JOURNAL OF APPLIED POLYMER SCIENCE VOL. 18, PP. 555–565 (1974)

Use of Multiple Regression Analysis to Develop Predictive Models for Failure Times of Adhesive Bonds at Constant Stress

WILLIAM C. JONES III, ELISE MCABEE, RAYMOND F. WEGMAN, and DAVID W. LEVI, *Materials Engineering Division, Picatinny Arsenal, Dover, New Jersey 07801*

Synopsis

Predictive models were developed for failure times of adhesively bonded titanium and aluminum under a constant stress. The independent variables in the equations were temperature, relative humidity, and stress. The predictive equations were developed using multiple regression analysis which was performed by computer. A comparison of actual failure times with those predicted by the models gave at least an order-of-magnitude agreement. A previously developed reaction rate model was also verified by regression techniques.

INTRODUCTION

A question of vital interest to the design engineer who wants to use an adhesive bond is, "How long will the bond last under a given set of conditions?" Experimentally determining the answer to this question is an expensive and time-consuming process.

As the number of conditions (or variables and parameters) increase, the amount of experimentation rapidly increases. It then becomes highly desirable to express the relationship between the variables in a functional relationship or model. The model can be used to predict the variables of interest when values of the associated variables are known. Such predictions should be useful in estimating the stress, temperature, humidity, or other condition that can be tolerated if the bond is to last for a certain required time.

Since a regression model tends to smooth the experimental irregularities in data, use of a model will often result in a more accurate prediction than reliance on a single experimental point. A regression model may also be used for interpolation between experimental values, thus greatly extending the usefulness of a limited amount of experimentation.

In this paper, the functional relationship among the variables is developed by using the multiple regression analysis technique. Least-squares methods are used to determine the best functional relation among the variables, and correlation methods are used to determine the degree of fit.

TABLE I
Failure Data[a] for AF126 Adhesive (CP Titanium Adherends) Under Constant Stress

Alkaline cleaned				Phosphate-fluoride treated			
Temp., °K	t_f, hr	S, psi	Relative humidity, %	Temp., °K	t_f, hr	S, psi	Relative humidity, %
296	21	1980	95	296	116	1980	95
	5.1	2200			100	2200	
	4.9	1210			120	2420	
	7.5	2640			28	2640	
	2	2860			15.5	2860	
322	160	880	95	322	234	1100	95
	76	1100			187	1320	
	72.5	1320			80	1540	
	1.4	1540			53	1760	
	0.3	1760		333	34	1100	95
333	175	660	95		27	1320	
	72	880			19	1440	
	22	1100			9.3	1760	
	5	1320			1.2	1980	
	2	1540		344	30	660	95
344	6	880	95		19	880	
	1.5	1100			13.5	1100	
	0.3	1320			9.5	1320	
322	70	1760	50		3.5	1540	
	20.5	1980		322	10	1980	50
	3.8	2200			7	2200	
	0.7	2420			3.5	2420	
	1.1	2640			0.8	2640	
333	9	1320	50	333	138	1540	50
	1.8	1540			25	1760	
	1.0	1760			10	1980	
	0.5	1980			1	2200	
	0.3	2200		344	12	1100	50
344	4.0	1100	50		4	1320	
	4.0	1320			1.5	1540	
	1.1	1540			0.8	1760	
	1.0	1760			0.5	1980	
	0.3	1980					

[a] Averages of four values.

These statistical methods are well established and have been programmed for calculation by digital computer, making it possible to evaluate many models with little computational effort.

With the computational burden removed, the problem is reduced to one of deciding what models to investigate. There are two conventional approaches in choosing the models.[1] These are (1) use of a functional relationship based on existing theories describing the phenomenon concerned and (2) examination of two-dimensional plots of each independent variable versus the dependent variable. The first method is preferred since the resulting model is often more reliable, particularly if extrapolations are

desired. However, if little is known about the basic mechanisms involved, the second method provides an indication of the functional relationships among the variables. The investigator is thereby guided in his choice of which models to use for regression analysis.

RESULTS AND DISCUSSION

Source of Data

The data used in this study have been previously reported[2,3] but are given in Tables I to III for convenience. Two types of titanium adherends

TABLE II
Failure Data[a] for AF126 Adhesive (6,4 Titanium Adherends) Under Constant Stress

Alkaline cleaned				Phosphate-fluoride treated			
Temp., °K	t_f, hr	S, psi	Relative humidity, %	Temp., °K	t_f, hr	S, psi	Relative humidity, %
296	13	1980	95	296	580	1980	95
	1.9	2200			147	2200	
	8	2420			372	2420	
	8	2640			31	2640	
	2.5	2860			41	2860	
322	3.3	1320	95	322	161	1320	95
	5.3	1540			97	1540	
	1.7	1760			62	1760	
	6.0	1980			40	1980	
	1.2	2200			7	2200	
333	85	880	95	333	45	1100	95
	53	1100			49	1320	
	30	1320			36	1540	
	2.9	1540			8	1760	
	3.7	1760		344	32	880	95
344	44	440	95		31	1100	
	13	660			9	1320	
	6	880			7	1540	
	1.1	1100			2	1760	
	0.8	1320		322	60	1540	50
333	30	1100	50		38	1760	
	18	1320			55	1980	
	12	1540			61	2200	
	0.2	1760		333	16	1540	50
	0.3	1980			23	1760	
344	36	1100	50		1.5	1980	
	7	1320			16	2200	
	7	1540			0.3	2420	
	0.9	1760		344	8	1540	50
					17	1760	
					1.2	1980	
					0.7	2200	
					0.25	2420	

[a] Averages of four values.

TABLE III
Failure Data[a] for AF126 Adhesive (Anodized 2024-T3 Aluminum Adherends) Under Constant Stress

Temp., °K	t_f, min	S, psi	Relative humidity, %
296	533	3080	90
	4,320	2600	
	10,080	2200	
	38,880	1760	
322	15,840	1760	95
	19,440	1540	
	23,760	1320	
	9,360	1100	
	64,800	880	
344	860	1760	95
	900	1540	
	3,756	1320	
	2,700	1100	
	6,900	880	
	248	1980	
	308	1760	
	214	1540	
322	18,720	2420	50
	6,300	2640	
	340	2860	
333	11,200	1760	50
	7,510	1980	
	1,150	2200	
	350	2420	
344	14,200	1320	50
	10,720	1540	
	4,620	1760	
	300	1980	
	610	2200	
322	11,340	2420	20
	490	2640	
	150	2860	
	260	3080	
	230	3300	
333	56	2200	20
	130	2420	
	153	2640	
	147	2860	
	28	3000	
344	720	1760	20
	490	1980	
	160	2200	
	110	2420	
	75	2640	

[a] Averages of four values.

were used, either the "commercially pure" (CP) or the 6 aluminum–4 vanadium (6,4) titanium alloy. Each bonded titanium surface underwent one of two types of pretreatments, either alkaline cleaned (AC) or phosphate–fluoride (PF) treatment. Anodized 2024T-3 aluminum adherends were used. The adhesive was a commercial epoxy (AF126).

Bonded specimens (lap-shear) of titanium and of aluminum were placed under constant stress at various stress levels and environmental conditions, and the failure time (t_f) for each set of conditions was observed.

Titanium Adherends

Analytical Sequence and Development of Models

The independent variables for which experimental data were available, and which therefore were included in the model, were per cent relative humidity (H), stress in psi (S), and temperature in degrees Kelvin (T).

As mentioned previously, a theoretical relationship among the above variables would be desirable. However, the science of adhesion has not progressed to the point of being able to provide complete theories describing the dependence of bond failure time on these variables. Some success has been obtained[3] in applying reaction rate theory to predict failure times of adhesive bonds at constant stress. However, the application of this theory to adhesive bonds requires a number of assumptions.[3] Even then, the theoretical relationship includes only the variables failure time, stress, and temperature, with no theoretical way of expressing relative humidity. Although no purely theoretical approach was possible, "quasi-theoretical" models using reaction rate theory were developed for one data subset (PF-pretreated 6,4 titanium). The models were essentially the referenced reaction rate models with relative humidity terms included as follows:

$$\log t_f = a_0 + a_1 \log T + a_2(1/T) + a_3(S/T) + a_4H + a_5H^2 + a_6HT$$

With there being no real theoretical basis for the development of the models, most of the emphasis was put on the more empirical approach based on examining individual plots. The terms to be used in the models were determined from the individual plots of the dependent variable versus each of the independent variables. The best fit to smooth curves was obtained by plotting the logarithm of the failure time versus each independent variable. The use of logarithmic failure time was also indicated by reaction rate theory,[3,4] and experience has shown that models with logarithmic failure time will smooth data of this type. The plots for the variables temperature and stress were nonlinear, indicating that second-order terms may improve the correlation. Since the titanium data included only two levels for relative humidity, no conclusions could be drawn above the curvature of this variable, and it could only be included as a first-order variable.

Since the plots indicated a nonlinear relation between log t_f and stress or temperature, several possible relationships were screened. This was done

TABLE IV
Regression Equations and Correlation for PF-Pretreated 6,4 Titanium

Model no.		Corr. coeff.	Error,[a] %
	Empirical Models		
1	$\log t_f = a_0 + a_1S + a_2H + a_3T + a_4S^2 + a_5T^2 + a_6SH + a_7ST + a_8HT$	0.9450	28.4
2	$\log t_f = a_0 + a_1S + a_2H + a_3T + a_4S^2 + a_5T^2 + a_7ST + a_8HT$	0.9444	27.9
3	$\log t_f = a_0 + a_1S + a_2H + a_3T + a_4S^2 + a_7ST + a_8HT$	0.9444	27.4
4	$\log t_f = a_0 + a_1S + a_2H + a_3T + a_7ST + a_8HT$	0.9417	27.5
5	$\log t_f = a_0 + a_1S + a_2H + a_3T + a_8HT$	0.9417	27.0
6[b]	$\log t_f = a_0 + a_1S + a_2H + a_3T$	0.9388	27.2
	Theoretical Models		
7	$\log t_f = a_0 + a_1 \log T + a_2(1/T) + a_3(S/T) + a_4H + a_5HT$	0.9413	27.6
8[c]	$\log t_f = a_0 + a_1 \log T + a_2(1/T) + a_3(S/T) + a_4H$	0.9383	27.8
	Factorial Model		
9	$\log t_f = a_0 + a_1 \log S + a_2 \log H + a_3 \log T$	0.9210	30.8

[a] Standard error of the estimate as a per cent of the mean response.
[b] Best empirical model.
[c] Best theoretical model.

by fitting the data to computerized least-squares curves corresponding to the following equations:

$$Y = A + BX \quad \text{linear} \quad (1)$$

$$Y = AX^B \quad \text{parabolic curve (pos. B)} \quad (2)$$

$$\text{hyperbolic curve (neg. B)} \quad (3)$$

$$Y = Ae^{BX} \quad \text{exponential curve} \quad (3)$$

$$Y = A + BX + CX^2 \quad \text{quadratic} \quad (4)$$

where Y is $\log t_f$ and X is the independent variable. For both stress and temperature, the best fit was obtained by the quadratic equation, followed by the linear equation. These results confirmed those of the individual plots in suggesting second-order terms for the variables stress and temperature. The following second-order polynomial (with relative humidity as first order) is suggested by these results:

$$\log t_f = a_0 + a_1S + a_2H + a_3T + a_4S^2 + a_5T^2 + a_6SR + a_7ST + a_8HT.$$

Since it is possible and indeed probable that not all the terms in this model are significant, the backward elimination procedure,[5] which utilizes

the standard analysis of variance data to eliminate insignificant terms, was used.

A third type of model (factorial) has been used by Bussa, Sheth, and Swanson[6] for metal life under random fatigue loads. Applied to the variables in this study, their model would be

$$t_f = C(S)^{a_1}(H)^{a_2}(T)^{a_3}$$

where C, a_1, a_2, a_3 are constants. This equation, when linearized by taking logarithms of both sides, becomes

$$\log t_f = \log C + a_1 \log S + a_2 \log H + a_3 \log T.$$

This model represents the third and final approach to developing a predictive model for adhesive bond failure time.

Regression Equations and Correlations

Table IV illustrates the computer computations that were carried out in a single case. Table V shows the best of the models obtained for all the systems. The best models represent a trade-off between including enough

TABLE V
Best Regression Equation for Each System Studied

System	Equation[b]	Corr. coeff.	Error,[a] %
PF-Pretreated 6,4 titanium	$\log t_f = 23.4 - 0.151S - 0.0234H - 5.93T$	0.9388	27.2
AC-Pretreated 6,4 titanium	$\log t_f = 16.4 - 0.515S - 5.71H - 2.50T + 0.00252S^2 + 0.353SH$	0.7893	61.1
PF-Pretreated CP titanium	$\log t_f = 25.6 - 0.153S - 0.202H + 24.5T - 4.73T^2$	0.9270	26.8
AC-Pretreated CP titanium	$\log t_f = 28.4 - 0.395S - 2.00H - 6.55T + 0.00574S^2$	0.8955	61.5

[a] Standard error of the estimate as a per cent of the mean response.
[b] Regression coefficients based on data being scaled by 1/100 with t_f in hours.

variables to give a good correlation and excluding any insignificant variables in order to reduce the complexity of the models. The results indicate that the quadratic model is somewhat better than either the factorial or quasi-theoretical models.

Prediction Accuracy of the Models

According to the conventional F-test for significance of regression, all of the models given in Table V are significant at the 99% confidence level. Although this implies that the model will be a "good predictor," it does not reveal how good.

A quantitative measure of the accuracy of the model is the square of the multiple correlation coefficient. This value stated as a percentage measures

TABLE VI
Illustration of Data Smoothing by the Model for PF-Pretreated 6,4 Titanium at 296°K and 95% Relative Humidity[a]

S, psi $\times 10^{-2}$	Actual t_f, hr	Predicted t_f, hr
19.8	580	700
22.0	147	326
24.2	372	152
26.4	31	71
28.6	41	33

[a] Model: $\log t_f = 23.4 - 0.151S - 0.0234H - 5.93T$.

the percentage of the variation in the data that is explained by the model. The values of percentage squared correlation coefficients in the order in which they appear in Table V are 88%, 62%, 86%, and 80%. The models explaining 80% or better of the variation can be considered good predictors, especially for this type of data. Examination of the AC-pretreated 6,4 titanium data reveals a great deal of scatter; thus, it is not too surprising the the best model explains only 62% of the variation in the data.

For most of the data, at constant temperature and relative humidity, as stress increases the failure time decreases. However, there is a great deal of scatter in the data for PF-pretreated 6,4 titanium at 296°K and 95% relative humidity, as shown in Table VI. The predicted failure times are far more consistent than the actual data. It is highly probable that the scatter of the actual data is due to experimental difficulties and that the predicted data are a better estimate of the mean response than the actual data. As can be seen from this example, a model can be quite useful for smoothing the experimental inconsistencies of data.

Regrouping of Data by Parameters

By regrouping the data to combine one or both of the parameters, models can be developed for the following five cases; (1) all CP titanium data (both pretreatments); (2) all 6,4 titanium data (both pretreatments); (3) all PF pretreated data (both adherends); (4) all AC pretreated data (both adherends); and (5) all data (both adherends and pretreatments).

Regression analysis on the data regrouped in this way revealed that pretreatment was an important parameter whereas adherend was not. The correlation coefficient for the combined adherends was not significantly different from those for each adherend subset. When the pretreatments were combined for each adherend, however, there was a considerable decrease in correlation as compared to the individual pretreatment subsets, revealing that pretreatment is a necessary parameter. Likewise, when all the data were combined, there was a decrease in correlation, again due to the necessity of the pretreatment parameter.

Aluminum Adherends

Selection of Regression Model

Based upon the development of the titanium model in the foregoing sections, the development of the aluminum regression model was simplified. Both the empirical terms and those based on reaction rate theory were used in the development of the model. Rather than determining the best empirical and quasi-theoretical equations separately, the best overall equation was determined. This was accomplished by combining both the theoretical and empirical terms into one equation. The resulting model,

$$\log t_f = a_0 + a_1S + a_2H + a_3T + a_4S^2 + a_5H^2 + a_6T^2 + a_7SH + a_8ST\ a_9HT + a_{10}\log T + a_{11}\,(1/T) + a_{12}\,(S/T) + a_{13}\,(H/T) + a_{14}\,(S/H)\ a_{15}\ \ (1/H)$$

consists of a second-order polynomial (relative humidity could be included as a second-order variable since it was measured at three levels, for this case) with six theoretically based terms added. Starting with this 15-variable model, the most insignificant terms were eliminated by the backward elimination method until all terms were significant at a 95% confidence level. In accordance with the backward elimination procedure, the term with the lowest F-test value was eliminated from the previous model. After the eleventh elimination, the remaining terms were all significant since the F-test values for S, R, T, and H^2 exceeded the tabulated F-test value. It is interesting to note that all of the theoretically based variables were insignificant and were eliminated. In contrast to the titanium models, however, relative humidity was the only significant second-order term. In the titanium data, relative humidity was not varied at a sufficient number of levels to reflect a second-order term.

Regression Equation and Predictions

The final model with all significant variables was

$$\log t_f = 19.5 - 0.137S + 4.96H - 4.35T - 4.69H^2.$$

The correlation coefficient squared ($0.8935^2 = 0.7983$) indicates that the model will explain about 80% of the variation in the data. The standard error of the estimate was 13% of the mean response. From these statistics, the regression model can be considered a good predictor of the failure time.

In order to evaluate the predictability of the model, eight sets of data were excluded and the regression model recomputed from the remaining data. Using the recomputed regression equation, the failure times for the excluded data were predicted and compared with the actual failure times. The results are given in Table VII.

TABLE VII
Use of Model to Predict Failure Time for Aluminum Data[a]

S, psi $\times 10^{-2}$	$RH \times 10^{-2}$, %	T, °K $\times 10^{-2}$	Actual t_f, min	Predicted t_f, min
13.2	.95	3.22	23,760	13,900
17.6	.95	3.44	860	405
15.4	.95	3.44	214	802
19.8	.95	3.44	562	604
17.6	.50	3.44	4,620	2,488
28.6	.20	3.22	150	249
26.4	.20	3.33	153	167
22.0	.20	3.44	160	222

[a] Model: $\log t_f = 19.3 - 0.135S + 4.57H - 4.28T - 3.46H^2$.

Verification of Graphically Determined Reaction Rate Models

In a previous report,[3] a reaction rate model was fitted to the aluminum data. The model relates failure time to stress and temperature in the following form:

$$\log t_f = C - \log T + a(1/T) - b(S/T)$$

where C and b are constants and a is an activation energy term ($a = \Delta H\ddagger/2.3R$). The parameters of this model were previously evaluated by a technique involving a graphic extrapolation.[3]

The regression analysis technique provides an easy and precise method of checking the previous computations. The data can be divided into sets of like relative humidity, and a regression model can be created to be comparable to the reaction rate model. Since the log T term must have a coefficient of unity, that term must be excluded from the regression. Over the range of temperatures considered (296 to 344), the log T term has values from 2.47 to 2.54 and can thus be considered a constant with a value of 2.5. For the regression analysis, the term is incorporated with the constant term C and can later be separated to give the equation in the proper form. Table VIII compares the two methods of computation. The agreement is remarkably good. The coefficients for the 20% data were not computed by the graphic method owing to the uncertainty of the method caused by the excessive scatter in the data. Although the correlation coefficient for the computer regression method reflects this same scatter, the "best" fit was made and the coefficients reported.

The activation energy can be determined from the a coefficient. The activation energy was 24, 55, and 32 kcal/mole for the 95%, 50%, and 20% relative humidities, respectively. Since the earlier report considered the activation energy for only the two highest humidities,[3] it was inferred that more energy was required as the humidity decreased. This was attributed to less weakening of the bonds by the lower humidity. However, the 32 kcal/mole for the 20% relative humidity reveals that the trend is not a continuous decrease of activation energy with increased humidity but that

TABLE VIII
Verification of Graphically Determined Reaction Rate Model[a]

Method of calculation	Relative humidity	Coefficients			Correlation coefficient
		C	a	b	
Computer	95%	−7.7	5,240	−0.45	0.92
Graphical	95%	−7.4	5,130	−0.45	0.90
Computer	50%	−25.5	11,900	−0.75	0.88
Graphical	50%	−22.5	11,000	−0.71	0.86
Computer	20%	−13.1	6,970	−0.39	0.69
Graphical	20%	—	—	—	0.54

[a] Model: $\log t_f = C - \log T + a(1/T) - b(S/T)$.

there is a maximum. Thus, the effect of humidity on bond strength is more complex than a simple inverse relationship.

References

1. B. Ostle, *Statistics in Research*, Iowa State University Press, Ames, Iowa, 1963, pp. 159–205.

2. R. F. Wegman, National SAMPE Tech. Conf. Series, Vol. 4, 4th Natl. SAMPE Tech. Conf., 1972, pp. 425–429; *Appl. Polym. Symposium* **19**, 385 (1972); *Evaluation of the Adhesive Bonding Processes Used in Helicopter Manufacture*, Part I, Picatinny Arsenal Technical Report 4186, 1971.

3. E. McAbee and D. W. Levi, *Use of a Reaction Rate Method to Predict Failure Times of Adhesive Bonds at Constant Stress*, Picatinny Arsenal Technical Report 4105, 1970.

4. B. D. Coleman and A. G. Knox, *Text. Res. J.*, **27**, 393 (1957).

5. N. R. Draper and H. Smith. *Applied Regression Analysis*, Wiley, New York, 1966, pp. 163–169.

6. S. L. Bussa, N. J. Sheth, and S. R. Swanson, *Mater. Res. Stand.*, **12**(3), 31 (1972).

Received March 19, 1973
Revised August 14, 1973