationships.

On the other hand, use-rate accelerated tests (in which a product is used more frequently or continuously) are commonly used during product development. Such tests can be useful to accelerate certain failure modes with a system-level test. Examples of use-rate ALTs include printers that print continuously, washing machines that perform cycles continuously, engines and motors that run continuously, etc.

H. Inadequate monitoring of failures during an ALT

We have been involved in ALT applications where no useful lifetime data were available at the end of the test. For example, a company conducted a very expensive accelerated test of a component in environmental chambers that controlled temperature and humidity. They placed the test units into the chamber, set the desired levels of stress and then left the units unmonitored, without checking the status of the units at regular intervals. A week later, all of the units had failed. Because all of the units were left-censored, there was not sufficient information to estimate the life distribution of the component.

If possible, one should obtain time-to-failure information by using continuous monitoring so that exact times to failure can be determined. When continuous monitoring is not possible, periodic inspection at carefully chosen times (chosen such that the failures are assured to be spread out adequately over the different intervals between inspections) can provide nearly the same statistical accuracy as continuous monitoring Meeker (1986). Chapter 3 of Nelson (1990) provides examples of non-equally spaced inspection times that were used in testing microelectronic devices.

I. Not inspecting survivors at the end of an accelerated test (or not fully using the information gained from such inspections)

The product development program and subsequent serious reliability problems described in O'Boyle (1990) provides an enlightening example where ALT results were not interpreted properly because inspection information, although collected, was not used properly. In particular, a sample of refrigerator compressors was run for one year in an accelerated test with no failures. Because there were no failures in the test, a decision was made to launch the product. What was not reported to higher-level management was that the unfailed compressors, when disassembled for inspection, exhibited unexpected discoloration, providing evidence of a lubrication issue and suggesting that the unfailed units were well on their way to failure. After problems were discovered from field failures, it was determined that all of the refrigerators that had been sold would have premature compressor failures.

Of course it is also useful to inspect failures after an accelerated test, for various reasons, including determining if the unit failed from the expected failure mode.

J. Ignoring the impact of idle time

When planning use-rate accelerated tests, (e.g., a printer printing continuously), the effect of idle time is often overlooked (Meeker and Escobar 1998b, page 518). In the case of inkjet printer testing, the effect of
the ink drying up on the orifice plates cannot be observed when the printer is continuously passing ink through the printheads. Idle times of a week or more have been used to understand the effect of idle times in this application. Other consumer electronics systems with mechanical components need to be exposed to idle times to reveal failure modes otherwise not observed in continuous usage. For example, corrosion or thermal cycling that can cause failures that may be inhibited by constant use. Special attention has to be used when designing use-rate tests to make sure that such critical failure modes are not masked by continuous usage. In some applications, separate accelerated tests might need to be conducted to study failure modes that are not accelerated by use or to generate failures from mechanisms that are active only during idle times. For example, in the testing of automobile engines there are separate accelerated test protocols for continuous testing (where time is measured in hours of operation) and start-stop tests (where time is measured by the number of start-stop cycles) and these protocols are expected to generate different failure modes. Turning a system on and off will sometimes generate a failure mode (e.g., due to thermal cycling) that will not be seen in continuous testing.

K. ALTs that generate failures that will not arise in field use
We have seen a number of ALT and HALT programs that led to the discovery of a failure mode followed by an expensive redesign effort to eliminate the failure mode, only to learn later that the failure mode would never have occurred at use conditions. A particular example is described on page 38 of Nelson (1990). In one application a failure mode in an ALT was generated by an artifact of the test setup. It was discovered that this failure mode was different from the failure modes being seen in field tests. Further investigation showed that the ALT failure mode would never arise in the field.

4 Pitfalls during the analysis of ALT data
This section includes pitfalls that are associated with the data analysis stage of an accelerated test.

L. Attempting to estimate a life distribution from a HALT
Highly Accelerated Life Tests (also known as HALTs) serve a useful but different purpose than ALTs. HALTS are importantly useful tests that are often conducted in product development processes, especially in electronic products, to discover product design weaknesses, screen for failure modes and understand design and destruct limits (as opposed to estimating reliability). HALTs generally use a complicated combination of varying environmental variables such as temperature and voltage-level cycling and vibration. Also, HALTs often lead to modifications in the design of the product (e.g., replacing a weak component with a stronger one) whenever there is a change in a design (which could be triggered by either a HALT or an ALT), the data would no longer be relevant for estimating the product’s reliability. Because of the non-quantitative nature of these tests, it is generally not useful to try to use the results from a HALT to infer life at some specified use conditions. Engineers sometimes get pressure from management to combine results from HALTs and ALTs, in an effort to be more efficient. Because of the complicated stressing patterns in HALTs, it is doubtful that there would be a useful model (i.e., sufficiently accurate in extrapolation) relating the results of the complicated stressing to life at use conditions.

M. Using an inadequate acceleration model
As mentioned earlier, in accelerated testing, it is important to understand the physics of failure for the mechanism(s) that will be accelerated. Different accelerating variables will have different underlying physical relationships to the failure mechanism(s). For example, if temperature is to be used to increase the rate of a chemical reaction, the relationship between life and temperature might be modeled adequately with the Arrhenius relationship. We should point out, however, that the Arrhenius relationship is a simple model that will not adequately describe the effect that temperature has on all chemical reactions. Counter-examples are given in Chapter 18 of Meeker and Escobar (1998b). Also, there would be no physical basis for using the Arrhenius relationship in applications where temperature is a variable that can be controlled in the manufacturing process (as opposed to a variable being increased to increase the speed of a chemical reaction leading to failure). See the example in Meeker, Escobar and Zayac (2003) for a situation in which
temperature was a processing variable in the manufacturing of a metal spring. From this accelerated test, it was learned that the processing temperature had little effect on the fatigue life of the spring but that shot peening the springs (causing compressive stress at the surface to inhibit crack initiation) led to important improvements in fatigue life.

The inverse power relationship is commonly used to describe how physical stresses, such as pressure, load, or voltage stress will affect lifetime. The inverse power relationship assumes that log life is linearly related to log stress. Unlike the Arrhenius relationship, the inverse power relationship does not have any physical justification (although section 18.4.3 of Meeker and Escobar (1998b) gives a possible physical explanation for a specific situation). It is certainly possible that log life may be better described by some other transformation of stress (e.g., square root or reciprocal). Because model decisions involve uncertainty, sensitivity analysis is needed, as described in Meeker, Escobar and Zayac (2003), to assess the consequences of assuming different credible relationships. That is, one should compare the estimates for quantities of interest under different plausible model assumptions. When extrapolating to produce lifetime estimates (as is almost always done in accelerated testing), conclusions can be highly dependent on the particular model that is used.

N. Failure to recognize a non-monotonic accelerating variable relationship

The commonly-used accelerated test models have a monotonically decreasing relationship between life and the accelerating variable, implying that if the accelerating variable increases, life will decrease. These models have been described, for example, in Chapter 2 of Nelson (1990), Chapter 18 of Meeker and Escobar (1998b), and Escobar and Meeker (2006).

If possible, reliability engineers should assess the likely accuracy of the proposed relationship between life and the accelerating variable(s) before designing an accelerated test and analyzing accelerated test data. Most relationships can be identified if the physics or chemistry of failure is clearly understood. If that is not the case, screening or other pilot experiments could be conducted. As an example, an engineer was analyzing life versus temperature data from an ALT. As temperature increased, a different chemical reaction started and reversed the trend, as shown in Figure 1. This example is described more completely in Sarakakis, Gerokostopoulos, and Mettas (2011).

![Figure 1: Non-monotonic life versus temperature relationship](image)

O. Use of inaccurate information about an activation energy (for an Arrhenius model)

Many accelerated tests use elevated temperature to increase the rate of a chemical reaction or diffusion process and thereby reduce the life of the component. The data are then typically analyzed using the Arrhenius model relating life and temperature. A common yet dangerous practice is to presume that the effective activation energy for the Arrhenius model is known instead using experimental data to estimate this quantity. As described in Groebel, Mettas, and Sun (2001), using published or prior known values of activation energy can be misleading, because the effective activation energy may actually vary from one
production lot to another or because of design changes. Nelson (1990) refers to the choice of a given activation energy as input as “company tradition with unknown origins.” Some reliability handbooks (e.g., page 59 of Klinger, Nakada, and Menendez 1990), however, contain tables of activation energy values for different failure mechanisms. Due to extrapolation, estimates of lifetime at use conditions will be highly sensitive to the assumed value of activation energy.

Also, assuming that the activation energy is known will have the effect of making confidence intervals for life (e.g., a quantile of the failure-time distribution at use conditions) unreasonably narrow.

On the other extreme, analyzing limited accelerated test data in the usual statistical manner, taking into consideration the uncertainty in the activation energy, will generally result in confidence intervals that are extremely wide. Ignoring available information about the effective activation energy (e.g., based on previous experience with a similar product or based on a handbook value for a known failure mode) is wasteful. Examples of these two extremes for a single data set are shown in Figures 19.13 and 19.15 of Meeker and Escobar (1998b).

There are two useful methods to bring available information about the activation energy into an analysis. First, one can conduct sensitivity analyses with different given values of the activation energy over some plausible range and compare the results. Second, this approach can be formalized with a single analysis by using a probability distribution to describe the prior knowledge about the activation energy and doing a Bayesian analysis. This approach is described in Section 22.2 of Meeker and Escobar (1998b).

P. Ignoring or inadequately treating interactions

The effect of using more than one accelerating variable in a ALT can be complicated by interactions involving the accelerating variables (in statistical jargon, an interaction between two explanatory variables, sometimes known as a synergism, implies that the rate of change in the response with respect to one variable depends on the level of the other variable). An interaction would arise, for example, if the effective activation energy for temperature acceleration depends on the level of humidity. It is important to model possible interactions and understand the impact they might have on the failure mode of interest. Interactions, when they exist, tend to make interpretation and extrapolation difficult or impossible, unless there is an adequate physical/chemical model for the interaction effect. The example in Section 19.5.3 of Pascual, Meeker and Escobar (2006) shows how adding an interaction term to an ALT model led to implausible results because, in effect, the estimates at use conditions involved extrapolation with a quadratic model.

Chapter 6 in Nelson (1990) describes the planning of an ALT on a type of insulation where the experimental factors were insulation thickness (in units of mils) and voltage stress (in units of kV/mil). Assuming that there is no interaction between these two experimental variables (i.e., the effect that changing voltage stress has on life does not depend on thickness), it is easy to show that there would be a strong interaction between thickness and applied voltage (as opposed to voltage stress). Thus the proper choice of experimental factors is critical in accelerated testing where extrapolation is required. It is better to avoid strong interactions, if possible, by proper choice of experimental factors.

In another accelerated test application, described in ReliaSoft (2011), both temperature and relative humidity were used as accelerating variables. The failure mode, however, was caused by moisture penetration which only occurred at a certain temperature and relative humidity combinations. In order to account for the interaction between the two accelerating variables and model the accelerated test data appropriately, the team transformed the relative humidity to a physically-based pseudo stress as a function of both temperature and relative humidity that would better describe the test acceleration. Both dew point and vapor pressure were explored and one of them was found to provide the best physical representation from the combined variables. The choice would depend on the particular failure mechanism (e.g., drying out versus corrosion).

Another application, given in Nelson (1990), page 272, describes the modeling of fatigue failure data from an accelerated test on nickel-base alloy specimens where life was modeled as a function of a physically-motivated pseudo stress that was defined as the product of the specimen’s Young’s modulus (a measured material property describing elasticity and thus a covariate in the experiment) and strain (a controlled experimental factor). Combining these two variables into the pseudo stress (assuming that there is physical motivation for doing so) simplifies modeling and concern about possible interaction between
strain and Young’s modulus.

Q. Not comparing failure modes in the field with failure modes in the laboratory

In order to be valid, accelerated tests must generate failures in the same manner that they will be generated in actual use. This should be ascertained to the highest degree possible. As described in Meeker, Escobar and Hong (2009), the reliability of a turbine-device component in an appliance was assessed in an industry-standard accelerated test, using constant running at an increased load. This industry-standard accelerated test caused a cracking failure mode in all units that were tested. Subsequently, when field failures were inspected, only about 30% of the returned failures had crack failures. The other failures were from a different wear mechanism that was never seen in the industry-standard accelerated test. If field-failure predictions had been made on the basis of the industry standard accelerated test alone, then the number of failures would have been seriously underpredicted.

These phenomena can, indeed, be viewed as an example of the “Masked failure modes” pitfall described in Meeker and Escobar (1998a). Ultimately the manufacturer had to conduct a second accelerated test in which units were run continuously with only a light load and this test was able to reproduce the wear-type failures seen in the field.

The photodegradation experiments described in Vaca-Trigo and Meeker (2009) used Fourier transform infrared (FTIR) spectroscopy to study changes in chemistry between specimens tested under natural light/weather conditions and specimens tested in chambers exposed to controlled UV exposure and spectrum, temperature, and humidity. Using these methods it was possible to confirm that the chemical reactions leading to failure in the specimens did not differ between the laboratory and the outdoor exposure site.

R. Failure to utilize available degradation data

When available, degradation data on unfailed units can provide valuable information about the time at which units will fail. Alternatively, the degradation data can be modeled directly (see Chapters 13 and 21 of Meeker and Escobar 1998b for examples).

It is possible to take degradation data and map them into failure-time data. Previously, we have seen several examples where engineers were doing this because older reliability text books and commercial software offered reliability data analysis methods only for failure-time data. For an example, see the discussion in Meeker, Escobar, and Lu (1998). Here engineers were concerned because they had not seen any failures at the lowest level of temperature and a decision had to be made soon. It was pointed out that all of the units at the lowest level of temperature had accumulated amounts of degradation approaching the soft-failure definition and that this information, if used, would strengthen the analysis of the data.

In another example, a company measured the mechanical torque required to disassemble a joint after exposing the unit to a specific type of mechanical load. Having such degradation data allows for prediction of when the units are going to reach a critical failure threshold, which can reduce the amount of time it takes to conduct tests even more. Indeed, such degradation knowledge can be used to obtain reasonably precise reliability inferences even when there are no failures!

For the same amount of test time, degradation data always contain more information than the failure-time data and thus provide more precise estimates, particularly when there are few or no failures. Degradation data have another advantage in that they provide much more information for assessing the adequacy of an acceleration model. For example, under the assumption that the effective activation energy is the same for all test units, it is possible to estimate the activation energy by having repeated-measures degradation data on a single unit at each level of temperature. Of course it would be unwise to conduct such a single-unit test, because there would usually be an important amount of unit-to-unit variability in degradation rates.

5 CONCLUDING REMARKS

Accelerated testing is a difficult, challenging endeavor. Skill, knowledge, and great care are needed to conduct accelerated tests that provide correct and useful information about product reliability at use conditions. Due to developments in statistical methodology and their implementation in computer software packages, the planning of efficient accelerated tests and fitting models to the resulting data have become
relatively simple and routine. The difficult challenges arise in determining appropriate accelerating
variables and finding a model that will allow one to plan and conduct ALTs and ADTs appropriately and
use the results to produce reasonably accurate estimates of reliability at use conditions. This ease of
planning and analyzing the data from ALTs and ADTs creates risks of allowing those conducting these
tests to fall into some of the many pitfalls that have been outlined in Meeker and Escobar (1998a) and in
this paper. We hope that the outline of these many pitfalls will serve as a warning of the dangers of
conducting accelerated tests without careful consideration and due planning and help to promote better
accelerated tests.

ACKNOWLEDGMENT

We would like to thank Necip Doganaksoy, Harry Guo, Gerald Hahn, Mike Hamada, Yili Hong, Brad
Jones, Ming Li, Wayne Nelson, Paul Prew, and David Trindade for providing helpful comments on an
earlier version of this paper.

REFERENCES

30 Years Later.” Proceedings of the 26th Annual International Telecommunications Energy Conference,
Chicago, IL, 19-23.


552-577.

Degradation Tests: Data, Models, and Analysis.” Chapter 21 in Mathematical and Statistical Methods in

Using Accelerated-Test Data.” 2001 Proceedings Annual Reliability and Maintainability Symposium,
Philadelphia: IEEE.


Nostrand Reinhold.

Outdoor Optical Products.” 2007 Proceedings Annual Reliability and Maintainability Symposium,
Philadelphia: IEEE.

Meeker, W.Q. (1986) “Planning Life Tests in which Units are Periodically Inspected for Failure.” IEEE
Transactions on Reliability, 35, 571-578.


Reliability, 47, 114-118.

Inc.


