

ornl

**OAK RIDGE
NATIONAL
LABORATORY**

LOCKHEED MARTIN

ORNL/TM-13359

Prediction of External Corrosion for Steel Cylinders

B.F. Lyon

RECEIVED
MAR 12 1997
OSTI

MANAGED AND OPERATED BY
LOCKHEED MARTIN ENERGY RESEARCH CORPORATION
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

ORNL-27 (3-96)

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from 615-576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

ORNL/TM-13359

**PREDICTION OF EXTERNAL CORROSION
FOR STEEL STORAGE CYLINDERS**

B. F. Lyon

Date Issued--February 1997

Prepared by
Risk Analysis Section
Life Sciences Division
Oak Ridge National Laboratory

Prepared for
U.S. Department of Energy
Office of Environmental Restoration and Waste Management
under budget and reporting code EW 20

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6285
managed by
LOCKHEED MARTIN ENERGY SYSTEMS, INC.
under contract no. DE-AC05-96OR22464
with the U.S. DEPARTMENT OF ENERGY

CONTENTS

EXECUTIVE SUMMARY	vii
1. INTRODUCTION	1
2. BACKGROUND	2
3. MODELING METHODS	4
Modeling Maximum Pit Depth	4
Modeling the Initial Thickness	5
Calculation of Confidence Limits	5
4. SUMMARY OF DATA UTILIZED	6
K Yard at Oak Ridge	7
C-745-B Yard	7
C-745-F Yard	7
C-745-G Yard	8
C-745-K and L Yards	8
PORTS	9
5. DATA ANALYSIS	9
K-1066-K Yard	11
C-745-B/F/K/L Yards, Bottom Row	12
C-745-G Yard, Bottom Row	13
C-745-B/F/G/K/L Yards, Top Row	13
PORTS, Top Row	13
PORTS, Bottom Row	13
6. RESULTS	14
7. UNCERTAINTIES AND LIMITATIONS	18
8. CONCLUSIONS AND RECOMMENDATIONS	19
9. REFERENCES	20
APPENDIX A: FIGURES	A-1
APPENDIX B: METHODS	B-1
APPENDIX C: SIMULTANEOUS CONFIDENCE LIMITS ON THE PERCENTILES OF A NORMAL DISTRIBUTION	C-1

TABLES

1. Summary of thin-walled cylinder data utilized (numbers in parenthesis are the ages, or range of ages, of the cylinders when evaluated).	6
2. Summary of population datasets and modeling assumptions.	10
3. Comparison of statistics for corrosion rate for K-1066-K subpopulations.	11
4. Summary results for cylinder populations for which data are available.	16
5. Summary results for all cylinder populations.	17

EXECUTIVE SUMMARY

The United States Department of Energy (DOE) currently manages the UF₆ Cylinder Program (the program). The program was formed to address the depleted-uranium hexafluoride (UF₆) stored in approximately 50,000 carbon steel cylinders. The cylinders are located at three DOE sites: the K-25 site (K-25) at Oak Ridge, Tennessee; the Paducah Gaseous Diffusion Plant (PGDP) in Paducah, Kentucky, and the Portsmouth Gaseous Diffusion Plant (PORTS) in Portsmouth, Ohio.

The System Requirements Document (SRD) (LMES 1996a) delineates the requirements of the program. The appropriate actions needed to fulfill these requirements are then specified within the System Engineering Management Plan (SEMP) (LMES 1996b). The report presented herein documents activities that in whole or in part satisfy specific requirements and actions stated in the UF₆ Cylinder Program SRD and SEMP with respect to forecasting cylinder conditions. The wall thickness projections made in this report are based on the assumption that the corrosion trends noted will continue. Some activities planned may substantially reduce the rate of corrosion, in which case the results presented here are conservative. The results presented here are intended to supercede those presented previously (Lyon 1995,1996), as the quality of several of the datasets has improved.

The minimum wall thicknesses of interest used in this report are 0, 62.5 mils, and 250 mils (1 mil = 0.001 in.). These thicknesses are preliminary boundaries identified within the program that indicate loss of material, safe handling and stacking operations and safe off-site transport and contents transfer criteria, respectively. In general, these safety criteria are based on area of wall thinning. However, the minimum thickness predicted in this report is for a region with an area of about 0.01 sq. in, because this is the type of data used. Using minimum point thickness adds a considerable degree of conservatism to the results in this report.

The summary results are provided in Table 5 of this report. Most of the cylinders predicted to have a minimum thickness less than 250 mils in 1996 are located in K-1066-K yard at K-25 or were in the bottom row of C-745-G yard at PGDP. Of the few cylinders predicted to have a minimum thickness less than 62.5 mils or have a point breach in 1996 (using expected values), all are in these same two yards. Both of these yards have had a large fraction of cylinders that were in ground contact at one time, although they are no longer in ground contact. A single breach is predicted in 1996 using conservative upper confidence limits in a different cylinder population at PGDP (the PGDP bottom row population that is treated separately from C-745-G yard); however, this may be due to the small sample size for this population.

THIS PAGE INTENTIONALLY LEFT BLANK

1. INTRODUCTION

The United States Department of Energy (DOE) currently manages the UF₆ Cylinder Program (the program). The program was formed to address the depleted-uranium hexafluoride (UF₆) stored in approximately 50,000 carbon steel cylinders. The cylinders located at three DOE sites: the K-25 site at Oak Ridge, Tennessee (K-25); the Paducah Gaseous Diffusion Plant in Paducah, Kentucky (PGDP), and the Portsmouth Gaseous Diffusion Plant (PORTS) in Portsmouth, Ohio.

The System Requirements Document (SRD) (LMES 1996a) delineates the requirements of the program. The appropriate actions needed to fulfill these requirements are then specified within the System Engineering Management Plan (SEMP) (LMES 1996b). The report presented herein documents activities that in whole or in part satisfy specific requirements and actions stated in the UF₆ Cylinder Program SRD and SEMP with respect to forecasting cylinder conditions. The wall thickness projections made in this report are based on the assumption that the corrosion trends noted will continue. Some activities planned may substantially reduce the rate of corrosion, in which case the results presented here are conservative. The results presented here are intended to supercede those presented previously (Lyon 1995,1996), as quality of several of the datasets has improved.

System Requirement 1.2.2 states that performance shall be monitored and evaluated to identify potential risks within the program. The subsequent **SEMP Action 2.1.2** is to model corrosion to project cylinder integrity. This report establishes the technique for modeling corrosion rates used in the program to forecast cylinder wall thickness conditions in the future.

System Requirement 4.1.2 calls for cylinder conditions to be monitored. The subsequent **SEMP Action 3.1.2** is to statistically determine the baseline condition of cylinder populations by obtaining quantitative data. This report contains the statistical method used in the program to apply the available quantitative data to cylinder populations. Populations have been established based on historical storage locations (yard and position) and similarity of quantitative data. Wall thickness and corrosion pit depth data have been collected for several subpopulations of cylinders. Most recently, these data were collected between March 1996 and September 1996 at the PGDP and PORTS.

System Requirement 4.2.2 further states that cylinder conditions shall be forecast to direct surveillance and maintenance resources. **Technical Requirement 4.2.2a** is that specific information, as determined by the program, shall be tracked to project the current and future conditions of the system. In addition, **Technical Requirement 4.2.2.b** entails the development of mechanisms to consolidate information for summary level decision-making determinations. **SEMP Action 2.2.1** is to integrate cylinder condition elements to be forecast with cylinder categorization. **SEMP Action 3.1** is to forecast cylinder conditions using parameters identified. Wall thickness, the subject of this report, is one parameter identified in the program to forecast cylinder conditions. The available wall thickness data are used to forecast out year conditions.

SEMP Action 3.1.1 is to project the number of non-compliant cylinders. The disposition of any particular cylinder for storage, handling, and transfer is based on the condition of the cylinder, where "condition" is ultimately reflected by the minimum wall thickness of a cylinder. The wall thickness parameters (0, 62.5 mils, and 250 mils) used in this report are preliminary boundaries identified within the

program that indicate loss of material, safe handling and stacking operations, and safe off-site transport and contents transfer criteria, respectively. In general, these safety criteria are based on area of wall thinning rather than minimum thickness at any one point as used in this report. Using minimum point thickness adds a considerable degree of conservatism to the results in this report.

2. BACKGROUND

The basic problem addressed here is to estimate how many cylinders will have a minimum thickness below some value z by time t . The current analyzed data only allow estimating the minimum thickness at a small point, and not the thickness over a large area. Additional analysis of the available data could provide estimates of the thickness over a larger area.

Let $C_0(x)$ denote the initial wall thickness (mils) at a location x on the cylinder, and let $P(t,x)$ denote the amount of corrosion that has occurred (mils) at location x by time t . The minimum wall thickness at time t for a given cylinder, denoted here by $M(t)$, is given by

$$M(t) = \min_x \{C_0(x) - P(t,x)\}$$

where the minimum is over all points x on the cylinder.

If the only concern is about the minimum thickness for a given cylinder at a given time, then knowledge of both $C_0(x)$ and $P(t,x)$ is not necessary. One can simply estimate the minimum wall thickness directly, although there will be measurement error that depends on the type of equipment used, as well as uncertainty as to the exact location of the minimum wall thickness. However, in order to make predictions for unsampled cylinders, or to make predictions for future time periods, assumptions must be made about the nature of the quantities $C_0(x)$ and $P(t,x)$.

Because the thicknesses of the cylinder walls were not recorded when they were first delivered, it is not possible to determine $C_0(x)$ as a function of x . For this reason, the joint distribution (in x) of $C_0(x)$ and $P(t,x)$ cannot be known. Assumptions must then be made about $C_0(x)$. One such assumption is to treat it as an independent (from $P(t,x)$) random variable, in which case the minimum thickness $M(t)$ for a given cylinder is then also a random variable, defined by

$$\begin{aligned} M(t) &= C_0 - \max_x \{P(t,x)\} \\ &= C_0 - P(t) \end{aligned}$$

where $P(t)$ is defined as the maximum penetration depth for a given cylinder of age t . The corrosion rate is then dP/dt . Thus, even given the knowledge of the value of $P(t)$, there would be uncertainty in $M(t)$ due to uncertainty in the initial thickness where the maximum pit depth occurred, C_0 , for the given cylinder. The design range for the initial thickness of thin-walled cylinders is from 302.5 to 345.5 mils, and so it could be argued that, without sufficient supporting information, the lower end of the design range must be used to confidently bound the minimum thickness. However, if it is acceptable that where the maximum penetration

depth occurs the initial thickness is actually larger than 302.5 mils, then less conservative estimates may be possible.

The preceding discussion pertains to estimating the minimum thickness for a given cylinder. When estimating the thickness for a population of cylinders, there are two additional sources of variability: (1) variability across cylinders of the maximum penetration depth $P(t)$, and (2) variability across cylinders of the distribution of initial thickness C_0 . Variability in $P(t)$ can be due to random variations in the corrosion process, and to differences in the storage history of the cylinders. Variability in the distribution of initial thickness can be due to differences in the methods used by manufacturers.

Sampling of the cylinders provides estimates of the maximum penetration depth $P(t)$, and from these data predictions must be made for the maximum penetration depth for both the unsampled cylinders and for all cylinders as a function of age. Even for a fixed age, there will be variability in the maximum penetration depth. Given the uncertainty in the storage histories for the cylinders, conservative estimation of the variability in $P(t)$ can be desirable. When approximating the distribution for $P(t)$, it is natural to restrict oneself to methods of the form $P(t) = F(\alpha(t), \beta(t))$ where $F(\alpha, \beta)$ is a two-parameter random variable (e.g., the parameters α and β could be the mean and standard deviation). The goal is then to approximate what the functions $\alpha(t)$ and $\beta(t)$ are.

The minimum thickness for a given population of cylinders is then given by

$$M(t) = C_0 - F(\alpha(t), \beta(t))$$

and the probability that a particular cylinder of age t will have a minimum thickness below a given thickness z is

$$\text{Prob}\{M(t) < z\} = \text{Prob}\{C_0 - F(\alpha(t), \beta(t)) < z\}$$

With the exception of cylinders that are being purchased now, there is no way to know the distributions C_0 . The nominal wall thickness for the design specifications is 312.5 mils, and the data collected suggest that the wall thickness on relatively uncorroded areas of a cylinder is usually thicker than this.

Ultimately, for a given population of cylinders, the total number of cylinders with a minimum thickness below a given value z at time T is estimated by

$$\begin{aligned} \text{\# Cylinders with minimum thickness below } z \text{ at time } T = \\ \sum_t (\text{\# cylinders of age } t \text{ at time } T) \times \text{Prob}\{M(t) < z\} \end{aligned}$$

3. MODELING METHODS

There are two basic quantities that must be estimated in order to predict the minimum thickness: the maximum penetration depth as a function of time, and the initial thickness for the unsampled cylinders.

Modeling Maximum Pit Depth

An expected feature of the corrosion rate and penetration depths is that, in general, they decrease with time. The problem that must be addressed is defensibly quantifying just what the rate is. It may be that in many cases (i.e., many subpopulations of cylinders) the corrosion process has reached a condition in which, whatever the past corrosion history for each cylinder may have been, each cylinder is corroding at some relatively constant (over the year) rate. If this is true, then the modeling is fairly straightforward: one determines the current condition for each cylinder, estimates what the current "constant rate" is, and then projections can be made for future times. Conceptually, this idea is appealing; however, to do this successfully requires a reliable "picture" of the current conditions, and determination of the current "constant" corrosion rates for all relevant populations. Both of these factors rely on the quantity and quality of the data available.

Prediction of the distribution of penetration depths $P(t)$ across cylinders is critical to the whole process of estimating the number of cylinders that have a minimum thickness below a given level. The simplest manner in which to do this is to assume that the general shape of the distribution of penetration depths is the same for all ages, but the mean or median (measures of central tendency) and standard deviation (i.e., the amount of "spread" in the distribution) is changing in a specified fashion. Focus is then directed to determining exactly how these factors depend on age.

The main approach utilized here allows modeling of the "leveling off" commonly observed, and is of the form $P(t)=A t^n$ where A and n are constants. This is often referred to as the "linear bilogarithmic law," and there are many applications of this model in long-term corrosion prediction (Felieu et al. 1993a; Felieu et al. 1993b; Legault and Preban 1975; Pourbaix 1982; Mughabghab and Sullivan 1989; Romanoff 1957). Determination of A and n are performed by doing a linear regression of the form $\ln P = \ln A + n \ln t$. According to Pourbaix (1982), Passano (1934) was the first to use such a relationship in corrosion prediction. This law is considered to be valid for different types of atmospheres (rural, marine, industrial) and a number of materials. The parameter A can be interpreted as the corrosion in the first year, and the parameter n represents the attenuation of the corrosion because of the passivation of the material in the atmosphere (Pourbaix, p.115). It is also possible to discuss this model in terms of the mean (or age-averaged) corrosion rate, since the mean corrosion rate is given by $P/t = A t^{n-1}$. If $n=1$ then this implies that the age-averaged corrosion rate is constant, while if $n < 1$ (which is usually the case) then the corrosion rate decreases with time. Mechanistic interpretations of n have also been made (Horton 1964). If $n=0.5$, then the relationship is said to be parabolic, with the corrosion rate controlled by diffusion through the rust layer. If $n < 0.5$, then this implies that the rust layer is showing protective properties, while if $n > 0.5$, then the rust layer is not protective because of factors that may be preventing the homogeneous thickening of the rust layer. This approach is used in several Department of Energy models to predict time to breach due to external corrosion for carbon steel containers in soil. This method can be rather sensitive to deviations from the "leveling off" pattern usually expected for the penetration depth, and so must be used with caution. Indeed, in cases where no "leveling off" is observed (due, for example, to a narrow age range for cylinders), the simpler method used in Lyon (1995,1996) is utilized.

In order to address the variability inherent in the corrosion process, it is assumed that the penetration depths are lognormally distributed at each time. This can also be expressed as $\ln P(t) \sim N(\ln A + n \ln t, \sigma_t)$, where $N(\mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ . For this model, the median is equal to At^n , the arithmetic mean μ is $At^n \exp[0.5\sigma^2]$, and the arithmetic standard deviation is $At^n \exp[0.5\sigma^2] [\exp(\sigma_L^2) - 1]^{1/2}$. The coefficient of variation (ratio of the standard deviation to the mean) is constant with time, and is equal to $[\exp(\sigma_L^2) - 1]^{1/2}$.

Given that the data consist of what are considered to be maximum pit depths, it is natural to apply extreme-value statistics to this problem. Indeed, application of this method (without confidence limits) is discussed in several papers in the literature and has also been suggested for use within this program in Rosen and Glaser (1996). The basic premise underlying this theory is that the distribution of extreme values, under rather general assumptions, should have a specific (parametric) form. However, for the present analyses simpler methods are used because calculation of confidence limits is more straightforward, and the lognormal distribution has many of the same qualitative properties as the applicable extreme value distribution.

Modeling the Initial Thickness

The initial thickness assumed is an important feature of the analysis. In Lyon (1995 and 1996) this was dealt with in a simple fashion. In this report, the initial thickness is modeled using a distribution that accounts for the variability in the initial thickness. According to design specifications, the nominal wall thickness when the cylinders were delivered was 312.5 mils, with a design range of 302.5 to 345.5 mils. In this analysis, the initial thickness is approximated using a truncated normal distribution, which is a normal distribution that is defined on a finite range. The parameters for the distribution are estimated based on the data available. These data consist of wall thickness measurements made on the cylinders evaluated in relatively uncorroded regions of the cylinder. These measurements were made using either an automated scanner or a hand-held probe, depending on the particular dataset. This initial thickness estimate does not include any general corrosion that may have occurred across the entire cylinder surface. This is motivated in part by concerns within the program that the variability in initial thickness could be a critical factor (e.g., Rosen and Glaser 1995).

Calculation of Confidence Limits

The method used to calculate confidence limits is discussed in Appendix B. It is stressed that the confidence limits are relevant only when the samples upon which they are based come from a random sample. As such, because few of the datasets currently available come from a random sample, the defensibility of the confidence limits presented here is suspect, except in the case of the PORTS cylinders and perhaps to some degree C-745-G bottom row cylinders, since random sampling was utilized for these data. The confidence limits are presented for all populations in this report because it may assist program management by providing an upper envelope, and implementation of the methods necessary to calculate confidence limit facilitates subsequent modeling when random sampling has been employed for all populations.

4. SUMMARY OF DATA UTILIZED

Using an automated scanner called a P-Scan system (see Schmidt et al. 1996 for a description of the equipment), there have been three efforts that have collected pit depth and wall thickness data for various thin-walled cylinder populations at the PGDP in Paducah, Kentucky, the K-25 site in Oak Ridge, Tennessee, and the PORTS in Portsmouth, Ohio. The first effort was performed during 1994 at K-1066-K yard at K-25. The second was performed during the fall of 1995 at the PGDP, and the third and most recent was conducted between March and September 1996 at both the PORTS and PGDP sites as part of the cylinder relocation efforts. The pit depth data utilized here consist of measurements made with the automated scanner for a square region of width and height of about 2.54 mm (0.1 in). The wall thickness data used for the initial thickness consisted of either data collected with the automated scanner near where the maximum pit occurred (with a width and height of approximately the same size as the pit data), or was collected using a hand-held probe for a circular region with a radius of about 2 mm (0.08 in).

Both "thick-walled" (nominal thickness 625 mils) and "thin-walled" cylinders (nominal thickness 312.5 mils) were evaluated in FY 1996. Only thin-walled cylinders are considered in this report. Application of the methods discussed here to the thick-walled cylinders will be summarized in a subsequent memorandum.

With the exception of the data collected at PORTS and C-745-G at PGDP during 1996, the cylinders selected for measurement were not selected randomly. This seriously undermines the representativeness of the data when making predictions for the unsampled population.

Table 1 provides a summary of the data utilized in this analysis.

Table 1. Summary of thin-walled cylinder data utilized (numbers in parenthesis are the ages, or range of ages, of the cylinders when evaluated).

Site	Yard	Row	Number of Cylinders			
			1992	1994	1995	1996
K-25	K-1066-K	Top	1 (29)*	60 (31-36)		
		Bottom	1 (34)*	55 (31-38)		
	C-745-B	Top			4 (39)	
		Bottom			2 (39)	
	C-745-F	Top			13 (31-36)	
		Bottom			13 (32-36)	6 (36-37)
PGDP	C-745-G	Top			9 (33-36)	137 (18-37)
		Bottom			17 (33-36)	98 (5-37)
	C-745-K	Top			16 (15-18)	
		Bottom			23 (15-19)	6 (16-37)
	C-745-L	Top			1 (13)	
		Bottom			2 (14)	
PORTS	X-745-C	Top				221 (6-40)
		Bottom				252 (6-40)

* These are the two cylinders that were determined to breach from external corrosion (Barber et al. 1994).

K Yard at Oak Ridge

K-1066-K yard, located at the K-25 plant in Oak Ridge Tennessee, contains 2942 cylinders, ranging in age (in 1996) from about 32 years to 39 years. These cylinders were initially stored at K-1066-G yard at Oak Ridge starting at about 1966, and relocated in 1983 (Barber et al. 1994). During the six month period 12/93 to 5/94, pit depth and wall thickness measurements were made for 136 cylinders (Philpot 1995) using an automated scanner. It was intended that the cylinders selected for measurement were chosen at random, although a random number generator was not used to select the cylinders, and there were limitations imposed by the automated scanner (e.g., length of power cord, clearance between adjacent cylinders). For these reasons, it is not possible to objectively conclude that the cylinders selected are a representative sample from the population, although this actually may be the case. For the first 21 cylinders evaluated, only minimum wall thickness data were recorded, while pit depth data were also recorded for the rest of the cylinders. There is concern about the accuracy of the wall thickness data for the first group of cylinders. Further, since no pit depth data were recorded for these cylinders that would allow estimating how much corrosion had occurred, these cylinders are not included in this analysis.

It is noted that the accuracy of the equipment used when these data were collected was such that only increments of 5 mils were recorded for pit depth. As a result, there may be several cylinders with the same pit depth and this cannot be seen in Figure 2 (figures are presented in Appendix A).

Also included in the dataset are two breached cylinders discovered on K-1066-K yard in 1992 (Barber et al. 1994)¹.

C-745-B Yard

This yard contains about 1500 cylinders manufactured between 1954 and 1988. In 1995, six cylinders were inspected, all type 48T and manufactured in the period 1956-57. These cylinders had been stored on 10" high concrete piers above the yard surface since April 1967 (Blue 1995a). Four top row and two bottom row cylinders were evaluated, this particular choice of cylinders being a matter of convenience for the material handlers (Blue 1995a).

C-745-F Yard

C-745-F yard contains approximately 4500 cylinders. The top and bottom rows of this yard were interchanged in 1992 when all bottom row cylinders were put on concrete chocks, rather than wood as had been previously used (each row was also relocated south one row). It is likely that some fraction of the bottom row cylinders were in water contact for extended periods of time, although none are now.

In 1995 twenty-six (26) cylinders were evaluated, with 13 from both the bottom and top rows. Both the pit depths and wall thickness were recorded for these cylinders. It is noted that hand-held measurements

¹There have been five other breaches discovered (two at K-25 in 1992, two at PORTS in 1990, and one at PGDP in 1992), but it was concluded that the breaches were induced by mechanical damage at the time of stacking rather than to external corrosion. In particular, the breaches were caused by a lifting lug of an adjacent cylinder that induced a small crack near a stiffening ring.

using a 2 mm probe were made in 1994 to estimate the minimum wall thickness for 21 cylinders in C-745-F yard, but the pit depths were not recorded (Blue 1994). These data are not included in this analysis because it is not possible to reliably estimate the pit depths. These data were used in the analysis discussed in Lyon (1995) because these were the only data available for this yard at that time.

Six cylinders were also evaluated in 1996 from C-745-F yard.

C-745-G Yard

The population modeled as C-745-G yard actually consists of cylinders that were originally in C-745-G yard but have been relocated to various other nearby yards at PGDP during the past year in preparation of the refurbishment of C-745-G yard. The purpose of this move was to allow conversion of C-745-G yard from a poor gravel yard to a concrete yard with sufficient drainage.

There are two datasets available for C-745-G yard that were utilized, both of which were collected using the automated scanner. The first consists of data for 26 cylinders that were evaluated in 1995 (Blue 1995a, 1995b). The second dataset consists of measurements made between March and September 1996 on cylinders set aside as part of the relocation efforts performed during 1995 and 1996. A total of 235 cylinders were evaluated (137 from the top row, and 98 from the bottom row). These cylinders are a subset of the approximately 390 cylinders set aside from the first 3900 cylinders moved out of the C-745-G yard. Because of the manner in which these cylinders were selected, these cylinders are a systematic sample only from the first 3900 cylinders moved out of G yard. This weakens the statistical defensibility of statements made for the whole C-745-G yard population based on the trends observed for these data.

A painting program was initiated for the cylinders moved from C-745-G to C-745-S yard in May or early June 1996. The cylinders that had been painted (about 1000 cylinders) were not included in the present modeling.

It is noted that hand-held measurements using a 2-mm probe were made in 1994 to estimate the minimum wall thickness for eight cylinders in C-745-G yard, but the pit depths were not recorded (Blue 1994). These data are not included in this analysis because it is not possible to reliably estimate the pit depths. These data were used in the analysis discussed in Lyon (1995) because these were the only data available for this yard at that time.

C-745-K and L Yards

The C-745-K and C-745-L yards contain a total of about 9000 Type OM and G cylinders manufactured in the period 1958-1992. These cylinders have been stored on five-inch concrete saddles in gravel yards constructed with an underground drainage system. Data were collected from these yards in 1995 and 1996. The sampling in 1995 was limited to those cylinders that were manufactured during the period 1976-1982 that had lost large portions of their protective coating (Blue 1995a). A total of 42 cylinders were inspected (39 from K yard, 3 from L yard). Twenty-five cylinders were from the bottom row, and 17 were from the top row. In 1996, 6 cylinders from the bottom row of C-745-K yard were evaluated.

PORTS

There are approximately 14000 thin-walled cylinders ranging in age from a few years to over 40 years old located at the PORTS site. Prior to FY 1996, there were four cylinder storage yards at PORTS. These yards were designated X-745-A, X-745-C, X-745-E, and X-745-F. The X-745-A and X-745-C yards were essentially the same yard, but were separated into different sections. The X-745-C yard had six sections, while the X-745-A yard had three sections. The X-745-A and X-745-C yards had a two tier stacking configuration. The cylinders from the X-745-F yard were single stacked cylinders. The X-745-E yard was a compacted gravel storage area, but was reconstructed during FY 1995-1996 to a reinforced concrete storage yard. In FY 1996, a total of 5708 cylinders were relocated at PORTS to meet the new storage requirements.

Cylinders at PORTS were moved from single row storage to a two tiered arrangement around 1976. Prior to this, there were no top row cylinders at PORTS. The cylinders had been in their current location until movement activities in FY 1996. Thus, the "top" row cylinders at PORTS discussed here have only been in the top row for about 20 years.

In FY 1996, 10% of the cylinders that were relocated were selected using a random number generator to evaluate the wall thickness using ultrasonic thickness (UT) measurements. The 10% evaluation criterion was required based on the Consent Decree with the Ohio Environmental Protection Agency. These cylinders, as well as other cylinders with handling or storage damage, were evaluated using the automated scanner P-Scan system and hand-held measurements. A total of 609 cylinders were evaluated at PORTS in FY 1996.

For the purpose of modeling, all of the thin-walled cylinders at PORTS are considered as one yard, and the top and bottom rows are treated separately. During FY 1996, 473 thin-walled (i.e., nominal thickness 312.5 mils) were evaluated (221 from the top row, 252 from the bottom row), with an age range of 6-40 years.

In Lyon (1995), a different dataset was used for the PORTS site. These data consisted of hand-held ultrasonic thickness measurements made in 1994 on 125 cylinders. These data are not used in the present analysis because the measurements were not taken at areas known to have accelerated corrosion, such as the saddle interface. Further, the evaluation techniques currently used are more stringent and provide more accurate data than that obtained previously.

5. DATA ANALYSIS

Table 2 shows the basic groupings made with the current data. These groupings are then applied to larger populations that include unsampled cylinder populations. The median maximum pit depth for each of the models used is provided in Figure 1 (figures are presented in Appendix A). Scatterplots for the pit depth for each population are shown in Figures 2-7. A discussion of each dataset is provided below.

Table 2. Summary of population datasets and modeling assumptions.

Dataset	Sample Size	Population Dataset Is Used for	Comment	Model for Penetration Depth $P(t)$ [†]	Predicted Median Pit Depth	Model for Initial Thickness ^{††}	$\overline{\ln(t)}$ *	Sum of Squares*
K-1066-K	117	K-1066-K	Narrow age range of cylinders (manufacture dates 1956-1964), many of which were previously in ground contact for at least 15 years while in K-1066-G yard	$t \times \log(0.53, 0.46)$	1.70 t	$N(315, 19.8)$ on [302.5, 340]	NA	NA
C-745-B/F/K/L bottom row	52	C-745-B/C/D/F/K/L/M/N/P bottom row	Bottom row PGDP cylinders that, for the most part, have not been in ground contact	$\log(1.87 + 0.57 \ln t, 0.37)$	$6.5 t^{0.57}$	$N(332, 0.9.5)$ on [302.5, 357]	3.1	7.0
C-745-G bottom row	115	C-745-G bottom row	Many cylinders were in ground contact for extended periods	$\log(1.89 + 0.69 \ln t, 0.37)$	$6.6 t^{0.69}$	$N(336, 3.10.0)$ on [302.5, 363]	3.4	9.4
C-745-B/F/G/K/L top row	180	C-745-B/C/D/F/G/K/L/M/N/P top row	Cylinders primarily in top row for most of their storage history	$\log(2.04 + 0.50 \ln t, 0.25)$	$7.7 t^{0.50}$	$N(332, 8.8.6)$ on [302.5, 356]	3.3	13.0
PORTS top row	221	PORTS and K-1066-B/E/J top row	Cylinders primarily in top row for most of their storage history	$\log(2.48 + 0.39 \ln t, 0.20)$	$11.9 t^{0.39}$	$N(344, 3.12.4)$ on [302.5, 378]	3.4	43.2
PORTS bottom row	252	PORTS and K-1066-B/E/J bottom row	Cylinders primarily in bottom row, but not ground contact, for most of their storage history	$\log(2.63 + 0.41 \ln t, 0.27)$	$13.9 t^{0.41}$	$N(335, 5.13.4)$ on [302.5, 375]	3.3	62.6

* These terms are used in the calculation of upper confidence limits. The term "Sum of Squares" term is defined by $\sum_{i=1}^N (\ln(t_i) - \overline{\ln(t)})^2$, where N is the sample size, t_i is the age of the i th cylinder evaluated, and $\overline{\ln(t)}$ is the mean of the natural logarithm of the ages. How these values are used in determining confidence limits is shown in Appendix C.

[†] Log(a,b) denotes a lognormal distribution with mean and standard deviation of the natural logarithm of the values of a and b , respectively.

^{††} "N(μ, σ) on [a,b]" denotes a normal distribution with mean μ and standard deviation σ defined on the range [a,b].

K-1066-K Yard

This population is treated separately from the other populations. A large portion of these cylinders were in ground contact for extended periods while they were in a previous yard (K-1066-G yard). The narrow age range, coupled with a lack of data for the same cylinders at significantly different times, makes it difficult to determine and/or defend any dependence of corrosion on age. In fact, it may be that the age of the cylinder is no longer relevant for predicting future corrosion. However, given that cylinders were evaluated in 1994 on this yard, it may be possible with additional sampling to determine what this current corrosion rate may be (this is suggested in Lyon and Lykins 1996).

When utilizing these data, there are two features of this dataset that require assumptions to be made: (1) how to incorporate the two cylinders discovered in 1992 that were deemed to breach due to external corrosion (Barber et al. 1994), and (2) whether or not the top and bottom row populations should be modeled separately. In this report, the breaches are included and the top and bottom row populations are combined. More discussion is provided below.

It is not clear how the two breached cylinders discovered in 1992 should best be incorporated in the analyses. One of these cylinders was in the top row and one was in the bottom row at the time they were discovered, but the prior location of these cylinders in K-1066-G yard is not known. Because these two cylinders were not evaluated as part of a random sample, it is natural to deem them as the most extreme cases, and hence exclude them. However, this then omits potentially important information about this population of cylinders; namely, that extremely high corrosion rates can occur. For conservative purposes, it is considered appropriate to include these cylinders in the analysis, and that is what is assumed here.

An analyses of the top and bottom row cylinders suggests that there is little statistical difference between the two populations, although the average corrosion rate is slightly higher for the bottom row cylinders. The main reason that this difference is not larger may be a reflection of the nature of the relocation efforts that were conducted when K-1066-G yard was moved to K-1066-K yard in 1982. In particular, if the top and bottom row cylinders were not kept together, then "different" populations (from a corrosion standpoint) may have been effectively "shuffled", thereby obscuring any row-effects that may have been present.

The statistics for the corrosion rates for the different subpopulations within K-1066-K yard are shown in Table 3.

Table 3. Comparison of statistics for corrosion rate for K-1066-K subpopulations.

		Log of Age-Averaged Corrosion Rate	
		Sample Size	Mean, St.Dev.
Top Row	without breach	60	0.48, 0.40
	with breach	61	0.51, 0.47
Bottom Row	without breach	55	0.53, 0.39
	with breach	56	0.56, 0.45
Top and Bottom Row	without breaches	115	0.50, 0.39
	with breaches	117	0.53, 0.46

Using the *F*-test (see, e.g., Snedecor and Cochran 1978, pp.116-117), one can accept that the variances of the logarithms of the pit depths for the top and bottom row populations are equal with 5% significance, whether or not the breaches are included. Similarly, using the same test as described in Lyon (1996), namely the *t*-test with unequal variances (Casella and Berger 1990), one can conclude that the medians of the distributions are the same; i.e., there is not a statistically significant difference between the medians of the distributions for the top and bottom row cylinders. Both of these results support treating the top and bottom rows as a single population, whether or not the breaches are included.

When a power law is fit to the maximum pit depths, the resulting model is unrealistic, whether or not the breaches are included: the median maximum pit depth would then be given by $P(t)=0.5 t^{1.4}$ (with breaches included) or $P(t)=0.1 t^{1.7}$ (without breaches included). This would imply that the corrosion rate is *increasing* for these cylinders at a very high rate, which seems unlikely (note that an example has been found in the literature where a value above 1 is reported², but the conditions do not seem relevant).

For the present analysis, the same method for pit depth discussed in Lyon (1995, 1996) is used for the K-1066-K yard cylinders. With this method, the distribution of penetration depth is given by $P(t)=R t$, where R is the distribution of age-averaged corrosion rates. This method is most applicable for short-term prediction due to the uncertainty about the current corrosion rates and the narrow age range for the cylinders: any model that would fit the current data should agree for near-term predictions since they must agree with the data available. If long-term predictions are a priority for this yard, then it is imperative that advantage be taken of the fact that cylinders were evaluated in 1994 and hence the current corrosion rates may be estimated. This would possibly eliminate much of the uncertainty regarding long-term prediction for this population, assuming that the yard conditions do not substantially change. This is not necessary if it is rather certain that the condition of the cylinders is going to be improved in the near future (e.g., by painting).

C-745-B/F/K/L Yards, Bottom Row

This population is used to represent cylinders at PGDP that were not in ground contact for extended periods. The dataset consist of two basic age groups: cylinders between 11 and 18 years old (from K and L yards), and cylinders between 31 and 39 years old (from B and F yards). If data for older cylinders were available for K or L yard bottom row (or any of the other better condition yards, such as M or N yard), then the F yard cylinders would not be included in this dataset. Use of the F yard bottom row cylinders is conservative because some may have been in water contact, and hence more corrosion may have occurred. C-745-G yard, which was the only other population of cylinders from PGDP for which data are available, was not considered appropriate due to the poorer conditions of the yard compared to the C-745-F yard.

A power law model that fits (via least squares) the pit depth data is $\text{Log}(1.87 + 0.57 \log t, 0.37)$, which has a median predicted pit depth of $6.5 t^{0.57}$.

² A value above 1 was reported for galvanized steel in an industrial environment in Chicago Pourbaix (1982).

C-745-G Yard, Bottom Row

The C-745-G yard cylinders represent the worst conditions at the PGDP site. Many of these cylinders were in ground contact for extended periods. Unlike K-1066-K yard at Oak Ridge, there is a wide range of ages for these cylinders, and cylinders from the entire range were evaluated as part of the most recent data collection efforts.

A power law model that fits (via least squares) the pit depth data is $\text{Log}(1.89 + 0.69 \log t, 0.37)$, which has a median predicted pit depth of $6.6 t^{0.69}$. This is interesting when compared to the other PGDP bottom row cylinders because the “spread” of both distributions and the predicted first year corrosion are essentially the same. The only difference is the power, with the higher power for C-745-G yard bottom row due to the poorer conditions for this yard.

C-745-B/F/G/K/L Yards, Top Row

Few, if any, of the cylinders in the top rows of these yards were ever in extended ground contact, and this is assumed to be the case for all of the top row cylinders at PGDP (note that what is modeled here as the C-745-F top row cylinders are currently in the bottom row, and vice versa, due to the relocation that took place in 1992).

A power law model that fits (via least squares) the pit depth data is $\text{Log}(2.04 + 0.50 \log t, 0.25)$, which has a median predicted pit depth of $7.7 t^{0.50}$. The power 0.50 is smaller than the similar power for both of the bottom row populations at PGDP discussed above, indicative of the lower corrosion rates for these cylinders.

PORTS, Top Row

As for the top row cylinders at PGDP, few of the PORTS top row cylinders have ever been in extended ground contact. The data collected for the top row cylinders at PORTS are used to represent both the PORTS top row cylinders as well as the top row cylinders in the K-1066-B/E/J yards at the K-25 site at Oak Ridge until data for the K-25 yards are collected.

A power law model that fits (via least squares) the pit depth data is $\text{Log}(2.48 + 0.39 \log t, 0.20)$, which has a median predicted pit depth of $11.9 t^{0.39}$. The power 0.39 is smaller than the similar power for the top row population at PGDP discussed above, and this implies a lower corrosion rate eventually; however, the higher value of 11.9 can result in higher predicted corrosion in the short-term.

PORTS, Bottom Row

Few of the PORTS bottom row cylinders have ever been in extended ground contact. The data collected for the bottom row cylinders at PORTS are used to represent both the PORTS bottom row cylinders as well as the bottom row cylinders in the K-1066-B/E/J yards at the K-25 site at Oak Ridge until data for the K-25 yards are collected.

A power law model that fits (via least squares) the pit depth data is $\text{Log}(2.63 + 0.41 \log t, 0.27)$, which has a median predicted pit depth of $13.9 t^{0.41}$. The power is similar to that for the top row cylinders at PORTS, although the power is slightly higher for the bottom row cylinders (0.41 versus 0.39), as is to be expected.

The difference between the power curves for the PORTS and PGDP cylinders bears noting. The higher value for the "A" term in $P(t)=At^n$ for the PORTS cylinders implies a higher predicted corrosion rate early in the life of the cylinder, while the lower value for "n" implies that the long-term corrosion rate will be smaller for the PORTS cylinders (see Figure 1). Because this behavior occurs for both the top and bottom row cylinder populations when comparing across sites, there may be something peculiar to one of the sites that can account for this interesting difference. At this time it is not clear what specific factors could defensibly explain these interesting differences, and it is suggested that further statistical analyses be conducted to verify that these differences are significant.

6. RESULTS

For each cylinder population considered, estimates were made of the number of cylinders as a function of time that will have a minimum point wall thickness of either 0 mils (a breach), 62.5 mils, or 250 mils. These wall thicknesses (0, 62.5 mils, and 250 mils) used in this report are preliminary boundaries identified within the program that indicate loss of material, safe handling and stacking operations and safe off-site transport and contents transfer criteria respectively. In general, these safety criteria are based on area of wall thinning. However, the minimum thickness predicted in this report is for a region with an area of about 0.01 sq. in, because this is the type of data used. Using minimum point thickness adds a considerable degree of conservatism to the results in this report.

Table 4 shows the cumulative predicted number of cylinders that have a minimum thickness below 250 mils, 62.5 mils, and breached cylinders as a function of time for the six populations from which the data are taken. Graphs of these results are shown in Figures 8-10. Table 5 shows the results for all populations of cylinders; these results are based on grouping unsampled cylinder populations with similar populations that have been sampled (e.g., K-1066-B yard is added to the PORTS population). There is generally not a big difference between the results in Tables 4 and 5 because the unsampled yards contain mostly younger cylinders.

The numbers in the columns labeled "Estimate" are the expected values based on the maximum likelihood estimate of the parameter for K-1066-K yard and the least squares estimate for the other populations. The confidence limits are upper bounds on the upper 95% confidence limit. How these confidence limits are calculated is described in Appendix B. For the populations other than the K-1066-K yard cylinders, the difference between the upper confidence limits depends on two factors: the total number of cylinders sampled, and the ages of the cylinders sampled. Basically, the larger the "spread" in ages of the cylinders sampled, the smaller the difference between the confidence limits and the estimated values. The confidence limits for the K-1066-K yard cylinders depend on the number of samples but not on the spread of the ages of the cylinders sampled.

From Figure 8, the results for predicted cylinders with minimum thickness below 250 mils show that the populations can be naturally divided into three groups: (1) K-1066-K yard and C-745-G yard bottom row, (2) C-745-B/F/K/L and PORTS bottom row, and (3) C-745-B/F/G/K/L and PORTS top row. The results for

the number of cylinders with minimum thickness below 62.5 mils or breaches (Figures 9 and 10) are similar, although there is a more pronounced difference between the K-1066-K and C-745-G yard bottom row cylinders.

For two cylinder populations, K-1066-K at K-25 and the cylinders that were in the bottom row of C-745-G at PGDP, a large number of cylinders are predicted to have a minimum point thickness below 250 mils at present. Both of these populations have had a large fraction of cylinders that were in ground contact at one time, although they are no longer in ground contact. For the other cylinder populations, bottom row cylinders are generally predicted to have a greater number less than 250 mils than top row cylinders.

Table 4. Summary results for cylinder populations for which data are available.

Population	Sample	Predicted number of cylinders with a minimum point thickness below 250 mils										2020
		1996	2000	2004	2008	2012	2016	Upper CL*	Estimate	Upper CL*	Estimate	
K25 K-1066-K yard top and bottom rows(2942 cylinders)	117	1281	1980	1537	2201	1767	2375	1967	2511	2139	2614	2233
PGDP C-745-G yard bottom row(3167 cylinders)	115	1109	1881	1345	2129	1571	2345	1780	2528	1969	2677	2137
PGDP C-745-B/F/K/L yards bottom row(7144 cylinders)	52	241	1249	349	1631	486	2077	652	2567	844	3082	1056
PGDP C-745-B/F/G/K/L yards top row(10102 cylinders)	180	70	283	113	443	172	659	250	914	350	1200	472
PORTS thin-walled bottom row(6681 cylinders)	252	283	621	390	830	521	1074	676	1332	850	1598	1039
PORTS thin-walled top row(6681 cylinders)	221	31	111	48	162	71	227	101	301	140	404	189
Total [†]	937	3016	3782	4588	5427	6291	7176	7176	8076	8076	8076	8076
Predicted number of cylinders with a minimum point thickness below 62.5 mils												
Population	Sample	1996	2000	2004	2008	2012	2016	Upper CL*	Estimate	Upper CL*	Estimate	Upper CL*
		Size	Estimate	Upper CL*								
K25 K-1066-K yard top and bottom rows(2942 cylinders)	117	3	35	7	50	13	71	22	106	34	151	51
PGDP C-745-G yard bottom row(3167 cylinders)	115	1	19	2	37	4	42	7	61	11	96	16
PGDP C-745-B/F/K/L yards bottom row(7144 cylinders)	52	0	9	0	22	0	35	0	48	0	64	0
PGDP C-745-B/F/G/K/L yards top row(10102 cylinders)	180	0	0	0	0	0	0	0	0	0	0	0
PORTS thin-walled bottom row(6681 cylinders)	252	0	0	0	0	0	0	0	0	0	0	0
PORTS thin-walled top row(6681 cylinders)	221	0	0	0	0	0	0	0	0	0	0	0
Total [†]	937	5	9	17	28	45	67	67	96	96	96	96
Predicted number of breached cylinders												
Population	Sample	1996	2000	2004	2008	2012	2016	Upper CL*	Estimate	Upper CL*	Estimate	Upper CL*
		Size	Estimate	Upper CL*								
K25 K-1066-K yard top and bottom rows(2942 cylinders)	117	1	7	1	16	3	31	5	41	9	59	14
PGDP C-745-G yard bottom row(3167 cylinders)	115	0	3	0	5	1	10	1	21	2	36	3
PGDP C-745-B/F/K/L yards bottom row(7144 cylinders)	52	0	1	0	1	0	3	0	9	0	20	0
PGDP C-745-B/F/G/K/L yards top row(10102 cylinders)	180	0	0	0	0	0	0	0	0	0	0	0
PORTS thin-walled bottom row(6681 cylinders)	252	0	0	0	0	0	0	0	0	0	0	0
PORTS thin-walled top row(6681 cylinders)	221	0	0	0	0	0	0	0	0	0	0	0
Total [†]	937	1	2	3	6	11	17	17	26	26	26	26

*These are bounds on the upper 95% confidence limit. The details about how they are calculated are provided in Appendix B.

[†]Apparent discrepancies between the totals and the summands are due to the fractional contributions of the summands.

Table 5. Summary results for all cylinder populations.

Population	Sample	Predicted number of cylinders with a minimum point thickness below 250 mils										Predicted number of cylinders with a minimum point thickness below 62.5 mils													
		1996	1996	Estimate	Upper CL*	2000	2000	Estimate	Upper CL*	2004	2004	Estimate	Upper CL*	2008	2008	Estimate	Upper CL*	2012	2012	Estimate	Upper CL*	2016	2016	Estimate	Upper CL*
K-1066-K yard, top and bottom rows(2942 cylinders)	117	1281	1980	1337	2201	1767	2375	1967	2511	2139	2614	2283	2693	2403	2153										
C-745-G yard, bottom row(3167 cylinders)	115	1109	1881	1345	2129	1571	2345	1780	2528	1969	2677	2137	2796	2284	2889										
C-745-B/C/D/F/K/L/M/N/P yards, bottom row(10665 cylinders)	52	300	1594	434	2096	610	2694	829	3373	1089	4113	1385	4878	1709	5638										
C-745-B/C/D/F/K/L/M/N/P/G yards, top row(13623 cylinders)	180	78	318	126	499	193	742	281	1035	395	1377	538	1821	712	2312										
K-1066-B/E/J yards and all PORTS yards, bottom row(8212 cylinders)	252	363	803	502	1067	671	1373	869	1701	1088	2037	1324	2405	1572	2774										
K-1066-B/E/J yards and all PORTS yards, top row(8212 cylinders)	221	40	140	61	205	90	292	129	392	179	522	242	675	318	843										
Total [†]	937	3171	4005	4902	5855	5855	6859	7909	7909	8997	8997	8997	8997	8997	8997	8997	8997	8997	8997	8997	8997	8997	8997	8997	8997
Population	Sample	Predicted number of breached cylinders										Predicted number of breached cylinders										2020			
		1996	1996	Estimate	Upper CL*	2000	2000	Estimate	Upper CL*	2004	2004	Estimate	Upper CL*	2008	2008	Estimate	Upper CL*	2012	2012	Estimate	Upper CL*	2016	2016	Estimate	Upper CL*
K-1066-K yard, top and bottom rows(2942 cylinders)	117	3	35	7	50	13	71	22	106	34	151	51	206	73	271										
C-745-G yard, bottom row(3167 cylinders)	115	1	19	2	37	4	42	7	61	11	96	16	140	23	188										
C-745-B/C/D/F/K/L/M/N/P yards, bottom row(10665 cylinders)	52	0	12	0	27	0	43	0	59	0	78	0	94	0	117										
C-745-B/C/D/F/K/L/M/N/P/G yards, top row(13623 cylinders)	180	0	0	0	0	0	0	0	0	0	0	0	0	0	0										
K-1066-B/E/J yards and all PORTS yards, bottom row(8212 cylinders)	252	0	0	0	0	0	0	0	0	0	0	0	0	0	0										
K-1066-B/E/J yards and all PORTS yards, top row(8212 cylinders)	221	0	0	0	0	0	0	0	0	0	0	0	0	0	0										
Total [†]	937	5	9	17	28	45	45	67	67	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	

^{*}These are bounds on the upper 95% confidence limit. The details about how they are calculated are provided in Appendix B.[†]Apparent discrepancies between the totals and the summands are due to the fractional contributions of the summands.

Despite the fact that a more sophisticated approach is used for the C-745-G yard bottom row cylinders, the predicted median maximum pit depth is higher than that for the K-1066-K yard cylinders for ages below about 60 years. Given that the oldest cylinders on both yards are about 40 years old (in 1996), this means that the median predicted pit depth will be higher for the C-745-G yard cylinders until about 2020. However, the available data also suggest that the K-1066-K yard cylinders are, on average, about 18 mils thicker on uncorroded areas. Since the estimate of the initial thickness is based on the wall thickness on uncorroded areas of the cylinder, this results in slightly fewer thinner cylinders predicted for the C-745-G yard bottom row than for K-1066-K yard.

The impact of the small sample size is shown when the results for the C-745-B/F/K/L bottom row cylinders (52 samples) are compared to the results for the PORTS bottom row cylinders (252 samples) in Table 4. For example, although the expected number of cylinders with a minimum point thickness below 62.5 mils or a breach is zero for both these populations through the year 2020, the upper confidence limits for the C-745-B/F/K/L population are significantly larger than that for the PORTS cylinders. Generally, the confidence limits will increase with increasing time in a manner that depends on the number of samples and the spread of the ages of the cylinders evaluated. Increasing either the sample size or the spread of the ages will result in a decrease in the difference between the expected value and the upper confidence limit.

7. UNCERTAINTIES AND LIMITATIONS

The purpose of this report is to estimate the extent of corrosion on populations of cylinders as a function of time, and this requires that assumptions be made regarding the dependence of the distribution of corrosion rates on time. The data currently available consist of wall thickness measurements made on different cylinders of different ages. There are no data from the same cylinder at substantially different times. This makes defensibly using more realistic and less conservative models difficult, and one is forced to either make conservative assumptions regarding the change of corrosion with time or to use measurements from different cylinders at different ages to estimate corrosion with time. Implicit in any trend derived from the current data is the assumption that the older cylinders had similar corrosion as the younger cylinders when they were younger; i.e., the distribution of pit depths for 10 year old cylinders in a given population is the same no matter when it is measured. This assumption is unavoidable without data for a fixed cylinder at different ages.

Another area of uncertainty is that the storage conditions have changed, thereby altering the corrosion characteristics of the steel. This complicates predicting corrosion when the cylinders have been moved numerous times throughout their history.

The results are also based on the assumption that the data are representative of the population from which they are taken, which is difficult to ascertain for some populations (e.g., data for C-745-B and L yard cylinders). The PORTS data are probably most representative because they were randomly sampled when the cylinders were relocated during FY 1996. Implementation of a statistically-based sampling plan could reduce this uncertainty.

8. CONCLUSIONS AND RECOMMENDATIONS

The United States Department of Energy (DOE) currently manages depleted uranium hexafluoride that is stored in approximately 50,000 carbon steel cylinders located at three DOE sites. Using either a hand-held ultrasonic transducer or an automated scanner, wall thickness and corrosion pit depth data have been collected for several subpopulations of cylinders. Most recently, these data were collected between March 1996 and September 1996 at the PGDP and PORTS sites.

In this report, the most recently collected wall thickness data were utilized, along with previously collected data, to make projections about the condition of the cylinders located at K-25, PGDP, and PORTS. The data now available for some populations allow the use of methodologies that more realistically address the behavior of the corrosion process with time. The results presented here are intended to supersede those presented previously (Lyon 1995,1996). The projections made in this report are based on the trends noted to date continuing. Some activities planned may substantially reduce the corrosion, in which case the results presented here are conservative.

Most of the cylinders predicted to have a minimum point thickness less than 250 mils in 1996 are located in K-1066-K yard at K-25 or were in the bottom row of C-745-G yard at PGDP. Of the few cylinders predicted to have a minimum thickness less than 62.5 mils or have a point breach in 1996 (using expected values), all are in these two yards. Both of these yards have had a large fraction of cylinders that were in ground contact at one time, although they are no longer in ground contact. A breach is predicted in 1996 using conservative upper confidence limits in a different cylinder population at PGDP (the PGDP bottom row population that is treated separately from C-745-G yard); however, this may be due to the small sample size for this population.

Recommendations

1. It is critical that a sampling plan be implemented that specifically documents the questions that the data collected are to address, and how well the answers must be known. An initial plan is discussed in Lyon and Lykins (1996), which could serve as a basis for future plans.
2. If long-term estimates are necessary for K-1066-K yard, then sampling during FY1997 or FY1998 could substantially improve the defensibility of the modeling by allowing determination of the current corrosion rates for this unique yard. In particular, some fraction of the cylinders evaluated in 1994 could be reevaluated to determine the amount of corrosion that has occurred during the last few years.
3. There are few data for many of the older populations at PGDP, and sampling should be directed towards these populations in order to better characterize their condition.
4. The values in this report are based on the assumption that the historical trends will continue, and thus represent all baseline projections. Many of the yards are being improved, in which case the corrosion rates will probably be reduced. It is suggested that future analyses incorporate these changes, if they can be defensibly quantified. This will require additional discussion with the site technical personnel.

5. It is suggested that further statistical analyses be conducted to determine if the differences in the corrosion trends observed at the PORTS and PGDP sites are statistically significant, and to determine what factors may account for these differences if they are significant.

9. REFERENCES

Barber, E.J., J.H. DeVan, J.M. Goigin, and M.S. Taylor (1994). *Investigation of Breached Depleted UF₆ Cylinders at the K-25 Site*. ORNL/TM-12840, Oak Ridge, Tennessee, October 1994.

Blue, S.C. (1994). Facsimile from S.C. Blue to M. Taylor, October 27, 1994.

Blue, S.C. (1995a). Memorandum from S.C. Blue to A.K. Balding, September 25, 1995.

Blue, S.C. (1995b). Facsimile from S.C. Blue to B.F. Lyon, October 30, 1995.

Casella, G. And R.L. Berger (1990). *Statistical Inference*. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, California.

Feliu, S., M. Morcillo and S. Feliu, Jr. (1993a). The Prediction of Atmospheric Corrosion from Meteorological and Pollution Parameters--II. Annual Corrosion in *Corrosion Science*, Vol. 34, No.3, pp.403-414.

Feliu, S., M. Morcillo and S. Feliu, Jr. (1993b). The Prediction of Atmospheric Corrosion from Meteorological and Pollution Parameters--II. Long-term Forecasts in *Corrosion Science*, Vol. 34, No.3, pp.415-422.

Horton, J.B. (1964). The Composition, Structure and Growth of the Atmospheric Rust on Various Steels. Thesis Lehigh University 1964, Bethlehem, PA.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.

Legault, R.A. and G. Preban (1975). *Corrosion-NACE*, Vol. 31, p.117.

Lockheed Martin Energy Systems, Inc. (LMES) (1996a). *UF₆ Cylinder Program System Requirements Document (SRD)*. K/TSO-001, Rev.2. April 1996. U.S. Department of Energy.

Lockheed Martin Energy Systems, Inc. (LMES) (1996b). *UF₆ Cylinder Program System Engineering Management Plan (SEMP)*. K/TSO-017. March 1996. U.S. Department of Energy.

Lyon, B.F. (1995). *Prediction of External Corrosion for UF₆ Cylinders: Results of an Empirical Method*, ORNL/TM-13012. June 1995. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Lyon, B.F. (1996). *Prediction of External Corrosion for Steel Cylinders at the Paducah Gaseous Diffusion Plant: Application of an Empirical Method*, ORNL/TM-13192. February 1996. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Lyon, B.F. and M.L. Lykins (1996). *Ultrasonic Thickness Sampling Plan for the Depleted Uranium Hexafluoride Program*, ORNL/TM-13280. July 1996. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Mughabghab, S.F. and T.M. Sullivan (1989). Evaluation of the Pitting Corrosion of Carbon Steels and Other Ferrous Metals in Soil Systems. *Waste Management*, Vol. 9, pp. 239-251.

Philpot, H.E. (1995). Memorandum from Halen Philpot to Valerie Newman, February 17, 1995.

Pourbaix, M. (1982). The Linear Bilogarithmic Law for Atmospheric Corrosion, in *Atmospheric Corrosion* (W.H. Ailor, ed.). John Wiley and Sons, New York.

Romanoff, M. (1957). Underground Corrosion. *NBS Circular 579*, National Bureau of Standards, Washington, D.C.

Rosen, R.S. and R.E. Glaser (1995). Letter to C.E. Bradley from R.S. Rosen and R.E. Glaser, November 8, 1995.

Rosen, R.S. and R.E. Glaser (1996). Letter to B.F. Lyon from R.S. Rosen and R.E. Glaser, April, 10, 1996.

Schmidt, M.A., J.K. Harper, and J.A. Broders (1996). *A Comparison of Wall Thickness Measurements on UF₆ Cylinders using Scanning vs Hand-held Ultrasonic Probes*. K/TSO-019, November 1996. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Snedecor, G.W. and W.G. Cochran. (1978). *Statistical Methods*, Sixth Edition, Iowa State University Press, Ames, Iowa.

APPENDIX A: FIGURES

Fig. 1. Predicted median maximum pit depths	A-2
Fig. 2. Measured pit depths for K-1066-K yard top and bottom rows	A-3
Fig. 3. Measured pit depths for C-745-G yard bottom row	A-4
Fig. 4. Measured pit depths for C-745-B/F/K/L yards bottom row	A-5
Fig. 5. Measured pit depths for C-745-B/F/G/K/L yards bottom row	A-6
Fig. 6. Measured pit depths for Portsmouth yards bottom row	A-7
Fig. 7. Measured pit depths for Portsmouth yards top row	A-8
Fig. 8. Predicted number of cylinders with minimum point thickness below 250 mils	A-9
Fig. 9. Predicted number of cylinders with minimum point thickness below 62.5 mils	A-10
Fig. 10. Predicted number of cylinders with a point breach	A-11

Fig. 1. Predicted median maximum pit depths.

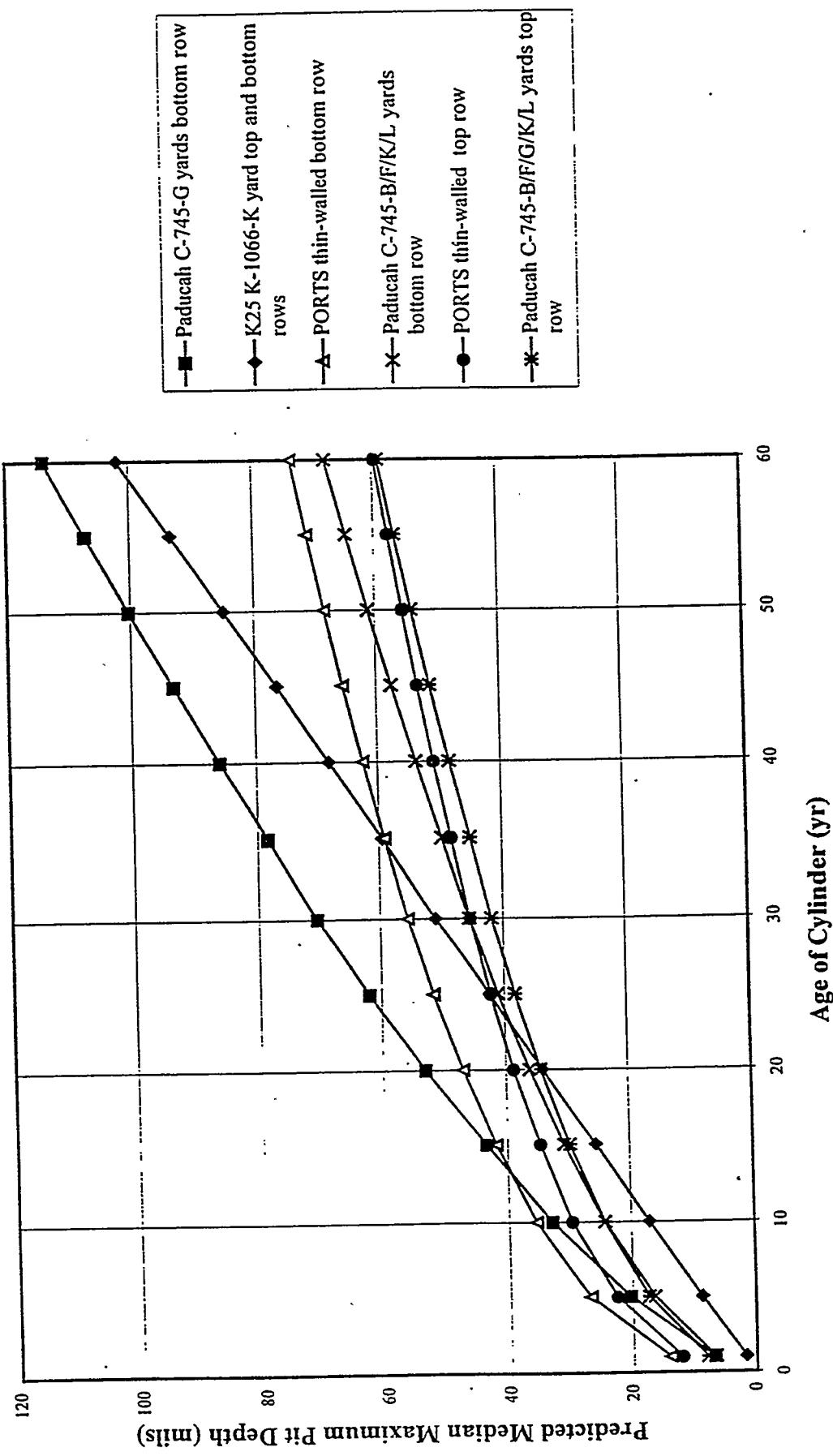


Fig. 2. Measured pit depths for cylinders in K-1066-K yard top and bottom rows. Total of 117 cylinders.

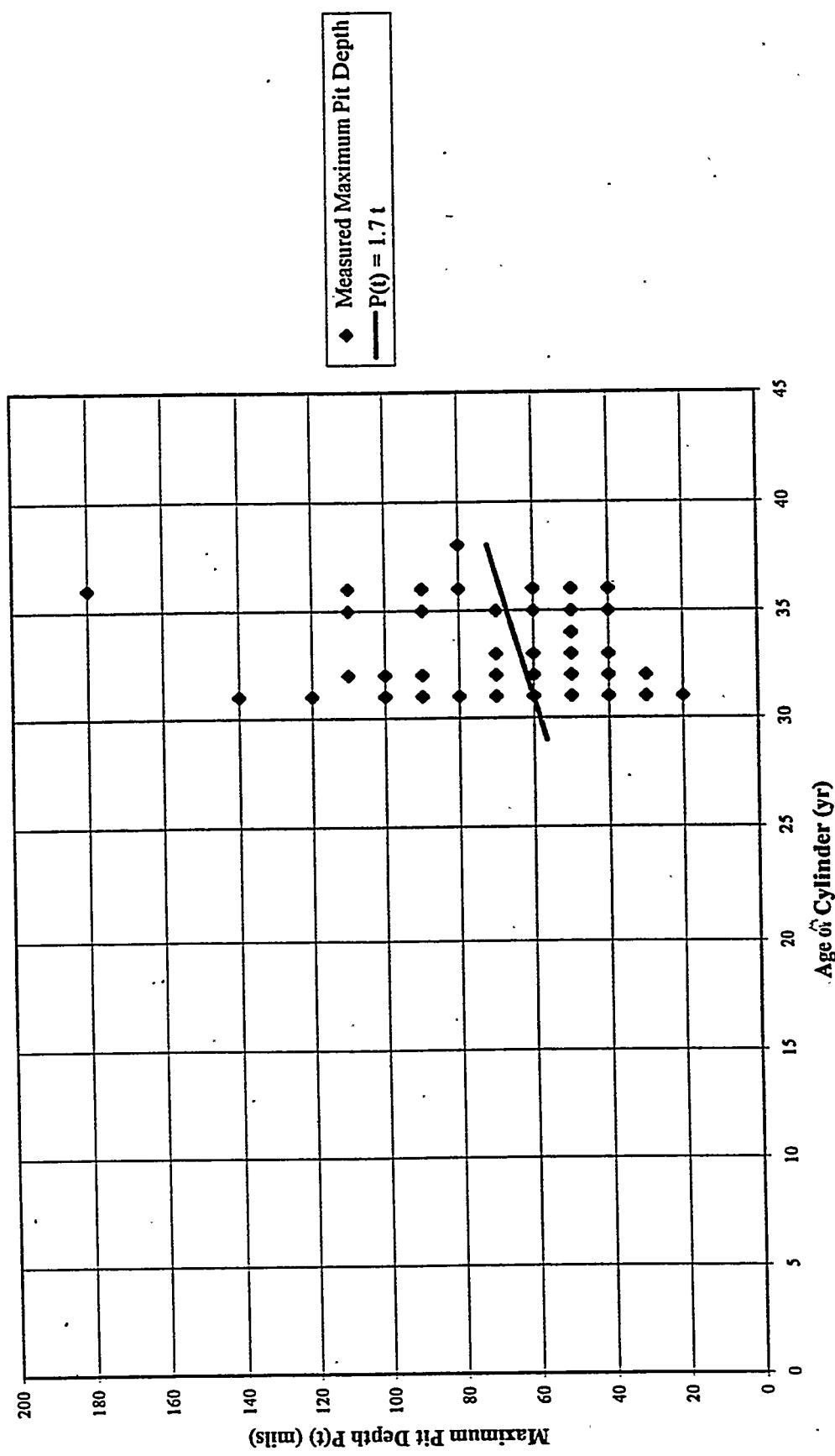


Fig. 3. Measured pit depths for cylinders in C-745-G yards bottom row. Total of 115 cylinders.

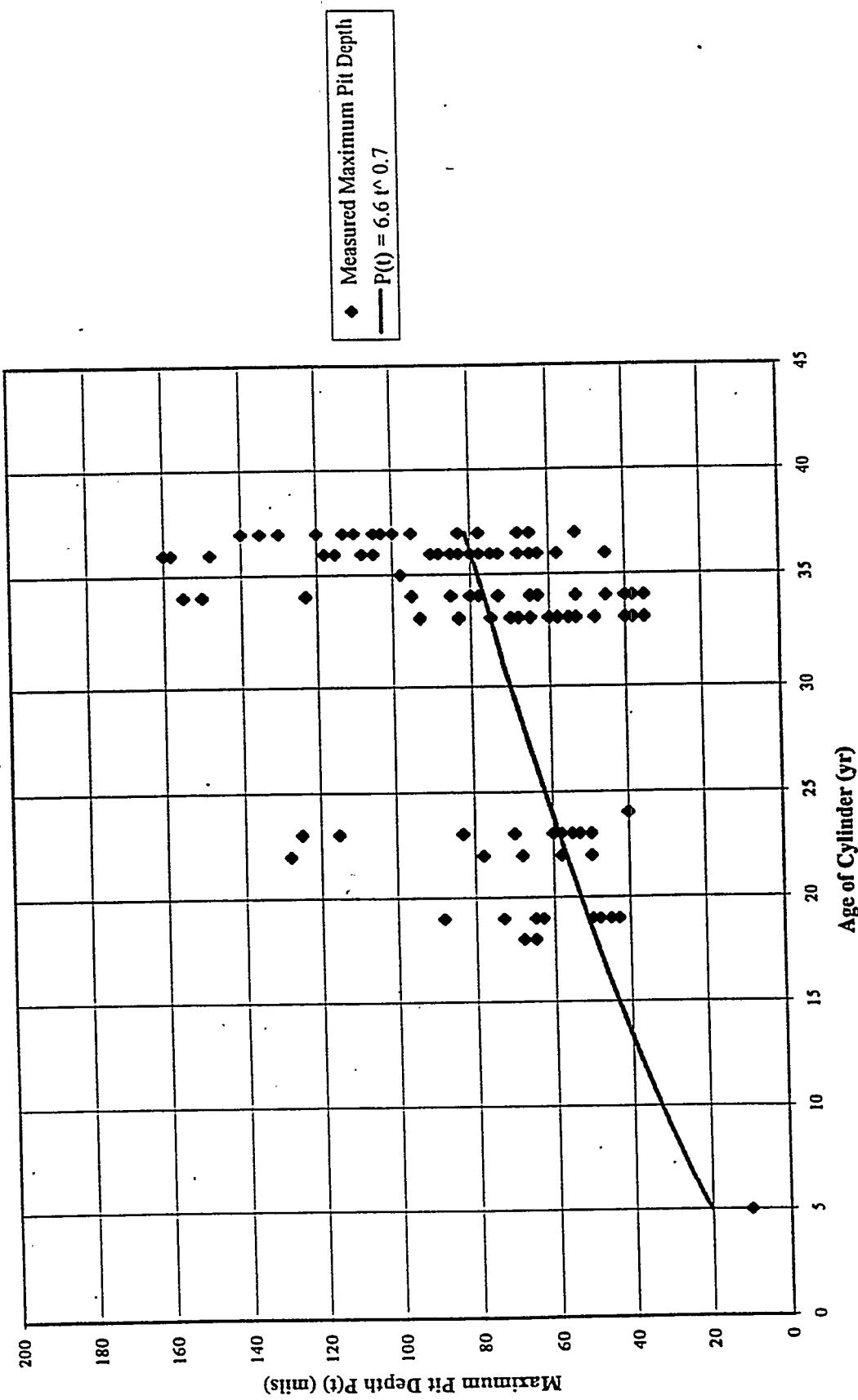


Fig. 4. Measured pit depths for cylinders in C-745-B/F/K/L yards bottom row. Total of 52 cylinders.

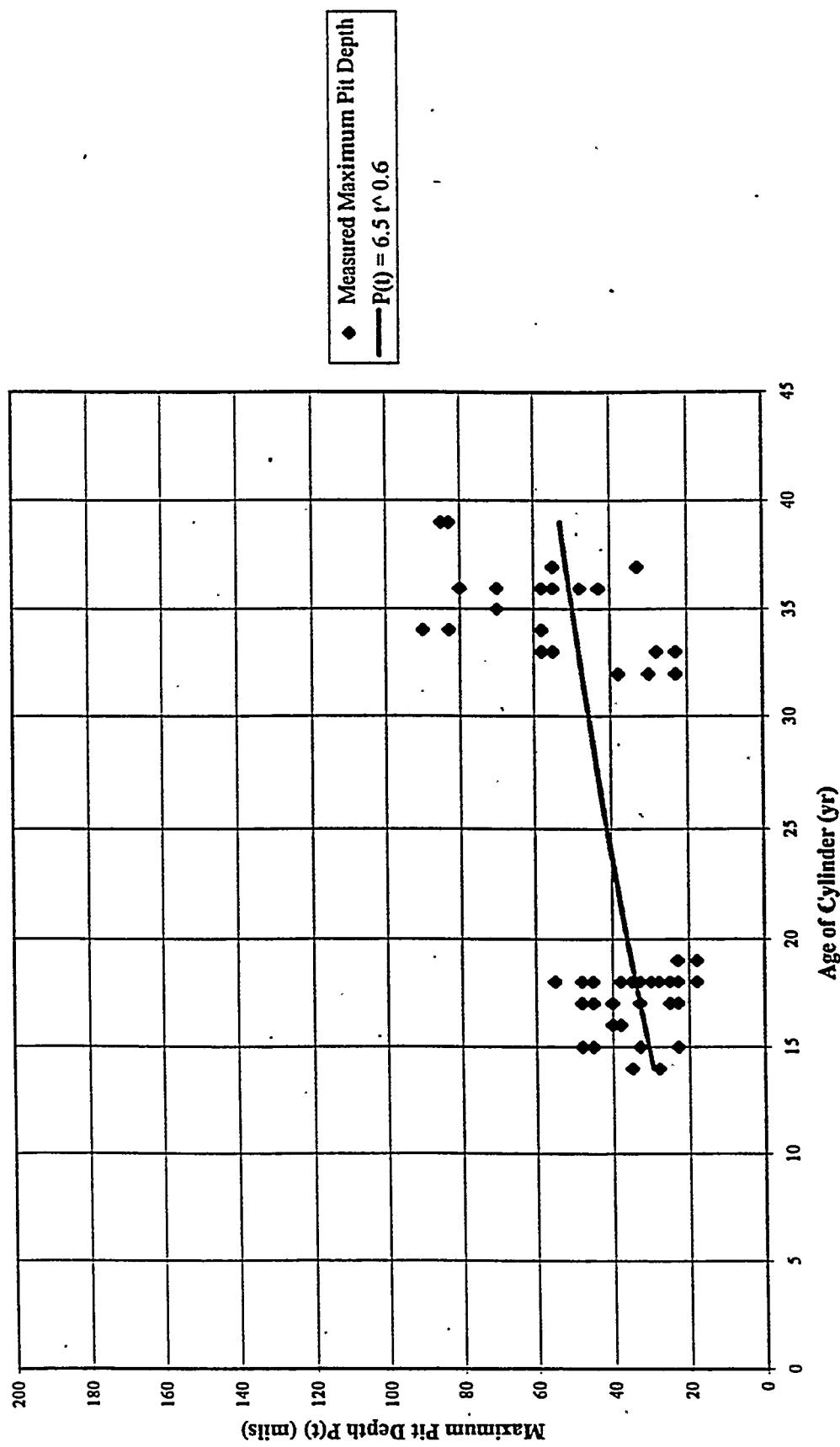


Fig. 5. Measured pit depths for cylinders in C-745-B/F/G/K/L yards top row. Total of 180 cylinders.

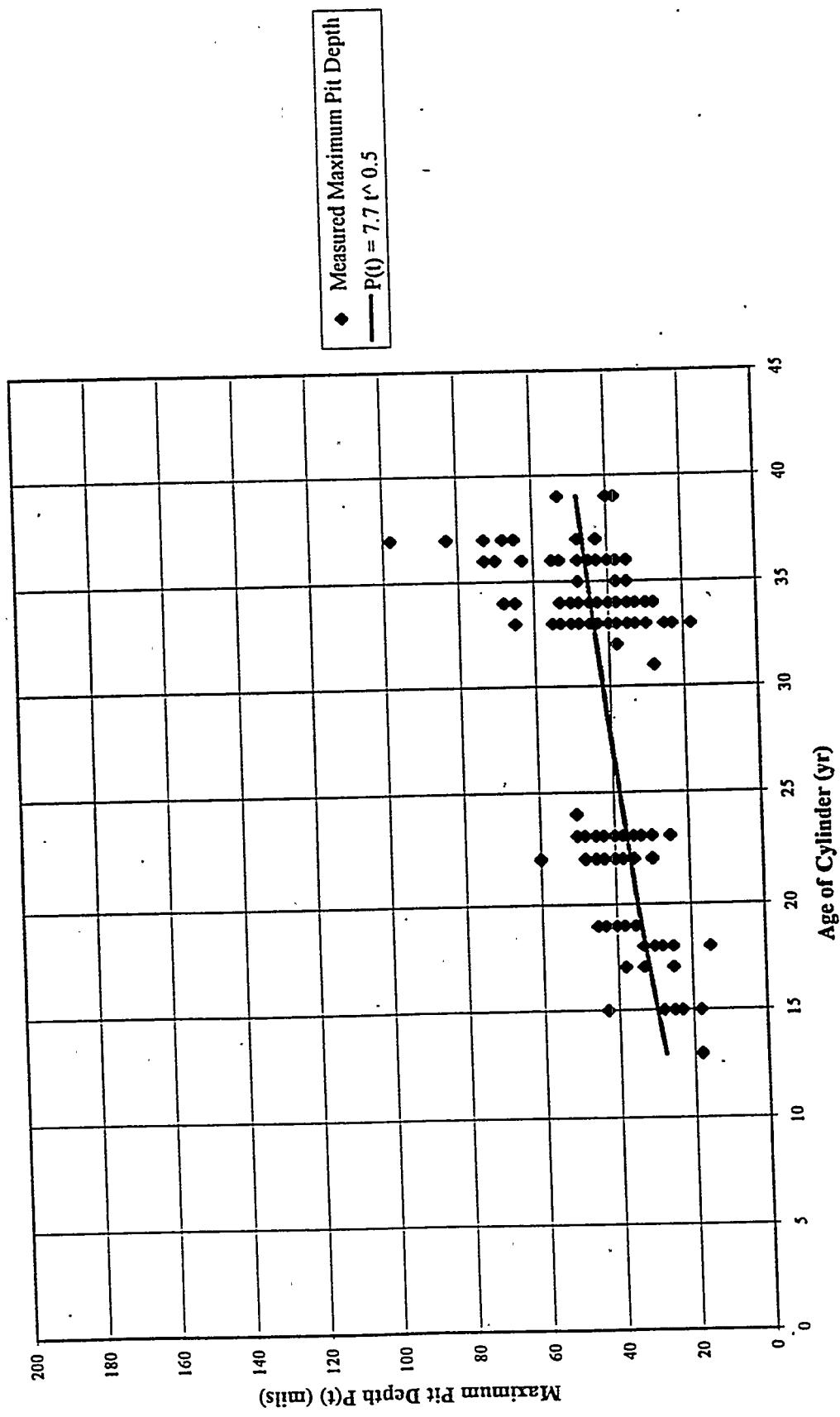


Fig. 6. Measured pit depths for cylinders in PORTS thin-walled bottom row. Total of 252 cylinders.

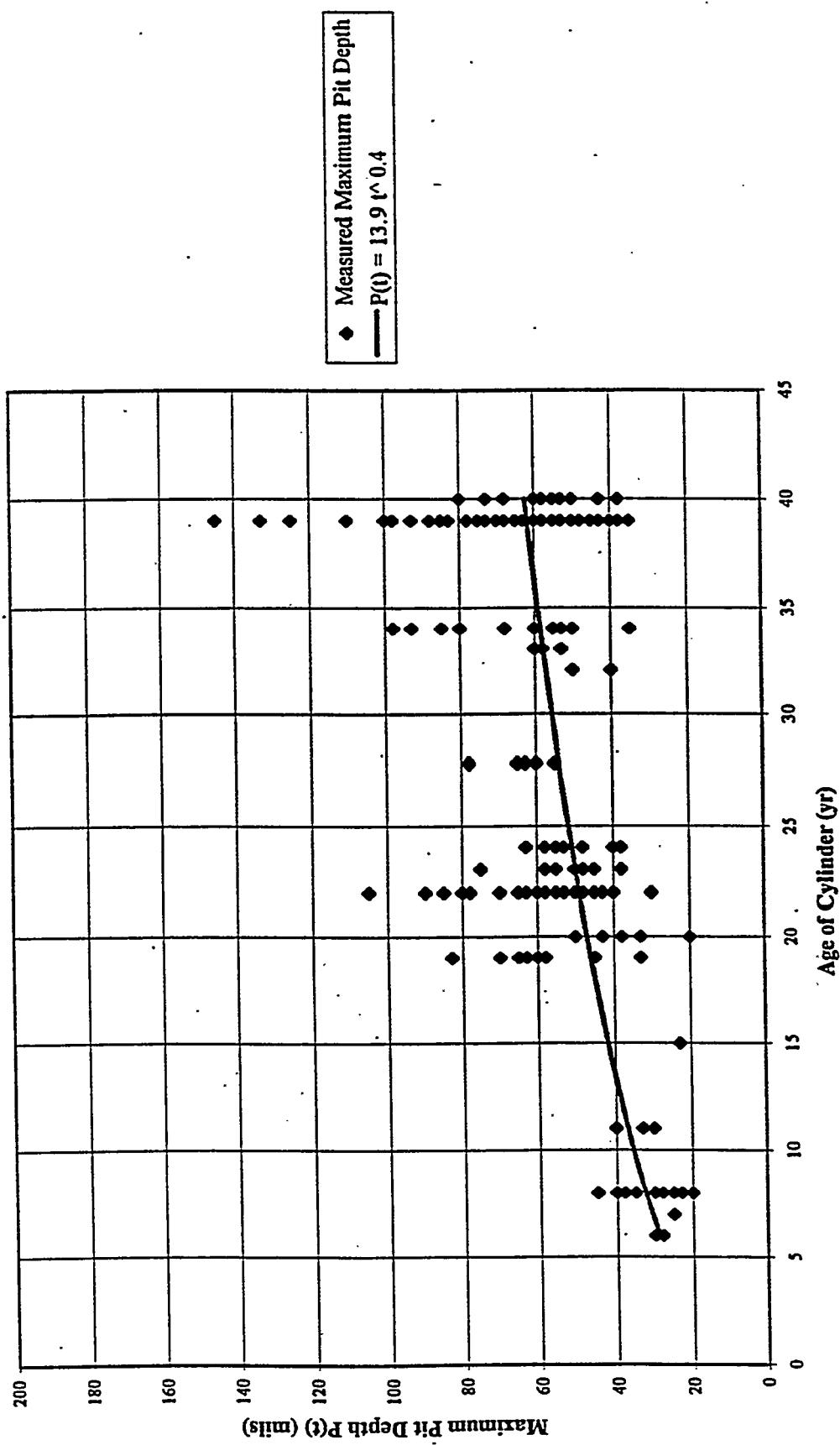


Fig. 7. Measured pit depths for cylinders in PORTS thin-walled top row. Total of 221 cylinders.

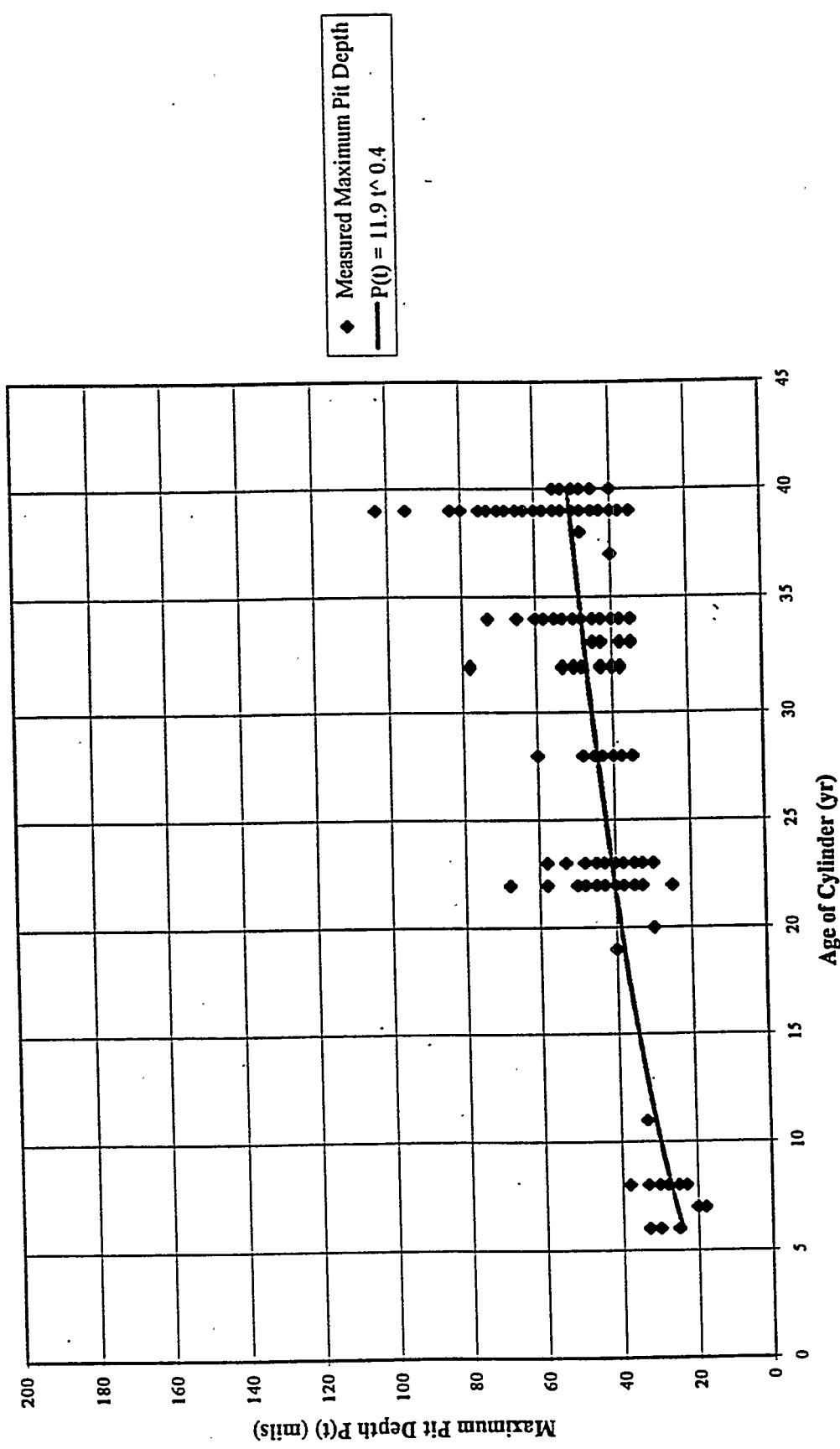


Fig. 8. Predicted number of cylinders with minimum point thickness below 250.0 mils.

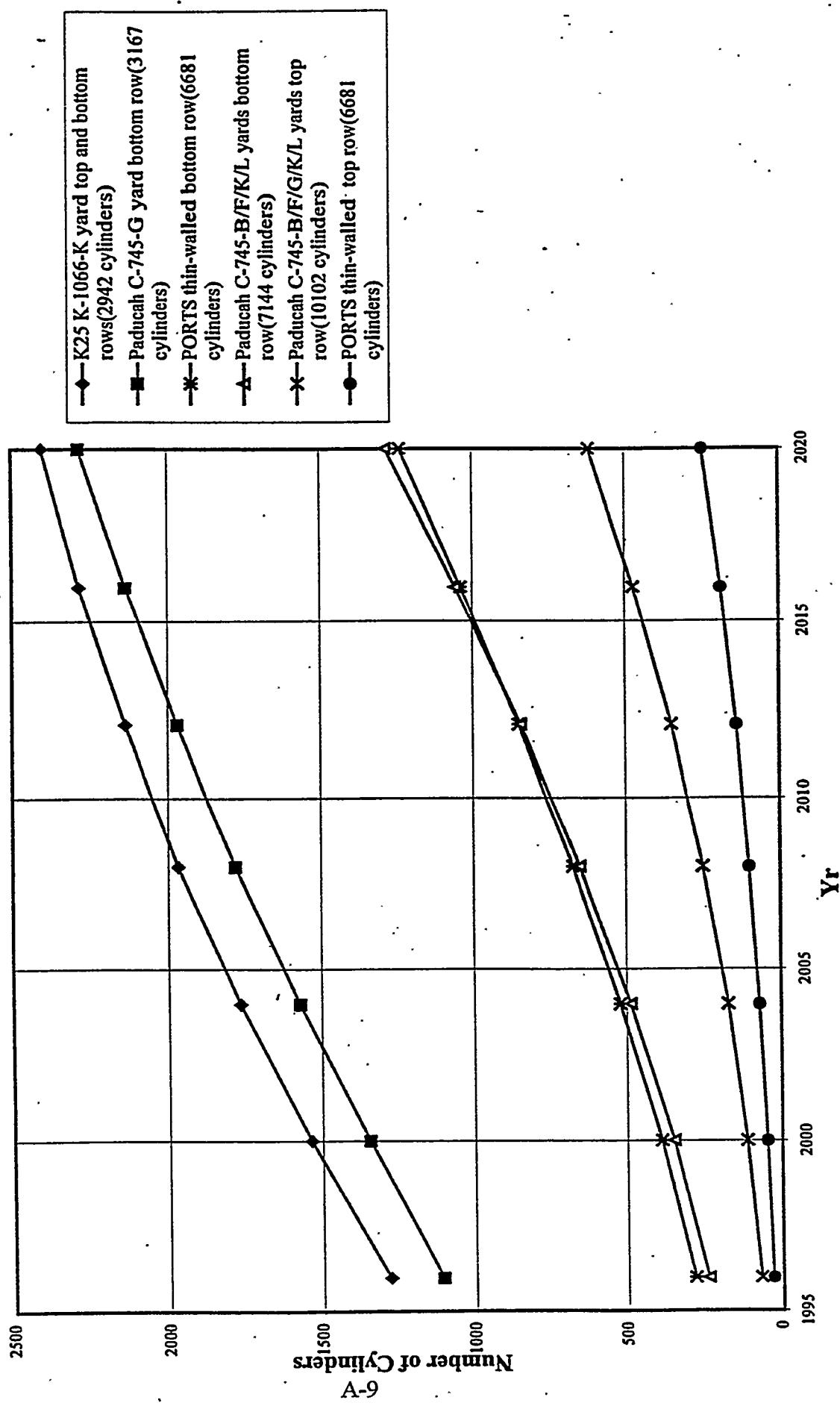


Fig. 9. Predicted number of cylinders with minimum point thickness below 62.5 mils.

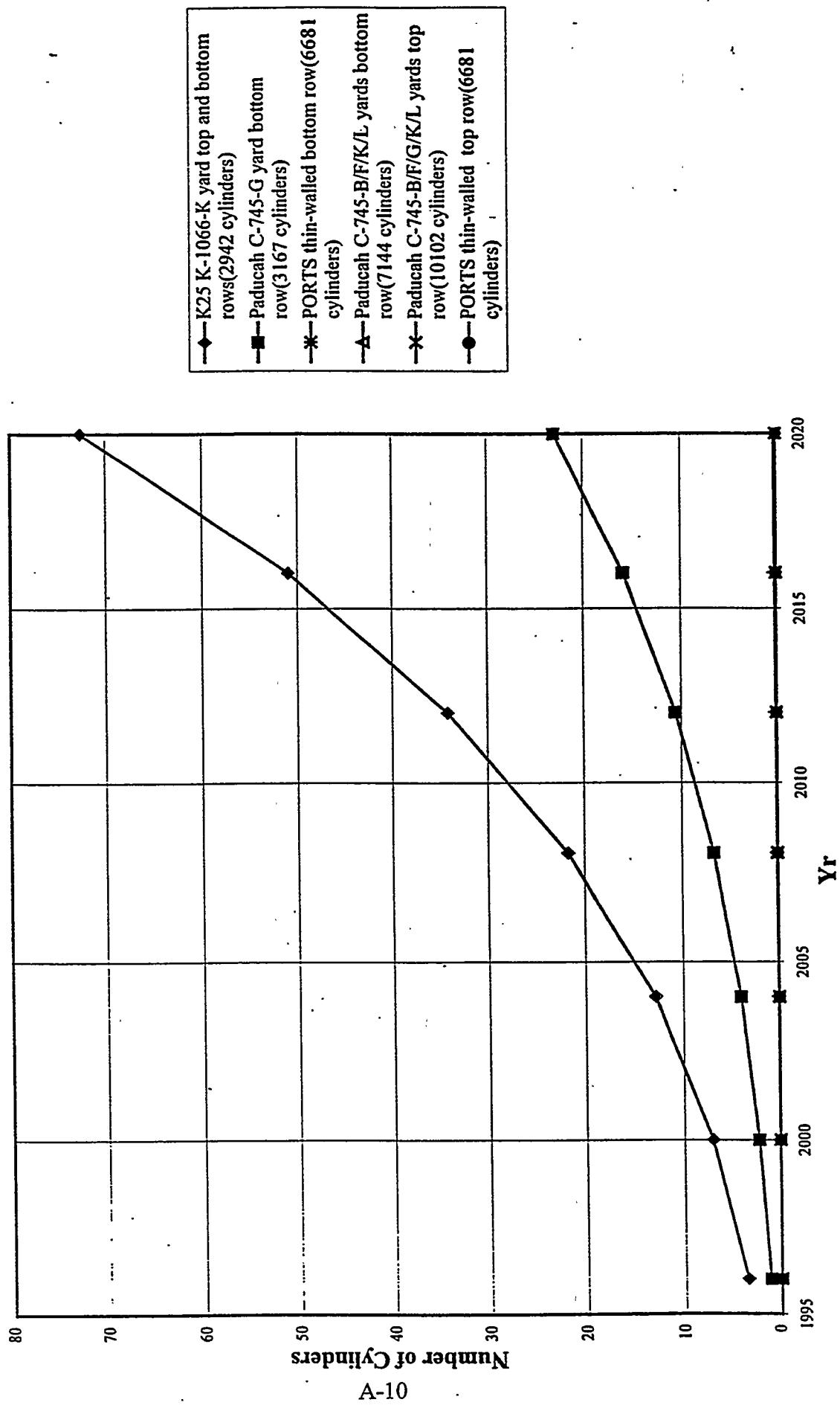
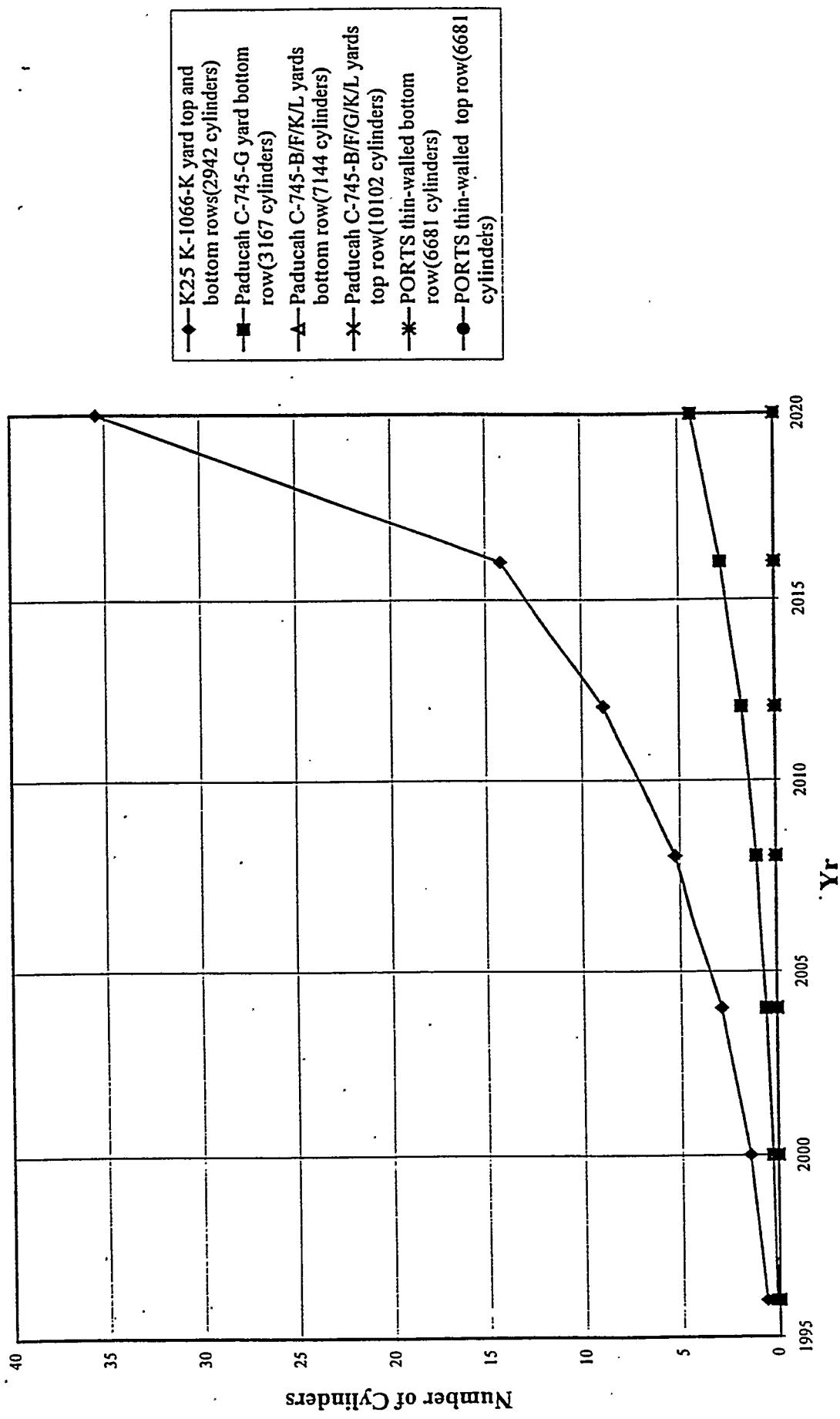


Fig. 10. Predicted number of cylinders with point breaches.



APPENDIX B: METHODS

Calculating the cumulative distribution function for the difference of two distributions

All of the methods discussed in this report are of the form

$$M(t) = C_0 - P(t) \quad (1)$$

where $P(t)$ is the amount of corrosion that results in the minimum wall thickness (mils) at time t , and C_0 is the initial thickness (mils) where the minimum wall thickness occurs. Both $P(t)$ and C_0 are distributions, and calculation of the number of cylinders that have a minimum thickness below a certain thickness z then requires calculating the probability that $M(t) < z$. Assuming a model of the form in Eq. 1, this is equivalent to calculating the probability that $C_0 - P(t) < z$. Since both C_0 and $P(t)$ are distributions, calculation of this probability is not as straightforward as calculating the probabilities of C_0 and $P(t)$ separately, except for certain special cases (e.g., when $P(t)$ and C_0 are both normal distributions, in which case the difference is also a normal distribution). In this section, the method of calculating the needed probabilities are provided.

General Formula

If the random variable W is defined by $W = X - Y$, where X and Y are independent random variables, then any sample w from W can be written in the form (not necessarily uniquely)

$$w = F^{-1}(p) - G^{-1}(q) \quad (2)$$

where p, q are in $[0,1]$, and F^{-1} and G^{-1} are the inverse cumulative distribution functions for X and Y , respectively. Determination of the probability that $W < z$ is then equivalent to evaluating

$$\iint_{A(z)} dp dq \quad (3)$$

where $A(z)$ is the set defined by $A(z) = \{(p,q) \mid F^{-1}(p) - G^{-1}(q) < z\}$. Since F^{-1} and G^{-1} are inverse cumulative distribution functions, they are both nondecreasing functions, and so the function $h(p,q) = F^{-1}(p) - G^{-1}(q)$ is a nondecreasing function of p and nonincreasing function of q . This makes evaluation of the integral in Eq. 3 relatively straightforward. First,

$$\begin{aligned} A(z) &= \{(p,q) \mid F^{-1}(p) - G^{-1}(q) < z\} \\ &= \{(p,q) \mid F^{-1}(p) < z + G^{-1}(q)\} \\ &= \{(p,q) \mid p < F(z + G^{-1}(q))\} \end{aligned} \quad (4)$$

and so

$$\begin{aligned}
 \iint_{A(z)} dp dq &= \int_0^1 \int_0^{F(z+G^{-1}(q))} dp dq \\
 &= \int_0^1 F(z+G^{-1}(q)) dq
 \end{aligned} \tag{5}$$

Therefore,

$$\text{Prob}\{W=X-Y<z\} = \int_0^1 F(z+G^{-1}(q)) dq \tag{6}$$

The integral over the interval $[0,1]$ is evaluated using the adaptive quadrature method described in Burden and Faires (1989). With this method, subintervals are determined so that the integral is approximated with the desired accuracy using Simpson's rule on each subinterval. This method is generally faster than simpler integration methods to achieve the same accuracy because the ultimate subdivision that is used need not be uniformly spaced over the entire interval of integration; the subintervals can be selected based on the desired accuracy and the variability of the function to be integrated.

Application

In this report, F is the cumulative distribution function (cdf) for the initial thickness C_0 which has a truncated normal distribution, and G is the cdf for the penetration depth $P(t)$ at a fixed time t which has a lognormal distribution with mean of the logarithm of the values of $\mu_L(t)$ and standard deviation of the logarithm of the values of σ_L .

Let $N(u)$ denote the cdf for the standard normal distribution (this is the normal distribution with mean 0 and standard deviation 1), and denote the q th quantile of the standard normal distribution by n_q . Then by the formula above it follows that

$$\text{Prob}\{C_0-P(t)<z\} = \int_0^1 F(z+e^{\mu(t)+n_q\sigma(t)}; \mu, \sigma) dq \tag{7}$$

where

$$F_{[a,b]}(x; \mu, \sigma) = \begin{cases} 0 & \text{if } x < a \\ (N(x; \mu, \sigma) - N(a; \mu, \sigma)) / (N(b; \mu, \sigma) - N(a; \mu, \sigma)) & \text{if } a < x < b \\ 1 & \text{if } x > b \end{cases}$$

where $N(x; m, s) = N((x-m)/s)$, where $N(z)$ is the standard normal distribution.

Calculation of Upper Confidence Limits

In the methods used in this report, the maximum penetration depth $P(t)$ is modeled using a lognormal distribution, with $P(t) \sim \text{Log}(\mu_L, \sigma_L) * t$ (for K-1066-K yard) or $P(t) \sim \text{Log}(\log A + n \log t, \sigma_L)$ and the parameters are fit with the data available. The number of cylinders with a minimum thickness below a certain thickness z by a given time T is calculated by a sum of the form

$$\sum_i \text{Prob}\{C_0 - P(t_i) < z\} \times \{\text{Number of cylinders of age } t_i \text{ at time } T\}$$

where the sum is over all age classes for the cylinder population of interest. An upper confidence limit is calculated for this sum by first determining a confidence level α such that if an upper $100\alpha\%$ confidence limit is used for each term in the sum, then the final sum will be bounded with at least 95% confidence. For the purpose of this effort, the Bonferroni inequality is used to determine α . This is conservative because there is structure in the model that is not exploited. If necessary, further analysis will investigate the implementation of less conservative methods (e.g., extension of Working-Hotelling Bands (Miller 1981)). The specific value of α depends on the number of age classes, and increases with the number of age classes; generally, α is at least 0.99 for the cases here.

Because a distribution is assumed for the initial thickness C_0 , it is somewhat involved to calculate confidence limits for the terms $\text{Prob}\{C_0 - P(t) < z\}$, even for a fixed time t . This is because one must obtain an upper confidence limit on an integral of the form

$$\int_0^1 F(z + G^{-1}(q, t)) dq \quad (8)$$

where $F(z)$ is the cumulative distribution function for the initial thickness and $G(z, t)$ ³ is the cumulative distribution function for the penetration depth at time t (i.e., $G(z, t) = \text{Prob}\{P(t) < z\}$).

In order to obtain an upper confidence limit on the integral above, it is not sufficient to use the confidence limits for $G^{-1}(q, t)$ for a fixed q . In particular, neglecting the uncertainty in the initial thickness distribution, a curve $H(q; q_1, q_2)$ must be found such that

$$\text{Prob}\{G^{-1}(q, t) < H(q; q_1, q_2) \text{ for } q \in (q_1, q_2)\} = \alpha$$

where α is the desired confidence level (e.g., 0.95 for an upper 95% confidence limit). The details of calculating the function $H(q; q_1, q_2)$ when $P(t)$ is either normally or lognormally distributed are provided in Appendix C.

³This is a slight abuse of notation that hopefully will help more than confuse the reader.

It is not possible to obtain simultaneous confidence limits that hold for q in the closed interval $[0,1]$ if $P(t)$ is normally or lognormally distributed. However, for arbitrary q_1 and q_2 , it can be concluded with at least $100\alpha\%$ confidence that

$$\int_0^1 F(z+G^{-1}(q,t)) dq < \int_0^{q_1} F(z+G^{-1}(q,t)) dq + \int_{q_2}^1 F(z+G^{-1}(q,t)) dq + \int_{q_1}^{q_2} F(z+H(q;q_1,q_2)) dq \quad (9)$$

Since $F(z) \leq 1$,

$$\int_0^1 F(z+G^{-1}(q,t)) dq < q_1 + (1-q_2) + \int_{q_1}^{q_2} F(z+H(q;q_1,q_2)) dq \quad (10)$$

If it is assumed that $P(t) \sim \text{Log}(\ln A + n \ln t, \sigma)$, then

$$H(q;q_1,q_2) = \hat{A} t^{\hat{n}} e^{S_L z_q} e^{S_L r_U(t,\alpha,q_1,q_2)}$$

where $\ln \hat{A}$ and \hat{n} are the least squares estimates of $\ln A$ and n , respectively, S_L is an unbiased estimate of σ_L , z_q is the q th percentile of the standard normal distribution, and $r_U(t,\alpha,q_1,q_2)$ is the factor that makes the curve $H(q;q_1,q_2)$ an upper confidence limit valid over the entire interval (q_1, q_2) . In the end, one can conclude with at least $100\alpha\%$ confidence that

$$\text{Prob}\{C_0 - P(t) < z\} < q_1 + (1-q_2) + \int_{q_1}^{q_2} \hat{A} t^{\hat{n}} e^{S_L z_q} e^{S_L r_U(t,\alpha,q_1,q_2)} dq \quad (11)$$

Note that the practical effect is to "increase" the term \hat{A} , and this facilitates calculation of the confidence limits because previously implemented integration routines can be employed, after replacing \hat{A} with $\hat{A} e^{S_L r_U(t,\alpha,q_1,q_2)}$, to calculate the integral.

The limits q_1 and q_2 are completely arbitrary, and can be chosen so as to minimize the upper confidence limit on the integral. The term $q_1 + (1-q_2)$ becomes smaller the closer (q_1, q_2) is to the whole interval $(0,1)$. However, this results in an increase of the term $H(q;q_1,q_2)$ since the interval is wider. At the

present, time constraints prevented a two-dimensional search algorithm for finding q_1 and q_2 . Instead, q_1 is fixed and q_2 is found by finding where the derivative with respect to q_2 is zero.

3. REFERENCES

Burden, R.L. and J.D. Faires (1989). *Numerical Analysis*, Fourth Edition, PWS-Kent Publishing Company, Boston.

Evans, M., N. Hastings, and Brian Peacock (1993). *Statistical Distributions*. Second Edition, John Wiley & Sons, New York.

Miller, R.G., Jr. *Simultaneous Statistical Inference*. Second Edition, Springer-Verlag, New York..

APPENDIX C: SIMULTANEOUS CONFIDENCE LIMITS ON THE PERCENTILES OF A NORMAL DISTRIBUTION

In order to calculate upper confidence limits for the case when both the penetration depth $P(t)$ and the initial thickness C_0 are treated as a distribution, it is necessary to find a curve $h(q)$ such that the percentiles P_q of $P(t)$ satisfy $P_q < h(q)$ for all q in a given interval $[a, b]$ for a specified confidence. In this case, it is assumed that $P(t)$ is lognormally distributed, and so the q th percentile of $P(t)$ is of the form $\exp[\mu_L(t) + \sigma_L z_q]$, where z_q is the q th percentile of the standard normal distribution. Since the exponential function is an increasing function, it is sufficient to find a curve that bounds the term $\mu_L(t) + \sigma_L z_q$ for all q in a given interval $[a, b]$ for a specified confidence. In this appendix, how this curve is determined is described, first for the special case when the mean $\mu_L(t)$ is constant, and then in the case for linear regression when $\mu_L(t)$ is assumed to be a linear function of t . The former case reduces to the problem of determining simultaneous confidence limits for the percentiles of a normal distribution. The latter case follows from the former with only a few modifications based on differences in the relevant sampling distributions.

Simultaneous Confidence Limits for the Percentiles of a Normal Distribution

If μ and σ are the mean and standard deviation of a normal distribution, then the p th percentile of the distribution is $\mu + \sigma z_p$, where z_p is the p th percentile of the standard normal distribution. If m and s are a sample mean and standard deviation from this distribution with a sample size of n , then an upper 100 α % confidence limit for $\mu + \sigma z_p$ is given by $m + s(z_p + e_p(\alpha))$, where $e_p(\alpha)$ is calculated from the percentage points of the noncentral t -distribution (Owen 1968). A lower 100 α % confidence limit for $\mu + \sigma z_p$ is the same as an upper 100(1- α)%; i.e., $m + s(z_p + e_p(1-\alpha))$. These confidence limits are not simultaneous in p ; i.e., one cannot state with 100 α % confidence that these bounds hold for a subinterval $[a, b]$. Here we describe determination of exact confidence limits that hold uniformly over a fixed interval. In particular, we describe how to determine numbers $\varepsilon_U(\alpha, a, b)$ and $\varepsilon_L(\alpha, a, b)$ such that

$$\text{Prob}\{\mu + \sigma z_p \leq m + s(z_p + \varepsilon_U(\alpha, a, b)) \text{ for all } p \in [a, b]\} = \alpha \quad (1)$$

and

$$\text{Prob}\{\mu + \sigma z_p \geq m + s(z_p + \varepsilon_L(\alpha, a, b)) \text{ for all } p \in [a, b]\} = \alpha \quad (2)$$

The numbers $\varepsilon_U(\alpha, a, b)$ and $\varepsilon_L(\alpha, a, b)$ can be determined using the distributions

$$T_U = \max_{p \in [a, b]} \left[\frac{(\mu + \sigma z_p) - (m + s z_p)}{s} \right] \quad (3)$$

and

$$T_L = \min_{p \in [a,b]} \left[\frac{(\mu + \sigma z_p) - (m + s z_p)}{s} \right] \quad (4)$$

respectively, because

$$\begin{aligned} \text{Prob}\{\mu + \sigma z_p \leq m + s(z_p + \varepsilon_U(\alpha, a, b)) \text{ for all } p \in [a, b]\} &= \text{Prob}\left\{ \max_{p \in [a,b]} \left[\frac{(\mu + \sigma z_p) - (m + s z_p)}{s} \right] \leq \varepsilon_U(\alpha, a, b) \right\} \\ &= \text{Prob}\{T_U \leq \varepsilon_U(\alpha, a, b)\} \end{aligned} \quad (5)$$

$$\begin{aligned} \text{Prob}\{\mu + \sigma z_p \geq m + s(z_p + \varepsilon_L(\alpha, a, b)) \text{ for all } p \in [a, b]\} &= \text{Prob}\left\{ \min_{p \in [a,b]} \left[\frac{(\mu + \sigma z_p) - (m + s z_p)}{s} \right] \geq \varepsilon_L(\alpha, a, b) \right\} \\ &= \text{Prob}\{T_L \geq \varepsilon_L(\alpha, a, b)\} \end{aligned} \quad (6)$$

It is shown below that

$$\begin{aligned} \text{Prob}\{T_U < t\} &= F_n(t; a, b), \\ \text{Prob}\{T_L < t\} &= F_n(t; b, a) \end{aligned} \quad (7)$$

where

$$F_n(t; a, b) = \int_0^\infty H_n(u; t, a) du + \int_0^{\sqrt{v}} [H_n(u; t, b) - H_n(u; t, a)] du \quad (8)$$

the integrand is defined by

$$H_n(u; t, a) = \frac{1}{\Gamma(v/2) 2^{v/2-1}} G\left(\frac{u}{\sqrt{v}} (t + z_a) \sqrt{n} - z_a \sqrt{n} \right) e^{-u^2/2} u^{v-1} \quad (9)$$

$v=n-1$, and $G(u)$ is the cumulative distribution function for the standard normal distribution, given by

$$G(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-w^2/2} dw \quad (10)$$

Therefore, the numbers $\varepsilon_U(\alpha; a, b)$ and $\varepsilon_L(\alpha; a, b)$ are found by solving the equations

$$\begin{aligned} F_n(\varepsilon_U(\alpha; a, b); a, b) &= \alpha \\ F_n(\varepsilon_L(\alpha; a, b); b, a) &= 1 - \alpha \end{aligned} \quad (11)$$

Note that this implies, at least formally, that

$$\varepsilon_L(\alpha; a, b) = \varepsilon_U(1 - \alpha; b, a) \quad (12)$$

When $a=b=p$ these limits reduce to the usual factors used to calculate confidence limits on the percentile of a normal distribution, and

$$F_n(t; p, p) = \text{Prob}\{t_{n-1}(z_p \sqrt{n}) < (t + z_p) \sqrt{n}\} \quad (13)$$

where $t_v(\delta)$ is the noncentral t -distribution with noncentrality parameter δ . Denoting the upper α th percentile by $t_v^{(\alpha)}(\delta)$, we have that

$$\begin{aligned} \varepsilon_U(\alpha; p, p) &= t_{n-1}^{(\alpha)}(z_p \sqrt{n}) / \sqrt{n} - z_p \\ \varepsilon_L(\alpha; p, p) &= t_{n-1}^{(1-\alpha)}(z_p \sqrt{n}) / \sqrt{n} - z_p \end{aligned} \quad (14)$$

in which case the upper and lower confidence 100 α % limits on $\mu + \sigma z_p$ reduce to $m + s t_{n-1}^{(\alpha)}(z_p \sqrt{n}) / \sqrt{n}$ and $m + s t_{n-1}^{(1-\alpha)}(z_p \sqrt{n}) / \sqrt{n}$, respectively.

Finally, using the elementary fact that $G(-w) = 1 - G(w)$, we have that $F_n(t; a, b) = 1 - F_n(-t; 1 - a, 1 - b)$, and hence

$$\varepsilon_L(\alpha; a, 1 - \alpha) = -\varepsilon_U(\alpha; a, 1 - \alpha) \quad (15)$$

that is, the upper and lower confidence limits are symmetric about the curve $m + s z_p$ when $b = 1 - a$.

Numerical Methods

The first integral in Eq. 8 is that for the cumulative distribution of the noncentral t -distribution. In particular,

$$\int_0^{\infty} H_n(u; t, a) du = \text{Prob}\{t_{\sqrt{n}}(z_a\sqrt{n}) < t + z_a\sqrt{n}\} \quad (16)$$

where $t_{\sqrt{n}}(\delta)$ is the noncentral t -distribution with noncentrality parameter δ . This integral is evaluated using the method discussed in Owen (1968).

The second integral in Eq. 8 is evaluated using adaptive quadrature. However, this integral is first split into several pieces and adaptive quadrature is applied to each piece. This is done because the integrand has a maximum value near the upper limit of integration and is close to zero over most of the integration range. As a result, the adaptive quadrature integration routine will terminate prematurely unless the endpoints of the first few subintervals are close to where the maximum occurs. For this reason, the integral is broken up into integrals over the regions $[0, \sqrt{n}/2]$, $[\sqrt{n}/2, \sqrt{n}-1]$, and $[\sqrt{n}-1, \sqrt{n}]$.

A combination of the bisection method and the secant method is used to solve the equation

$$F_n(\varepsilon, a, b) = \alpha \quad (17)$$

Two initial guesses are required for the secant method. Since $z_b > z_a$, the inequality $H_n(u; t, b) - H_n(u; t, a) \leq 0$ holds for $0 \leq u \leq \sqrt{n}$. Using Eq. 8, this means that $\varepsilon_U(\alpha, a, b) > L = \zeta_a/\sqrt{n} - z_a$, where $\text{Prob}\{t_{n-1}(z_a\sqrt{n}) < \zeta_a\} = \alpha$, and so L can serve as a lower bound for $\varepsilon_U(\alpha, a, b)$. An upper bound U is determined by adding 0.2 to the lower limit L until $F_n(U, a, b) > \alpha$. In this manner, $L < \varepsilon_U(\alpha, a, b) < U$. Similarly, $\varepsilon_L(\alpha, a, b) < \zeta_{1-\alpha}/\sqrt{n} - z_a$, and a lower bound is determined by subtracting 0.2 from the upper bound until an interval is found that contains the root $\varepsilon_L(\alpha, a, b)$. After an interval is found containing the root, the secant method is then used to generate the next approximation. If the approximation falls outside of the interval containing the root, then the bisection method is used to generate the next approximation. This procedure is continued until an approximation $\hat{\varepsilon}$ satisfies

$$|F_n(\hat{\varepsilon}, a, b) - \alpha| < \text{Tolerance} \quad (18)$$

for the upper confidence limit, and

$$|F_n(\hat{\varepsilon}, b, a) - (1 - \alpha)| < \text{Tolerance} \quad (19)$$

for the lower confidence limit.

Derivation of Formula for Relevant Distribution

In this section we derive the integral representation of the cumulative distribution for the functions T_U and T_L as shown in Eq. 8 above.

The function defined by

$$\frac{(\mu + \sigma z_p) - (m + s z_p)}{s} \quad (20)$$

has its maximum and minimum on $[a, b]$ at one of the endpoints since its derivative with respect to p never vanishes (unless $s=\sigma$, in which case it is a constant, an event with probability 0). Further, this function is an increasing function of p if $\sigma > s$ and decreasing if $\sigma < s$ because z_p is an increasing function of p . Thus, the maximum will occur at the left endpoint if σ is overestimated by s , and *vice versa*. This means that

$$\begin{aligned} \max_{p \in [a, b]} \left[\frac{(\mu + \sigma z_p) - (m + s z_p)}{s} \right] &= \frac{\sigma}{s} \left(\frac{\mu - m}{\sigma} \right) + \left(\frac{\sigma}{s} - 1 \right) \begin{cases} z_b & \text{if } \sigma/s > 1 \\ z_a & \text{if } \sigma/s < 1 \end{cases} \\ \min_{p \in [a, b]} \left[\frac{(\mu + \sigma z_p) - (m + s z_p)}{s} \right] &= \frac{\sigma}{s} \left(\frac{\mu - m}{\sigma} \right) + \left(\frac{\sigma}{s} - 1 \right) \begin{cases} z_a & \text{if } \sigma/s > 1 \\ z_b & \text{if } \sigma/s < 1 \end{cases} \end{aligned} \quad (21)$$

This shows that once we have found the distribution for the maximum, the distribution for the minimum can be obtained by simply interchanging a and b . Subsequent derivations are only shown for the maximum.

The random variable $W = (\mu - m)/\sigma$ is normally distributed with mean 0 and standard deviation $1/\sqrt{n}$, and the random variable $X = s/\sigma$ is distributed as $1/\sqrt{\chi_v^2/v}$, and the distributions W and X are independent. Let $G(w)$ denote the cumulative distribution function for the standard normal distribution ($= \sqrt{n} W$) and $f(x)$ denote the probability density function for X . Then

$$\begin{aligned} G(w) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-w^2/2} dw \\ f(x) &= \frac{2v}{x^3} \frac{1}{\Gamma(v/2) 2^{v/2}} (v/x^2)^{v/2-1} e^{-v/(2x^2)} \end{aligned}$$

Therefore we have that

$$\begin{aligned}
& \text{Prob} \left\{ \frac{\sigma}{s} \left(\frac{\mu-m}{\sigma} \right) + \left(\frac{\sigma}{s} - 1 \right) \begin{cases} z_b & \text{if } \sigma/s > 1 \\ z_a & \text{if } \sigma/s \leq 1 \end{cases} < t \right\} = \text{Prob} \left\{ XW + (X-1) \begin{cases} z_b & \text{if } X > 1 \\ z_a & \text{if } X \leq 1 \end{cases} < t \right\} \\
& = \text{Prob} \left\{ XW\sqrt{n} + (X-1) \begin{cases} z_b\sqrt{n} & \text{if } X > 1 \\ z_a\sqrt{n} & \text{if } X \leq 1 \end{cases} < t\sqrt{n} \right\} \\
& = \int_0^1 f(x) G \left(\frac{t\sqrt{n} - (x-1)z_a\sqrt{n}}{x} \right) dx + \int_1^\infty f(x) G \left(\frac{t\sqrt{n} - (x-1)z_b\sqrt{n}}{x} \right) dx \quad (23) \\
& = \int_0^\infty f(x) G \left(\frac{t\sqrt{n} - (x-1)z_a\sqrt{n}}{x} \right) dx + \\
& \quad \left[\int_1^\infty f(x) \left[G \left(\frac{t\sqrt{n} - (x-1)z_b\sqrt{n}}{x} \right) - G \left(\frac{t\sqrt{n} - (x-1)z_a\sqrt{n}}{x} \right) \right] dx \right]
\end{aligned}$$

Making the substitution $u = \sqrt{v}/x$ yields

$$\text{Prob} \left\{ \frac{\sigma}{s} \left(\frac{\mu-m}{\sigma} \right) + \left(\frac{\sigma}{s} - 1 \right) \begin{cases} z_b & \text{if } \sigma/s > 1 \\ z_a & \text{if } \sigma/s \leq 1 \end{cases} < t \right\} = \int_0^\infty H_n(u; t, a) du + \int_0^{\sqrt{v}} [H_n(u; t, b) - H_n(u; t, a)] du \quad (24)$$

where

$$H_n(u; t, a) = \frac{1}{\Gamma(v/2) 2^{v/2-1}} G \left(\frac{u}{\sqrt{v}} (t + z_a) \sqrt{n} - z_a \sqrt{n} \right) e^{-u^2/2} u^{v-1}$$

Application to Linear Regression

Assume that we have pairs of samples p_i at various ages t_i : (t_i, p_i) , $I=1, \dots, N$, and we assume that $p(t) = N(at+b, \sigma)$. Estimates of the regression coefficients a and b are given by

$$\hat{a} = \frac{\sum_{i=1}^N p_i t_i - \frac{1}{N} \left(\sum_{i=1}^N p_i \right) \left(\sum_{i=1}^N t_i \right)}{\sum_{i=1}^N t_i^2 - \frac{1}{N} \left(\sum_{i=1}^N t_i \right)^2}$$

$$\hat{b} = \bar{p} - \hat{a} \bar{t}$$

where $\bar{p} = \frac{1}{N} \sum_{i=1}^N p_i$ and $\bar{t} = \frac{1}{N} \sum_{i=1}^N t_i$. An unbiased estimate of the standard deviation σ is

$$S^2 = \frac{1}{N-2} \sum_{i=1}^N (p_i - \hat{a}t_i - \hat{b})^2$$

Assuming that $p(t) \sim N(at+b, \sigma)$, the following is known about the sampling distributions for the estimates of the parameters (Casella and Berger, 1990; pp. 569-575):

$$\hat{a} + t \hat{b} \sim N(a + bt, \sigma e_N(t))$$

$$S^2 \sim \sigma^2 \frac{\chi_{N-2}^2}{N-2}$$

where we define $e_N(t)$ by

$$e_N(t) = \sqrt{\frac{1}{N} + \frac{(t - \bar{t})^2}{S_u}}$$

and

$$S_u = \sum_{i=1}^N (t_i - \bar{t})^2$$

and χ^2_v is the χ -squared distribution with v degrees of freedom.

Based on similarities in the sampling distributions, the preceding discussion can be applied with only slight modifications to derive simultaneous confidence limits for the p th percentile for $N(at+b, \sigma)$ valid uniformly for p in the range (q_1, q_2) . In particular, all that changes is that we replace $1/\sqrt{n}$ with $e_n(t)$ and $v=n-2$ instead of $v=n-1$. The result is that an upper and lower 100 $\alpha\%$ confidence limit on the p th percentile for $N(at+b, \sigma)$ are given by

$$\{Upper 100\alpha\% confidence limit on at+b+\sigma z_p, p \in (q_1, q_2)\} = \hat{a}t + \hat{b} + S(z_p + r_U(\alpha, q_1, q_2))$$

$$\{Lower 100\alpha\% confidence limit on at+b+\sigma z_p, p \in (q_1, q_2)\} = \hat{a}t + \hat{b} + S(z_p + r_L(\alpha, q_1, q_2))$$

where

$$F_n^{(reg)}(r_U(\alpha, q_1, q_2); q_1, q_2) = \alpha$$

$$F_n^{(reg)}(r_L(\alpha, q_1, q_2); q_2, q_1) = \alpha$$

$$F_n^{(reg)}(\omega; a, b) = \int_0^{\infty} R_n(u; \omega, a) du + \int_0^{\sqrt{n-2}} [R_n(u; \omega, b) - R_n(u; \omega, a)] du \quad (33)$$

and

$$R_n(u; \omega, a) = \frac{1}{\Gamma((n-2)/2) 2^{(n-2)/2-1}} G\left(\frac{u}{\sqrt{n-2}} (\omega + z_a)/e_n(t) - z_a e_n(t) \right) e^{-u^2/2} u^{n-2-1} \quad (34)$$

Finally, if in the case that $\ln p(t) \sim N(a \ln(t) + b, \sigma)$, i.e., $p(t)$ is lognormally distributed, then this same method can be applied; replacing t with $\ln(t)$ and taking the exponential of the upper confidence limits.

REFERENCES

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.

Owen, D.B. (1968). A survey of properties and applications of the noncentral t -distribution. *Technometrics*, Vol. 10, No.3.