ANALYSIS OF INFRASTRUCTURE CORROSION USING REAL PROPERTY PORTFOLIO CONDITION ASSESSMENT DATA

A Thesis

by

SVEN MARKUS ELLEFSON

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Patrick Suermann
Committee Members,	Sarel Lavy
	John Walewski
Head of Department,	Patrick Suermann

December 2020

Major Subject: Construction Management

ABSTRACT

Facility Condition Assessments are a tool for Facility Managers to assess the maintenance needs of facilities and installed equipment and plan sustainment, restoration, and modernization activities. As the use of these tools and the data contained within them grows, data analysis can be used to reveal new information for practical purposes and a better understanding of a real property inventory. Understanding how the environment affects buildings is also critical to making informed maintenance and repair decisions, especially in the context of climate change impacts. This research examines the suitability and implication of using existing corrosion risk assessment techniques for understanding the relationship between corrosion and heating, ventilation, and air conditioning equipment at the strategic infrastructure portfolio level.

I dedicate this to my wife, Jenna, and my daughter, Julia. Thank you so much for your love, support and understanding throughout this journey.

ACKNOWLEDGEMENTS

I would like to begin by thanking Dr. Patrick Suermann, my adviser and the Head of the Department of Construction Science, for all his support, advice, mentorship, advocacy, and leadership from the moment I arrived at Texas A&M. There is a delicate balance in enabling a student to figure things out on their own and in providing a vector on where to go next when help is needed. Time is precious and I appreciate his availability for me on both my research and other topics while running the nation's largest construction science department.

I am thankful to Dr. Sarel Lavy and Dr. John Walewski for agreeing to be part of my graduate committee and provide their perspectives. Diversity of thought is important in life and critical in academic research to avoid stove piping and blind spots.

Mr. John Garcia, Colonel, USAF (Ret.) and Mr. Steve Desrosiers of Alpha Facilities Solutions, LLC provided me with invaluable advice and support. They were a force multiplier on the completion of this research and helped accelerate my work.

Mr. Mark Vandeveer, Colonel, USAF (Ret.) and his staff at AFIMSC gave me the foundation and support to build my research and take their work with Alpha Facilities in a new direction.

Dr. Willa Chen, Professor of Statistics, gave me valuable insights on how to implement the tools from her Statistics in Research course for my data.

Finally, I want to thank Ms. Liz Smith, Dr. Phil Lew, and Dr. Changbum Ahn from the Graduate Office of the Department of Construction Science. Making the transition into graduate school after a long break from academia is never easy and you helped me greatly in smoothing the transition, learning the requirements, and meeting the deadlines necessary to graduate on time.

CONTRIBUTIONS AND FUNDING SOURCES

This work is supervised by a thesis committee chaired by Dr. Patrick Suermann and joined by Dr. Sarel Lavy of the Department of Construction Science (COSC) and Dr. John Walewski of the Zachry Department of Civil and Environmental Engineering.

The data was provided by the Air Force Installation Mission Support Center in partnership with Texas A&M University under their Cooperative Research and Development Agreement. Additionally, the COSC Department worked in partnership with the COSC Construction Industry Advisory Council (CIAC) member ALPHA Facilities Solutions, LLC who conducted analysis of the data for a separate project and provided the results to the author as noted in the paper.

All additional work conducted on this thesis was completed by the student independently. No outside funding was received for the research and writing of this document.

NOMENCLATURE

AFIMSC	Air Force Installation Mission Support Center
BCI	Building Condition Index
BIM	Building Information Modeling
CERL	Construction Engineering Research Laboratory
CI	Condition Index
CRV	Component Replacement Value
DoD	Department of Defense
FCA	Facility Condition Assessment
FCI	Facility Condition Index
FM	Facility Management
FSRM	Facility Sustainment, Restoration, and Modernization
HVAC	Heating, Ventilation, and Air Conditioning
ICCET	ISO Corrosivity Category Estimation Tool
IHA	Installation Health Assessment
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardization
KPI	Key Performance Indicator
PRV	Plant Replacement Value
O&M	Operations and Maintenance
SMS	Sustainment Management System
USACE	United States Army Corps of Engineers
USAF	United States Air Force

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGEMENTS iv

CONTRIBUTORS AND FUNDING SOURCES v

NOMENCLATURE vi

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES xi

LIST OF EQUATIONS xii

CHAPTER I INTRODUCTION 1

Background 1
Problem Statement 2
Research Questions 3

CHAPTER II LITERATURE REVIEW 4

Overview 4
Facilities Management 4
Facility Condition Assessments 5
Climate Change 13
Environmental Factors Affecting Buildings 14
Corrosion 17

CHAPTER III METHODOLOGY 22

Overview 22
Assumptions and Limitations 23
Data Collection 24
Data Cleaning 29
Methods of Analysis 47

Page

CHAPTER IV RESULTS 49

Condition Index 49
Remaining Service Life 54
Age 56

CHAPTER V DISCUSSION 61

CHAPTER VI CONCLUSION 65

Summary 65
Contribution 66
Recommendations 66

REFERENCES 69

APPENDIX 76

LIST OF FIGURES

FIGURE Page

FIGURE		Page
1	Literature Review Organization	4
2	BUILDER Implementation of UNIFORMAT II Hierarchy	7
3	Example Condition Model for a Component-section	12
4	Runway Damage, Reprinted (Raughton, 2018)	16
5	Corrosion Category and Minimum Shore Distance by Number of Locations	29
6	Equipment Details Quantity Type “Each” Histogram and Boxplot	32
7	Sum of *Age* and *InstallDate* Histogram and Boxplot	34
8	Age Histogram and Boxplot	36
9	Install Date Histogram and Boxplot	37
10	CRV Histogram and Boxplot	38
11	CRV Probability Plot	38
12	RSL Histogram and Boxplot	39
13	CI Histogram, Boxplot, and Probability Plot	41
14	Transformed CI Histogram, Boxplot, and Probability Plot	43
15	RSL Histogram and Boxplot	44
16	Age Histogram, Boxplot, and Probability Plot	45
17	Transformed Age Histogram, Boxplot, and Probability Plot	46
18	Correlation of Sample Size and Variance for CI Corrosivity Categories	50
19	CI Boxplots	50
20	CI ANOVA Results	51

FIGURE Page

21 CI Tukey HSD Results........ 51

22 CI log base 10 Boxplots........ 53

23 CI log base 10 ANOVA Results........ 53

24 Log base 10 CI Tukey HSD Results........ 54

25 Correlation of RSL and CI Raw Data........ 55

26 Correlation of RSL and CI Transformed Data........ 55

27 Correlation of Sample Size and Variance for *CI* Corrosivity Categories........ 56

28 Age Boxplots........ 57

29 Age ANOVA Results........ 57

30 Age Tukey HSD Results........ 58

31 Log base 10 Age Boxplots........ 59

32 Log base 10 Age ANOVA Results........ 60

33 Log base 10 Age Tukey HSD Results........ 60

LIST OF TABLES

TABLE Page

1 USAF Minimum Assessment Systems 9

2 Definitions for Direct Rating, Adapted (Uzarski et al., 2018, p. 49) 10

3 Comparison of ESI, ICCET, and ISO 9223 23

4 Data Summary 25

5 Data Summary by Service/MAJCOM 26

6 Frequency of Occurrence for Top 5 D305006 Component Types 27

7 Data Cleaning Summary, Adapted (Dept. of Construction Science, Texas A&M University, & ALPHA Facilities Solutions, LLC, 2017) 30

8 Chosen Fields for Data Cleaning 31

9 Phase 1 Data Removal for 'Component-Section Details' 34

10 Expected Service Life Quantities 40

11 CI Transformation Comparison 42

12 Age Transformation Comparison 47

13 Comparison of CI and Age Ranking 62

14 Component-Section Details, Adapted (Desrosiers, 2020) 76

15 Equipment Details 77

LIST OF EQUATIONS

EQUATION Page

1 Condition Index with Beta Shift 10

2 Condition Index with Beta Factor and Expected Service Life 11

3 Condition Index where Time = Expected Service Life 11

4 Null Hypothesis 22

5 Alternate Hypothesis 22

6 Square Root Transform of CI 42

7 Natural Log Transform of CI 42

8 Log Base 10 Transform of CI 42

9 Square Root Transform of Age 47

10 Natural Log Transform of Age 47

11 Log Base 10 Transform of Age 47

CHAPTER I

INTRODUCTION

Background

The USAF has developed a new real property investment decision making model known as the Infrastructure Health Assessment (IHA) tool to guide strategic budgeting decisions for its $263 billion real property portfolio (Toliver, 2019). The IHA tool makes use of the Sustainment Management System (SMS) BUILDER tool, a Computerized Maintenance Management System (CMMS) used by the Department of Defense (DoD). These tools involve gathering facility condition information through local facility condition assessments. The information is loaded into the BUILDER database which is used to populate the data used by the IHA tool. This information is then used to forecast infrastructure condition on long time scales (e.g. 10, 30, 50 years) to show the impact of spending on facility sustainment, restoration, and modernization (FSRM) (Toliver, 2019). A known shortfall of this tool is its lack of accountability for degradation due to climate factors. Colonel Shamekia Toliver's Air War College thesis states: "address this by adjusting the degradation curve calculations used in the IHA tool to account for environmental variations associated with climate changes; thus, affecting the lifecycle of infrastructure." (Toliver, 2019, p. 19)

The Intergovernmental Panel on Climate Change (IPCC) *Fifth Assessment Report* on *Climate Change 2013: The Physical Science Basis* (IPCC, 2013) concluded from the body of knowledge that "human activity is the dominant cause" of climate change with 95% certainty (p. v) and identified changes that have already occurred. On average, the global temperature across land and sea has increased 0.85 °C in the period from 1880 to 2012. Alexander et al. (2006) found that more than 70% of 2,291 global grids points sampled experienced "statistically significant" warmer nights in the period of 1951 to 2003 with some parts of the earth experiencing twice the number of warm nights than established in the study's index. Precipitation has generally increased, but further extremes can be identified by accounting for regional variation. Parts of the U.S. and South America experienced an additional "up to 2 days per decade" each year of "heavy precipitation" (Alexander et al., 2006, p. 12).

The National Centers for Environmental Information (2020) calculates the cost of 265 meteorological disasters (each exceeding $1 billion) at a total of $1.78 trillion dollars in the U.S. since 1980. They cite both climate change and increases in material wealth of society as reasons for the rising cost. In addition to natural disasters, weather plays a role in corrosive effects. Callister & Rethwisch (2010) note that estimates of the cost of corrosion are approximately 5% of national income to maintain or repair materials damaged by corrosion. The literature shows that buildings and their environment are inextricably linked. Buildings contribute to 39% of global CO_2 emissions and consume 36% of global energy (Abergel et al. 2017).

While there are multiple tools in literature for understanding the severity of corrosion in an environment, Silver & Gaebel (2017) found that none exist that adequately link infrastructure, to environment, to cost. A first step in better understanding this problem is to understand how the condition of infrastructure correlates with the corrosivity of its environment.

Problem Statement

Building CMMSs such as BUILDER are tools for understanding and predicting maintenance costs. However, after reviewing the literature, a method for linking environmental data to maintenance costs in a way that decision makers can use for planning FSRM does not exist. In order to make the dataset more manageable, and at the recommendation of AFIMSC, this research looked specifically at the case of atmospheric corrosion in air conditioning (A/C) equipment. Dillon & Herro (1997, p. 2) write that “corrosion, deposition and material defects account for the vast majority of failures in HVAC” (heating, ventilation, and air conditioning). Bhatia (2020, p. 2) states that “corrosion alone accounts for approximately 40% of all equipment failures in industrial facilities.” HVAC is used for the primary purposes of occupant comfort or cooling for equipment within a conditioned interior space (Howell et al., 2013). This makes HVAC important for the USAF which operates in a variety of environments across the globe and needs to provide cooling for human comfort or for equipment to correctly function such as in laboratories or data centers. As an example, HVAC is one of the building systems that requires “a major repair project” to be planned, programmed, and designed for the year after its Condition Index (CI) follows below 70 out of 100 (Allen, 2019). HVAC equipment can be used as starting point for understanding the impacts of corrosion on infrastructure condition and FSRM costs.

Research Question

The research question for this project is:

> RQ. How does HVAC equipment condition correlate by level of environmental corrosivity, as evidenced by BUILDER data on air conditioning equipment?

CHAPTER II

LITERATURE REVIEW

Overview

There are two primary tranches of literature review conducted in this paper. The topics of facilities management and facility condition assessment constitute the first part of the literature review, and the topics of environmental factors affecting buildings, climate change, and corrosion are part two. Figure 1 summarizes the literature review approach.

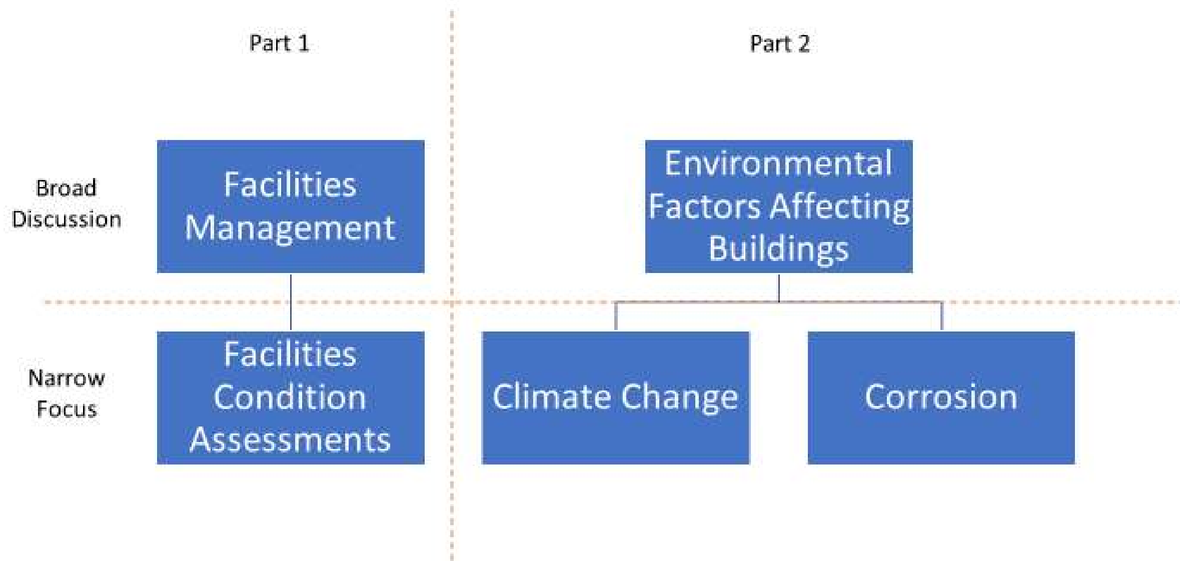


Figure 1. Literature Review Organization

Facilities Management

Roper & Payant (2014, p. 13) use the International Facility Management Association definition to define FM as “a profession that encompasses multiple disciplines to ensure functionality of the built environment by integrating people, place, process, and technology.”

Lavy et al. (2014) note that performance measurement is integral to holistically assessing how well a facility meets its functional purpose. They list tools for performance measurement such as "…benchmarking, balanced scorecard approach, post occupancy evaluation, key performance indicators (KPIs), and critical success factors" (p. 257). Lavy et al. (2010) identified 35 KPIs from literature and surveyed (n = 11) FM practitioners on a suggested categorization of these KPIs. Subsequent research by Lavy et al. (2014a) resulted in the identification of several "core KPIs" including facility condition index (FCI), functional index, maintenance efficiency, etc. They addressed two primary definitions of FCI as either (1) the ratio of the facility's or system's maintenance backlogs to their respective replacement values or (2) the weighted average of facility conditions. The latter of these definitions is also referred to as the building condition index (BCI). Lai & Man (2018) did not include FCI or BCI in their literature-review derived list of 71 performance indicators for facility operations and maintenance (O&M); they focused instead on other indicators such as thermal comfort, visual comfort, etc. to track the physical performance of a facility and relate it to occupant satisfaction.

Bröchner et al. (2018) researched future trends in FM and identified digitalization and sustainability as the most important factors for the future of the profession. Several papers address Big Data and building information modeling (BIM) as part of the digitalization trend. Ahmed et al. (2017) note that BIM has grown into a tool for use throughout the life of a facility and can be linked with "Big data analytics" (pp. 742-743). They clarify the distinction between Big Data and a "Small Data Set" (p. 728) by referencing a table from Japkowicz & Stefanowski (2016). This table distinguishes that Small Data is more homogeneous and can be analyzed through traditional means such as by use of a "relational database" whereas Big Data is very heterogenous, difficult to link, and grows too rapidly for analysis by traditional means. Munir et al. (2019) concluded from their literature review that information does not have an inherent value and that purpose and analysis are needed to make information useful.

Facility Condition Assessments

Overview

Among the tools of FM practitioners is the facility condition assessment (FCA). FCA can be defined as a method for comprehensively surveying all chosen elements of a building to provide data for FSRM funding decision-making (Karanja & Mayo, 2016). Karanja & Mayo (2016) found that several commercial platforms exist for managing assets such as "Archibus, IBM Maximo, IBM and Tririga" (p. 2). Another such system developed by USACE CERL belongs to the SMS family of tools, BUILDER, a "life-cycle asset management software application" (Uzarski et al., 2018, p. ii) and is used by the DoD to assess and report facility condition and update condition information through field inspections. PAVER, ROOFER, and RAILER are also systems under the SMS umbrella and are used for paved airfield and non-airfield surfaces, building roof inventory, and railway systems, respectively (Herrera et al., 2017). The use of BUILDER is mandated in the USAF by DoD policy for USAF Civil Engineers to record facility condition data (Henderson, 2019) and was part of a DoD-wide effort to meet the Department's Financial Improvement and Audit Readiness (FIAR) completion deadline for real property condition assessment by September 2017 (AFCEC, 2018).

Information in BUILDER is organized under the American Society for Testing and Materials (ASTM) Uniformat II standard to categorize building elements (Uzarski et al., 2018). Uniformat II breaks down the elements of a building into four levels (Charette & Marshall, 1999) and matches with the BUILDER level hierarchy used by Uzarski et al. (2018) as shown in the example in Figure 2. Uzarski et al. (2018) refer to the lowest level as the component-section level.

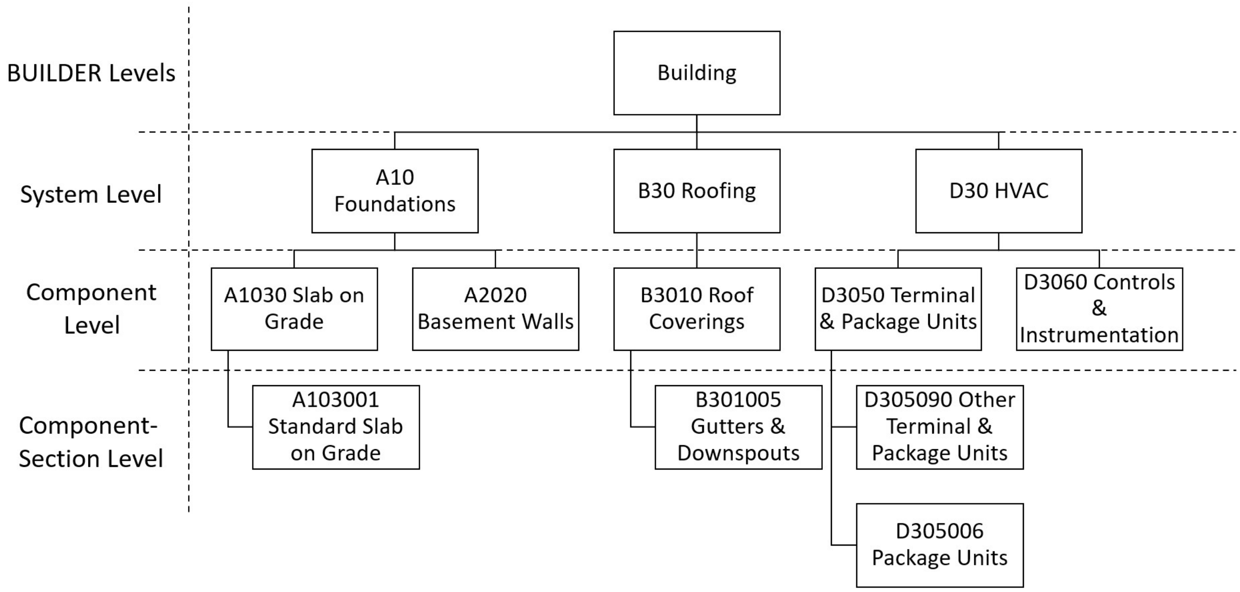


Figure 2. BUILDER Implementation of UNIFORMAT II Hierarchy

USAF Implementation

According to Herrera et al. (2017), the USAF began its pilot program for implementing BUILDER in 2010. The USAF utilized both in-house (DOD military and civilian personnel) and contractors to conduct the inventory, with the long-term plan of maintaining the BUILDER database condition assessments through in-house personnel. The findings of *A Review of the BUILDER Application for Assessing Federal Laboratory Facilities* report (Herrera et al., 2017) on implementing BUILDER in the federal government are summarized as:

- BUILDER is projected to provide cost savings but as of the report no study at been conducted to verify this

- In-house personnel are more knowledgeable about their facilities but may become overburdened by conducting FCAs or view them as having minor importance in comparison to their other duties
- Contracted personnel can provide a consistent assessment without overburdening in-house staff but they are not knowledgeable about an organization's facilities and may require additional escort or security clearance to assess restricted facilities. Contractors can also perform poorly "due to improper inventorying or inadequate system sectioning" (p. 30)
- BUILDER is meant to provide automated scheduling of when to conduct inspections to reduce the workload of unnecessary inspections on facilities of lesser importance, vice conducting inspections on a calendar-basis

The AFCEC Operations Directorate (AFCEC/CO) manages the USAF application of SMS with the FUELER and UTILITIES systems to be added in the future for managing fuel distribution systems and utility systems, respectively. FCAs are conducted on a recurring basis with the goal to assess 20% of an organization's square footage every year in accordance with DoD policy (AFCEC, 2018). The USAF only requires the assessment of 7 out of 13 total systems to count a building as completely assessed (Herrera et al., 2017). These are shown in Table 1; the assessment of C30 Interior Finishes is required for Military Family Housing (MFH) and Dormitories (AFCEC, 2018).

Table 1. USAF Minimum Assessment Systems

B20 Exterior Enclosure
B30 Roofing
C10 Interior Construction
D20 Plumbing
D30 HVAC
D40 Fire Protection
D50 Electrical

Rating Methods

There are two methods for conducting condition assessments of the building, direct rating and distress rating (Uzarski et al., 2018). The USAF utilizes the direct rating approach for assessing a facility (AFCEC, 2018). Direct rating requires "a seasoned inspector" (Bauer, 2020, p. 18) to quickly assess the condition of a component-section and assign it one of nine ratings, shown in Table 2. This allows for assessment to be conducted more quickly, such as when the "inventory that is being brought into BUILDER in bulk for the first time" (p. 18). Direct ratings are conducted visually, so AFCEC (2018) recommends only recommends tools that enhance visual inspections, such as digital cameras, flashlights, infrared thermometers, or thermal cameras as opposed to equipment-specific test tools.

A more thorough survey can be conducted using the distress rating method (Bauer, 2020). This method involves the visual evaluation of the severity and density of deterioration in at the sub-component-section level. Uzarski et al. (2018) account for 23 different types of distresses that can occur; examples include blistering, corrosion, cracks, leaks, excess noise, etc. An inspector must measure the quantity of distresses and estimate the severity based on pre-defined categories for each distress type. These values are then entered into BUILDER where the distress density is computed and a component-section condition index calculated based on the conditions of all the sub-component-sections.

Table 2. Definitions for Direct Rating, Adapted (Uzarski et al., 2018)

Rating	**Description**
Green (+)	No visible or identifiable deterioration
Green	No decrease in reliability or serviceability. Some noncritical parts are negligibly deteriorated, or few critical parts are negligibly deteriorated.
Green (-)	Small or no decrease to reliability or serviceability. Some noncritical parts have minor deterioration or more than one critical part negligibly deteriorated.
Amber (+)	Reduced reliability or serviceability. Few critical parts may have modest deterioration and few noncritical parts have serious degradation.
Amber	Reduced reliability or serviceability. Minor amount of critical parts exhibit partial deterioration with some noncritical parts exhibiting serious deterioration.
Amber (-)	Serious reduction in serviceability or reliability. Majority of parts partially degraded or few critical parts seriously degraded.
Red (+)	Serious reduction in serviceability or reliability. Most parts are significantly degraded or degraded to varying amounts.
Red	Critical reduction in serviceability or reliability. Majority of parts critically deteriorated.
Red (-)	Failed and mostly beyond salvage.

Condition Index

BUILDER makes use of the BCI metric and therefore the Condition Index (CI) of an individual section is the fundamental metric (Uzarski et al., 2018). The CI is a value from 0 to 100 where 0 is completely failure of the section and 100 is brand new. According to Grussing (2012, p. 5), it is difficult to implement a model describing lifespan for every building component. It is easier to amalgamate component-section condition data into higher levels across different building parts and materials if an "objective CI" model is used for consistency. Grussing prescribes the following model calculating CI as a function of time:

$$CI = A \times e^{-(t/\beta)^{\alpha}} \quad (1)$$

Where the parameters can be defined thusly:

- *A* is the parameter for initial steady state component-section index
- *α* is the parameter for accelerated deterioration factor
- *β* is the parameter for service life adjustment
- *t* is time in years

Desrosiers (2020) clarified in an e-mail to the author that CERL calculates CI in BUILDER with the following adjustments:

$$CI = A \times e^{-\left(\frac{t}{\beta \times E}\right)^{\alpha}} \tag{2}$$

Where:

- *β* is the service life adjustment factor
- *E* is the expected remaining service life

Equation (2) was used in this paper to describe the condition model. Colonel Toliver (2019) notes that typical starting values are $\alpha = 2.64$ and $\beta = 1$. Grussing & Marrano (2007) define $A = 100$ for new component-sections or fully replaced component-sections and $A = 95$ for fully repaired component-sections. At a CI of 40, a component is considered to have met its service life (approximately when $t = E$). It should be noted, that in situations where $A = 100$ and $\beta = 1$, a CI of 40 will occur slightly after ($t > E$) because when $t = E$ precisely the CI is approximately 36.788 because the equation reduces to:

$$CI = 100 \times e^{-\left(\frac{t}{1 \times t}\right)^{\alpha}} = 100 \times e^{-1} \tag{3}$$

When a component-section is repaired or replaced, a stepwise jump in condition occurs and a new degradation curve is calculated. When a repair occurs, the parameters are adjusted to speed up the degradation rate. This limits the number of repairs that can be done. Figure 3 provides an example of a condition curve where the input values are $A = 100$, $E = 15$, $\alpha = 2.64$, and $\beta = 1$. In

BUILDER, (Grussing, 2012) subsequent FCAs result in new degradation curves being fitted to the historical and the most recent CI point taken from a distress survey or direct rating. A regression is used and the α and β parameters are fitted via regression. This methodology allows for the condition model to account for feedback from real world conditions.

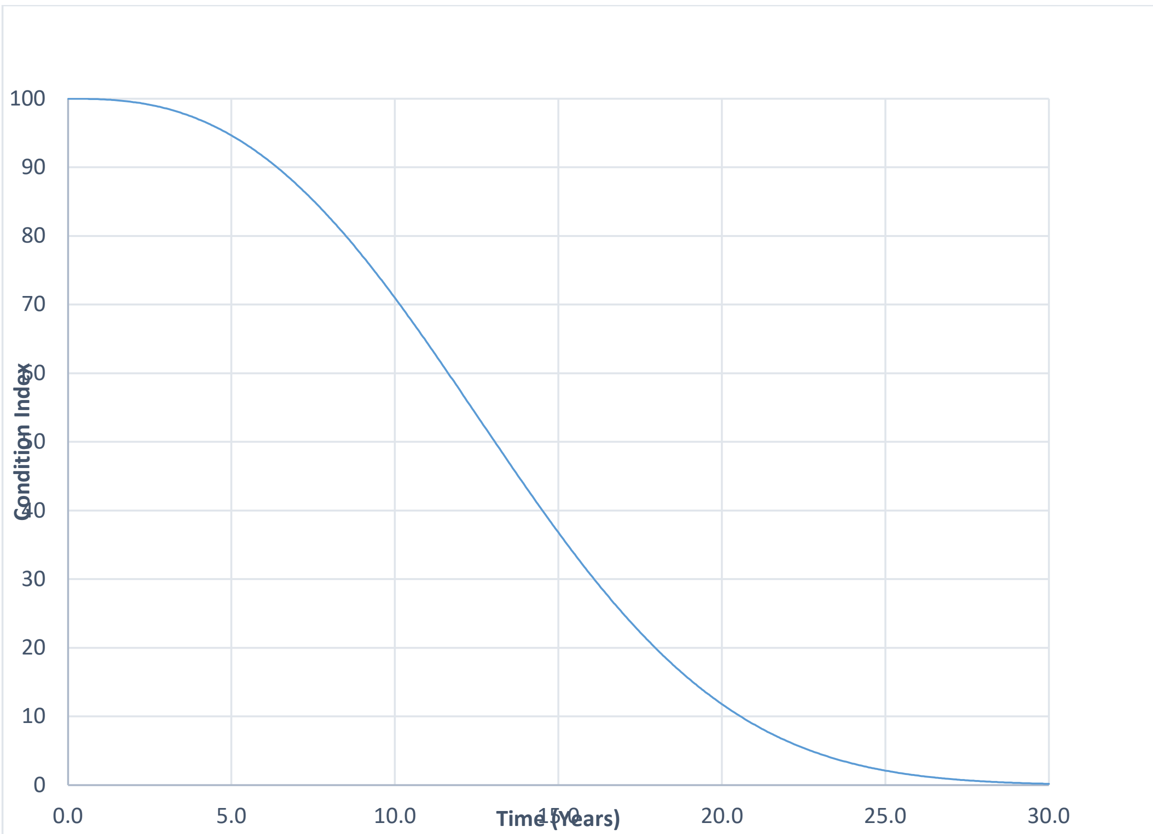


Figure 3. Example Condition Model for a Component-section

New Methods

Alley et al. (2017) proposed a probabilistic method to account for the probable failure rates of building component-sections as opposed to the deterministic method in Equations 1-2. This probabilistic method is not currently implemented in BUILDER but was found by Alley to produce statistically significant results for predicting component-section failures in a case study using actual work order data.

Bartels (2020) used association rule mining to identify frequently occurring component-sections in the USAF BUILDER database and identify condition associations between component-sections. This was done by data mining the inspection report details for condition assessments and developing associations between the component-sections inspected and the level of deterioration at the time of inspection. The result of this research is a methodology that can be used to predict the second-order impacts on the CI of related component-sections when a primary component-section CI is degraded and streamline FCA processes.

Climate Change

The overarching climate determines the cyclical nature of environmental and atmospheric characteristics. Climate change adds a long-term, changing dynamic to all the meteorological conditions we are familiar with for a given location. The 2014 IPCC Climate Change Synthesis Report (IPCC, 2014) states that "warming of the climate system is unequivocal." The 2014 Synthesis Report ties a rise of global temperature and oceanic acidification to emissions of greenhouse gases such as carbon dioxide. The IPCC (2014) projects with "high confidence" that warming is "likely" to exceed 1.5 °C by the late 21st century. IPCC (2012) projects with "high confidence" an increase in frequency, duration, or severity of heat waves (indexed against "late 20th-century extreme values") for North America, Europe (with the exception of Scandinavia), Africa, parts of South America, parts of Asia, and Australia. Heavy precipitation is defined as greater than the 95th percentile of precipitation days. The IPCC (2012) forecasts with "high confidence" that increases to heavy precipitation will occur in Alaska, parts of Canada, Iceland, Greenland, parts of Europe in winter, parts of Africa, and northern Asia.

Environmental Factors Affecting Buildings

There are examples in literature of how the environment affects built infrastructure and the role weather plays in deterioration and damage to buildings. They can roughly be divided into two classes of events: catastrophic short-term events acting up to several days (e.g. a hurricane) and slow-acting long-term events (years to the life-cycle of the building and its materials).

According to Bastidas-Arteaga & Stewart (2019) in *Climate adaptation engineering*, severe wind events (stand alone or in association with other meteorological phenomena) and elevated levels of precipitation pose significant risk of damage to structures. In particular, Khajwal & Noshadravan (2020) cite that hurricanes have had a devastating global impact on life and property at a cost of $938.2 billion in the period from 1980 to 2019. They developed an improved model for determining potential risk to infrastructure in terms of loss to life and property by making some of the parameters random variables. This model allows for the prediction of loss using wind speed and a minimum number of other parameters. Downton et al. (2005) pulled from National Weather Service damage to show that $50 billion in flood damage occurred in the U.S. in the 1990s.

Ginger et al. (2015) analyzed wind loads on industrial building frames and examined wind tunnel data to determine the effectiveness of structural design standards. The authors (Ginger et al., 2015) were able to identify weak points in the design of low hanging, large bay industrial buildings. Francis et al. (2010) examined secondary and tertiary effects of electrical grid loss or damage in the context of more severe weather events occurring under climate change. Their probabilistic extended life-cycle analysis cost allows them to model the broader impact to society of power distribution failure. Stewart & Deng (2015) produced a method for assessing hurricane wind damage of hurricanes and examined the impact of changing lumber fastening design standards as climate adaptation measures for probabilities of higher wind speed scenarios in climate change. Some research also shows that existing design standards can be adequate for meeting future needs. Lee & Ellingwood (2017) performed an extended +100-year analysis of a structural model subject to wind damage from hurricanes and found that buildings designed to last 50 years are capable of meeting current and future wind load resistivity needs beyond their design lifespans.

While temperature and moisture can contribute to meteorological phenomena that generate catastrophic hazards, they can also directly contribute to the long-term deterioration of infrastructure. Richardson (2001, p. 14) defines deterioration as "a natural process which may be unavoidable." He goes on to describe the effects of moisture and temperature on buildings. Temperature contributes to the thermal movement of infrastructure, and material type can play a significant role in determining temperature's overall effect. Differing thermal expansion coefficients of dissimilar materials can result in damage. For example, aluminum has a linear thermal expansion coefficient of 24 x 10-6 per °C while the coarse aggregate in concrete exhibits a thermal expansion coefficient of approximately 12-13 x 10-6 per °C. An example of thermal expansion in paved surfaces was reported by Raughton (2018) in Figure 4. A concrete taxiway at Osan Air Base, South Korea ruptured after three consecutive days of peak 100 °F (Iowa Environmental Mesonet, 2020)[1] weather in Korea and the intrusion of rocks and debris in the pavement slab expansion joints. This resulted in the concrete around the joint buckling due to a lack of room for expansion based anecdotally on local assessments.

[1] Weather data obtained from Meteorological Aerodrome Reports (METAR) collected by Iowa Environmental Mesonet (2020). The original article from Raughton (2018) did not include temperature values.

Figure 4. Taxiway Damage, Reprinted (Raughton, 2018)

Moisture contributes to the deterioration of porous materials such as masonry, concrete, and timber products (Richardson, 2001). Additionally, moisture is a part of numerous mechanical, thermal, biological, and chemical processes that degrade building materials. One such chemical process is corrosion. Nguyen et al. (2013) found that not only do moisture and temperature play a role in the corrosion of steel, but so do airborne pollutants such as sulfur dioxide albeit on a more localized basis to pollution sources. They refer to corrosion as a "complex discontinuous electrochemical process subjected to highly variable ambient conditions", noting that there are apparent paradoxes such as nighttime moisture layers driving corrosion, whereas regular rainfall can help remove corrosive pollutants from building materials. In contrast to these negative effects, Nguyen et al. found small but ultimately negligible benefit of increased atmospheric carbon dioxide to arresting the corrosion of zinc. Carbon dioxide concentrations of 1000 ppm are possible in the A1F1 climate change scenario, and Chung et al.

(2000) found that zinc plates had 20% higher corrosion resistance in an enclosed 1000 ppm CO_2 environment than in 350 ppm CO2 environment. Wang et al. (2011) write that temperature, humidity, and carbon dioxide play integral roles in the carbonation process of corroding concrete. They note that drier regions receive a significant deceleration of concrete carbonation due to lower relative humidity values whereas more humidity regions can experience faster rates of carbonation. Raised temperatures and carbon dioxide levels play an accelerating role in carbonation processes. In addition to carbonation, concrete can corrode due to chloride ingress (Bastidas-Arteaga et al., 2010). Similar to carbonation, moisture and temperature play an essential role in the chloride-ingress of concrete. They modeled the effects of chloride ingress, moisture, and heat flow from weather effects and used Karhunen-Loève expansions to introduce fluctuations for their humidity and temperature models. This enabled the simulation of weather extremes while still adhering to overall mean values. With this stochastic approach, they concluded that elevated temperature and humidity such as can be found in tropical environments reduces the initiation time of chloride ingress-based corrosion.

Corrosion

There are multiple ways to define corrosion. Callister & Rethwisch (2010, pp. 674-675) address the broader definition of corrosion but also provide a more typical definition that is used: "…the destructive and unintentional attack of a metal." Title 10 U.S. Code §2228 defines corrosion as "the deterioration of a material or its properties due to a reaction of the material with its chemical environment." Wang et al. (2011) modeled carbonation-based corrosion of concrete over a 100-year period. This type of corrosion is heavily influenced by atmospheric CO_2 and thus analysis over a 100-year period allows for climate change to be considered. Bastidas-Arteaga et al. (2013) modeled the corrosion of concrete reinforcement due to chloride ingress and carbonation damage of concrete and found that climate change had some impact on corrosion in highly corrosive environments. The results showed that even with climate change CO_2 effects considered, the corrosion in an environment will be dominated by other factors. Nguyen et al. (2013) modeled future corrosion effects of zinc and steel over a 100-year timespan under the A1FI climate change scenario. A key conclusion of their research is that it is difficult to predict secondary and tertiary effects because the data needed to make such predictions is not available. Rakanta et al. (2007) conducted a case study of damage done to a hotel in Greece. The closed-loop, chilled-water systems were corroded and factors that contributed to the corrosion were identified. These included build-up of salt in the chiller exchanger tubes, mixing of system water with freon and oils, and damage of the chiller motor due to contact with water. Water and pipe samples showed build-ups of chlorides, magnesium salts, and calcium salts contributing to corrosion. These latent problems were attributed to a lack of chemical science expertise in the maintenance department (either in-house or through consultation services).

Studies of metal sample plates are commonly used to assess the corrosivity of an environment. Fuente et al. (2007) and Fuente et al. (2011) conducted two studies over 13-year time periods to better understand the corrosion of zinc and mild steel in different environments in Spain. The 2007 study of zinc found that marine and industrial environments had the most dramatic effect on zinc. Additionally, facing direction of the zinc plates was related to corrosion in that downward facing plates exhibited less corrosion than skyward facing plates. Finally, slightly acidic rainwater had positive effects on limiting zinc corrosion. The 2011 study of mild steel found that marine environments had a most dramatic effect on their samples. They were only able to produce dimension, mass loss, and imaging data for 10 years in the marine environment because the samples had disintegrated by that time. Fuente et al. (2011) highlighted the compacting nature of corrosion layers in some environments due a combination of drier air and higher heat.

Grisham (2001) conducted a small-scale study of energy efficiency ratio (EER) and corrosion of metal samples on two HVAC units in Galveston, Texas over the course of one year to research the relationship between HVAC performance and saltwater exposure. Grisham found that the EER for the two units decreased by 11.5% and 8.3%. Aluminum and copper samples in the same atmospheric environment as the HVAC units were subjected to a microscopic analysis to discover pitting depths. The aluminum in the fin colors and in the galvanic couples exhibited pitting at 22% and 19%, respectively.

ISO 9223 (International Organization for Standardization) details standards for Corrosivity Category Classification using atmospheric and pollution data or through measurement of mass loss for standardized metal samples over a one-year period per ISO 9226 (Silver & Gaebel, 2017). This ISO standard is primarily meant to aid in the material selection process during design. It consists of six categories of corrosivity in increasing levels of corrosion: C1, C2, C3, C4, C5, and CX. CX is new in the 2012 variant of the standard. The older 1992 standard featured five categories and was scrutinized by Morales et al. (2005) in a study conducted in the Canary Islands. The Canary Islands have varied microclimates that cover a wide range of Corrosivity Categories. Morales et al. (2005) monitored atmospheric data and conducted metal plate testing and found that ISO 9223:1992 did not accurately predict the Corrosivity Category in numerous locations. Additionally, the 1992 version of the standard only had five corrosivity categories and Morales et al. (2005) found that some locations exceeded the parameters and mass loss rates of category C5. Following up in 2019, Santana et al. analyzed the updated ISO 9223:2012 standard in the Canary Islands, testing 74 different sites for carbon steel, copper, and zinc corrosion. They used the atmospheric and plate mass loss data they gathered to develop their own constants in the dose-response function of ISO 9223:2012 as well as using time of wetness instead of relative humidity to obtain more accurate results. Time of wetness is the time that the environment is greater than 0 °C and 80% relative humidity. The newly added CX Corrosivity Category in the 2012 update of ISO 9223 did not adequately describe corrosion in some locations.

The 2017 *Facilities Environmental Severity Classification Study* from Silver & Gaebel was conducted for the DoD to research the problem of classifying the severity of an environment in the context of DoD real property. Here is a summary of the major relevant findings from literature and other studies in this extensive report:

- A single method of corrosivity classification cannot adequately model corrosion in every environment
- The Battelle Columbus corrosion study of DoD properties showed an unclear correlation between steel mass loss and copper mass loss. Some modelling methods convert a certain metal's mass loss into an equivalent steel mass loss to simplify calculations
- There is a significant gradient of corrosion loss within the first mile of a saltwater shoreline (from maximum at the edge to minimum at one mile inland)
- Corrosion generally exhibits a piecewise behavior with maximum loss rate in the first 20 years and a lower, steady rate thereafter
- Corrosion occurs at the local and micro level. The local level can be defined as something such as a building, whereas the micro level is the actual site of

corrosion occurring on a specific material. The characteristics of these two environments are related but may differ

Additionally, Silver & Gaebel developed the ISO Corrosivity Category Estimation Tool (ICCET) for determining DoD installation Corrosivity Category using atmospheric data. This tool has some limitations. It does not incorporate pollution data due to a lack of availability (such as in ISO 9223:2012). It also uses atmospheric data from fixed weather stations, so if a location is selected without a nearby weather station, the nearest available one is used. ICCET also can only generally determine the Corrosivity Category for a location. Areas with large variations in corrosivity such as those near coasts or pollutant-producing industry may require multiple Corrosivity Categories to classify the area.

The DoD developed the Environmental Severity Index (ESI) based on a decade-worth of field data of steel and aluminum metal plate samples (Kendall, 2013). This system categorizes corrosivity on a scale of 1-20 either through metal plate samples or through environmental measurements of local salinity and time of wetness. ESI is also tied to cost by mapping installation ESIs to one of four categories (low, moderate, high, and severe) but does "not consistently correlate with the sustainment cost data" (Silver & Gaebel, 2017, p. 30). Silver & Gaebel (2017) recommend in their report that the DoD use ISO 9223 to classify environmental corrosivity.

CHAPTER III

METHODOLOGY

Overview

The null hypothesis (H_0) for the RQ is that no statistical difference can be identified among the average CI of D305006 Package Units by corrosivity category. This can be summarized in the following statement:

$$H_0: \mu_{C2} = \mu_{C3} = \mu_{C4} = \mu_{C5} \tag{4}$$

The only corrosivity categories represented in the data are C2, C3, C4, and C5. These are ordinal categories of increasing corrosivity. The alternate hypothesis (H_A) is that there is a statistical difference between the weighted averages of D305006 Package Units component-section condition indexes by corrosivity category. This can be stated as:

$$H_A: \text{At least one mean be different from the others} \tag{5}$$

Where μ_{C1}, μ_{C2}, etc. are the average of all D305006 Package Units component-section condition indexes for the respective ISO Corrosivity Categories. If there is a statistically significant difference in CI between corrosivity categories, a pairwise analysis of each category is conducted to determine if there is a similarity or decrease in increasing corrosivity categories.

The data was provided to the author in two spreadsheets. These were converted into comma separated value (CSV) files for better interoperability with other analysis tools. SQLite was used to import these files into a SQL-based relational database. Python was the chosen programming language for handling the data. It has an application programming interface with SQLite. Finally, JMP was used to for some of its statistical analysis features in presenting the results.

Based on comparisons of using ISO 9223 and ESI for classifying environmental corrosivity, the ICCET was chosen for conducting this research. While ESI and ISO 9223 offer categorization systems based on metal plate sampling, the ISO 9223-derived ICCET provides more accurate atmospheric model-based categorization because hourly weather factors and linear distance to saltwater are used (Silver & Gaebel, 2017). Finally, the ICCET has readily available data for the locations that were analyzed. The comparison between the methods is summarized in Table 3.

Table 3. Comparison of ESI, ICCET, and ISO 9223

	ESI	**ICCET**	**ISO 9223**
Metal plate samples	Yes	N/A	Yes
1-year mass loss data	Yes	N/A	Yes
# of categories	20	6	6
Temperature	No	Yes	Yes
Humidity	No	Yes	Yes
Time of wetness	Yes	No	No
Proximity to saltwater (salinity)	Binary (within 1 mile or not)	3 models depending on distance	Chloride deposition rate used

Assumptions & Limitations

There are several assumptions that were made to conduct this analysis, and some limitations to prevent scope creep. Assumptions included the following:

- Air conditioner condition degradation is primarily affected by corrosion (as noted in the literature review there is some research to support this assumption)
- It is assumed that different brands and models of air conditioner are similarly susceptible to corrosion

The justification for these limitations is that they not only involve the need for current data for several other parameters, but also historical data. The primary dataset that was analyzed in this paper only provides a current snapshot of building condition data. Additional historical data would be required to understand the cause of current conditions. This data is not available to the author and in some cases may not exist or be useable. The current BUILDER data can be used to understand the condition of items now, and project future conditions, but there is no historical data for this available to the author.

Limitations:

- Installations have equivalent levels of funding for their infrastructure
- Installations have similar preventive maintenance completion rates
- Installations are able to accomplish corrective maintenance in a similar timeframe

Data Collection

Condition Data

The primary dataset was already provided to the author in the form of two spreadsheets, (1) containing "Component-Section Details" and (2) the other containing "Equipment Details". This data was collected as part of the USAF implementation of BUILDER by USAF military personnel, civilians, and contractors conducting facility inventories and condition assessments and entering the inspections into the database. It is maintained and updated as part of the USAF FCA program. A location anonymized copy of the 2018 data was provided to Texas A&M under a Cooperative Research and Development Agreement in 2019. Please see Table 4 for a summary of the data contained in these spreadsheets. Additionally, two tables are contained in the Appendix that list the fields in each spreadsheet, data type, and clarifying notes about what each field is. Some of these notes were provided to the author, and others are the author's inference of the information contained therein. The “Component-Section Details” and “Equipment Details” spreadsheets contain anonymized location data, but the author was provided a reference by AFIMSC identifying each of the locations for the purpose of matching the anonymized data with the correct corrosivity category.

Table 4. Data Summary

	Component-Section Details	**Equipment Details**
# of Items	1,048,575	716,973
# of Locations	59	81
# of Fields	32	35
# of D305006 Package Units	21,657	39,084

Both the “Component-Section Details” and “Equipment Details” overlap in information. They were not combined or cross-referenced for the analysis or results. “Equipment details” contains further information on certain component-sections in “Component-Section Details” that are primarily concerned with mechanical, electrical, and plumbing systems. It is also the only data source that lists the brand of certain types of equipment. They cannot be combined because there is no unique identifier for line items that allows items to be matched. Finally, the spreadsheets were both pulled from BUILDER at different dates and therefore condition values differ by a few points.

“Component-Section Details” contains data from 59 locations and represents all eight USAF Major Commands (MAJCOMS) and the U.S. Space Force (USSF). The reserve components, Air National Guard and Air Force Reserve Command, are not represented by locations that they primarily operate but they may have facilities in some of locations operated by others. The data does not include a breakdown of owning units or commands by component-section. 47 of the 59 locations are in the U.S. or its territories and 12 are located outside the U.S. Table 5 shows the number of locations by MAJCOM as well as the USSF (Wright, 2020).

Table 5. Data Summary by Service/MAJCOM

Service / MAJCOM	# of Locations
Air Combat Command	12
Air Education and Training Command	7
Air Force Global Strike Command	4
Air Force Materiel Command	4
Air Force Special Operations Command	2
Air Mobility Command	8
Pacific Air Forces	8
U.S. Air Forces in Europe – Air Force Africa	7
U.S. Space Force	7

D305006 Package Units

The D305006 Package Units component-section consists of decentralized air conditioning and heat pump systems and are also described as "unitary systems" (Howell et al., 2013, p. 411). There are six examples of such systems listed in the *USAF Built Infrastructure Inventory and Assessments Manual Appendix for HVAC (D30)* (AFCEC, 2017):

- Computer room A/C unit
- Split systems with air cooled condenser
- Split systems with air cooled condenser and options for heat pumps or electric heat
- Evaporative cooler
- Through-wall A/C unit
- Water source heat pump

These systems are all exposed to the exterior environment because they necessitate a location outside of the conditioned space to dump heat (for A/C units) or to pull heat from (for heat pumps). An analysis of "Component-Section Details" shows that these units range in size from 1/2 ton to 120 ton HVAC units. The database has 196 descriptors under the *CompType* field and the top five most frequently occurring are summarized in Table 6. There are also 602 labeled as "General", 129 labeled as "Other", and 58 labeled as "Unknown". The D305006 component-section is generalizable for this research and analyzed as a whole component-section.

Table 6. Frequency of Occurrence for Top 5 D305006 Component Types

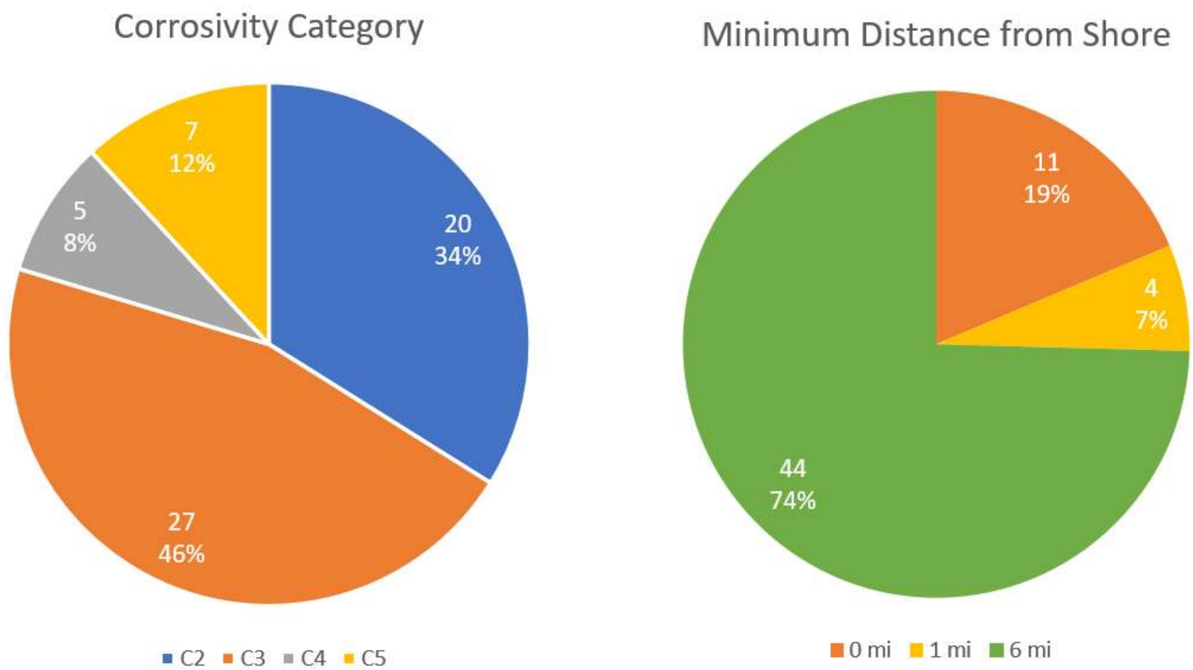
Component Type	**Frequency of Occurence**
A/C Unit, Split Systems w/ Air Cooled Condenser	3,109
A/C Unit, Split Systems w/ Air Cooled Condenser - 2 TN	1,898
A/C Unit, Split Systems w/ Air Cooled Condenser - 3 TN	895
A/C Unit, Computer Room	819
Evaporative Cooler	793

There are also decentralized A/C systems broken up across multiple component-sections to differentiate cooling generation, distribution, air handling, and terminal parts but they were not analyzed in this paper. "Component-Section Details" has 159,868 component-sections at the D30 HVAC system level and these include air handling units, fan coil units, boilers, gas supplies, hot water distribution, etc. Furthermore, there 5,360 D303001 Chilled Water Systems that include chillers and cooling towers ranging in size from 5 to 1,500 tons but some of the equipment under this category is located inside mechanical rooms and not exposed to the outside environment, such as centrifugal chillers (AFCEC, 2017).

Corrosivity Category Data

The secondary dataset involves the Corrosivity Category data. This data was obtained from two locations. The most accurate data is contained in Revision 2 of Appendix D of *Facilities Environmental Classification Study* (Silver & Gaebel, 2017), which consists of Corrosivity Categories derived from ISO 9223:2012 compliant one-year metal sample corrosivity tests. Appendix D is updated regularly, so the most recent version published to the Whole Building Design Guide on 24 August 2018 was used (Geusic, 2018). This dataset has information for 355 locations from the BUILDER dataset. For locations that do not have this data available, the Corrosivity Category was gathered using the ICCET (Gaebel, n.d.) with the final data pull on 22 September 2020. The ICCET utilizes weather data from nearby weather stations to approximate the Corrosivity Category based on ISO 9223:2012. The corrosivity category for 56 installations was obtained from the Appendix D values and the data for the remaining three installations was obtained from the ICCET.

Combining the two sources of data shows the following breakdowns for number of locations in each Corrosivity Category as well as by the minimum distance from a saltwater shoreline (Figure 5).



Corrosivity Category	Minimum Distance from Shore: 0 mi	1 mi	6 mi
C2			20
C3	2	2	23
C4	3	2	
C5	6		1

Figure 5. Corrosion Category and Minimum Shore Distance by Number of Locations

Data Cleaning

Overview

Although standardized practices were used to gather this information, the likelihood of error stemming either from incorrect condition assessments, misattribution, or data entry errors must be considered. Two primary approaches were used to eliminate data that is obviously or likely to be erroneous. The presentation *Lifecycle Management of Air Force Facilities and Assets* (Dept. of Construction Science, Texas A&M University, & ALPHA Facilities Solutions, LLC, 2019) provides a summary of various methods that were used to clean this specific data for another research project. They are summarized in Table 7.

Table 7. Data Cleaning Summary, Adapted (Dept. of Construction Science, Texas A&M University, & ALPHA Facilities Solutions, LLC, 2017)

Affected Field	**Description**
PRV	Buildings with below zero or empty values excluded
SEC_Beta	Negative values are erroneous
CRV	Component-sections with below zero or empty values excluded. Excessively high values (greater than $9.5 million) excluded. Component-sections with CRV greater than PRV excluded.
Age	Component-sections with ages 3 times greater than service life and CI greater than 39 excluded

These methods are the final result of an iterative process where ALPHA Facilities worked with AFIMSC and were very fruitful in selectively removing erroneous data They had 1,336,297 component-sections pulled from BUILDER and PAVER (the USAF SMS for paved surfaces) and 98.8% of the data was retained. In addition to the ALPHA Facilities method, data cleaning principles from Osborne (2012) and Ilyas & Chu (2019) were applied for the fields identified in Table 8. The chosen methods included outlier detection, identification of duplicates, data transformation, and rule-based data cleaning. The data cleaning was conducted in three separate phases. Phase 1 involved examining the totality of the data and removing data based on broadly applicable rules. In Phases 2 and 3, the scope of cleaning was narrowed to the D305006 PACKAGE UNITS component-sections. The data cleaning process was iterative, and so this paper presents the resulting end-process used. All phases were fully accomplished for "Component-Section Details", but only part of Phase 1 was completed for "Equipment Details." It was discovered that "Equipment Details" lacks a field for component-section age since installation date. "Equipment Details" was excluded from further analysis.

Table 8. Chosen Fields for Data Cleaning

Field	**Description**
CI	Component-section condition index
CRV	Component (component-section) replacement value
PRV	Building plant replacement value
Age	Age in years since installed or constructed
SEC_Beta	β for calculating CI
RSL	Remaining service life
InstallDate	Install or construct date
Expected_Service_Life	Overall expected service life (E in CI calculation)
SumOfRPATotalUnitOfMeasureQuantity	Component-section quantity numerical amount for RPATotalUnitOfMeasureCode.
RPATotalUnitOfMeasureCode	Component-section quantity category: each (EA), square foot (SF), square yard (SY), foot-pound (FP), or null.

Phase 1 Cleaning

One of the problem areas that arose out of initial viewing of the data concerned the *RPATotalUnitOfMeasureCode* and *SumOfRPATotalUnitOfMeasureQuantity* fields found in "Component-Section Details" and the similar fields *UoM* and *Qty* in "Equipment Details" (herein referred to as quantity type and quantity amount, respectively, for both spreadsheets). When summing or averaging CI, any component-section with a quantity type of "each" must be counted multiple times for any quantity amount greater than one. For example, a component-section for D305006 Package Units with 3 each at a CI of 75 means there are three package HVAC units with a CI of 75 in that building. They also share the same install date and age, and their combined replacement value is the CRV for that component-section. This was particularly problematic for D305006 Package Units in "Equipment Details" as summarized in the histogram and boxplot in Figure 6.

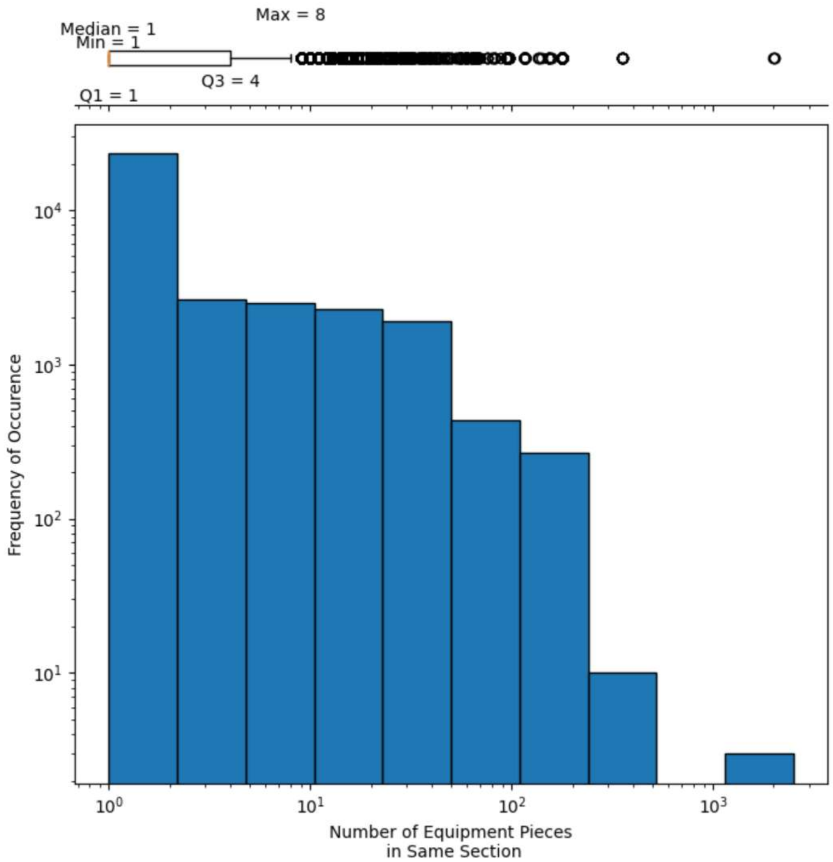


Figure 6. Equipment Details Quantity Type “Each” Histogram and Boxplot

Figure 6 was plotted log-log to best represent the extent of the data in a small space. The mode is one with a count of 20,130 component-sections, but some component-sections have zero “each” pieces of equipment and one has a value of 2007 “each”. The installation year for this piece of equipment is also 2007, so it can be concluded that the value was erroneously carried over into the wrong field. Other issues crop up as well. One building has 179 HVAC units that subsequently each have 179 HVAC units in their component-sections. Another facility has 10 HVAC units each with 355 units per component-section. It would require a per-building analysis for each building in the 81 locations represented in ‘Equipment Details’ as well as comparison against building size and type and assumptions about how the building is used. This is outside the scope of this thesis.

An addition difficulty with "Equipment Details" is that it lacks a field for age since installation. "Component-Section Details" has this field, along with a field for the year of installation. This allows for a data verification step where the sum of *Age* and *InstallDate* is the day the data was calculated and "rolled-up." There should be the same number for all component-sections to ensure consistency and that all component-sections are being compared in the same time snapshot. A graphical summary of the distribution for "Year Rolled-Up" in "Component-Section Details" can be found in Figure 7.

Due to the aforementioned problems with "Equipment Details" further analysis on this spreadsheet was discontinued and only "Component-Section Details" is fully cleaned and used for evaluation. The "rolled-up" year was set to 2018 and this eliminated 54.7% of the 1,048,575 component-sections. Additionally, the other fields mentioned in Table 8 were scrubbed for abnormalities that are broadly applicable. For *CRV* and *PRV*, all values less than or equal to zero were removed. *CRV*s greater than *PRV* were also removed because the sum of *CRV*s for a building should at most equal the *PRV* for that building. Finally, *SEC_Beta* values less than zero were removed because they imply a negative remaining service life. A summary can be found in Table 9.

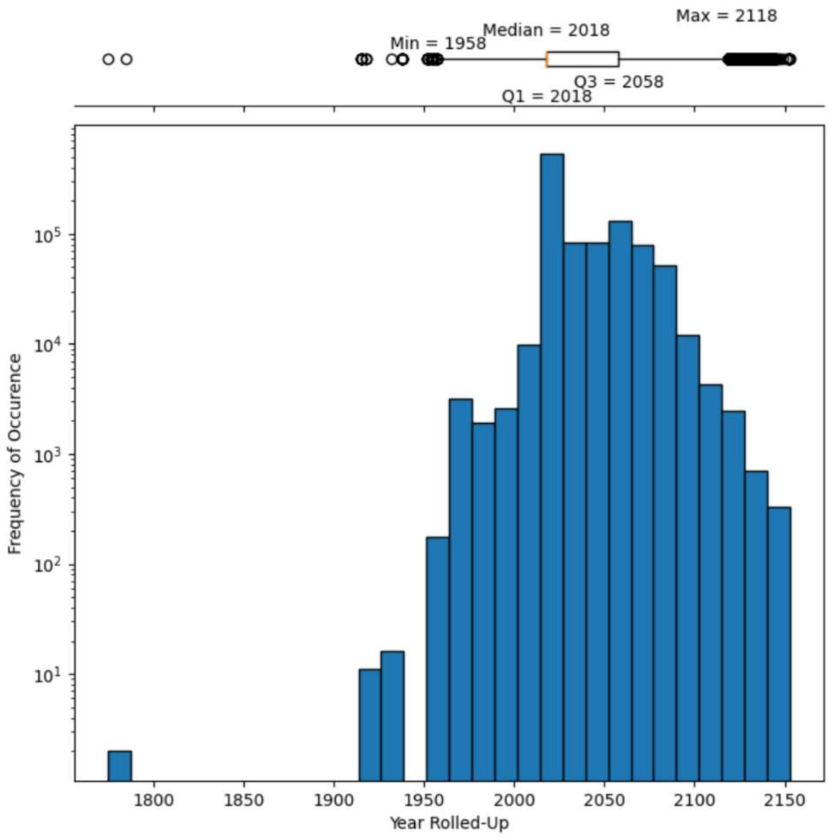


Figure 7. Sum of *Age* and *InstallDate* Histogram and Boxplot

Table 9. Phase 1 Data Removal for 'Component-Section Details'

Rule	**% All Removed**	**# All Removed**	**% D305006 Removed**	**# D305006 Removed**
Age + InstallDate = 2018	54.68	573,377	73.50	15,918
Cost Factors - CRV > 0 - PRV > 0 - PRV > CRV	8.16	85,584	4.39	951
SEC_Beta > 0	0.29	3,015	0.24	53
Result	56.32	590,377	74.26	16,083

Table 9 also includes the impact of Phase 1 of cleaning on the D305006 Package Units. Out of the 21,657 such units in the database, only 5,574 were retained for further cleaning and analysis. The effect of setting the "rolled-up" year to 2018 is more pronounced on the Package

Units component-section, with 73.5% removed compared to the overall 54.7% of all component-sections removed.

Phase 2 Cleaning

Phase 2 of cleaning was focused exclusively on examining the relevant fields from Table 8 for D305006 Package Units. Here, the approach was to explore the summary statistics and graphical representations of data. As in “Equipment Details”, “Component-Section Details” has apparent problems with the “each” type for measuring the number of equipment units per component-section. The analysis for this was saved until Phase 2 for “Component-Section Details” so that only one type of equipment, the D305006 Package Unit, is analyzed. Only 8 of 5,574 D305006 Package Unit component-sections has an “each” count. Seven of these have a count of one unit per component-section, and one is a count of 8,000 units per component-section. The component-section with 8,000 units was therefore removed. A further look at this specific building shows that it is unlikely to have 8,000 Package Units and this value is entered in error. Several other building component-sections such as C101001 Fixed Partitions, B202001 Windows, and D501001 Main Transformers also have a count of 8,000 units per component-section (47 component-sections total have this value for this building).

Age and Install Date were examined next. It can be seen in Figure 8 that Age has a strong positive skew of 1.22 and half of the units are less than 18 years old. The furthest outlier is 98 years old, putting this unit at an installation year of 1920. The arithmetic mean can be found in the third quartile at 23.6 years.

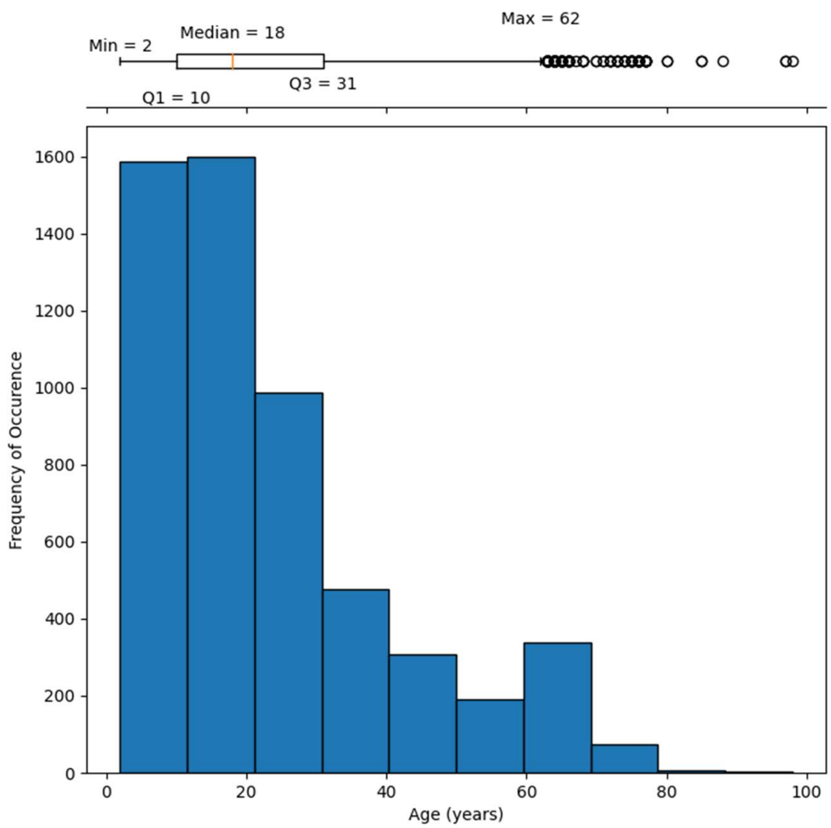


Figure 8. Age Histogram and Boxplot

Install Date is depicted in Figure 9 and complements the Age data in Figure 8 with a negative skew of -1.2. To limit the number of outliers, the maximum age is limited to units 76 years old or younger. This is equivalent to an Install Date of 1942. The age limit was based off the mean age of 23.6 years plus three standard deviations of 17.4 years (52.2 years). 50 component-sections were removed under this criterion. HVAC equipment used for comfort cooling has existed since the 1890s, and HVAC was installed in U.S. federal government buildings such as the Chamber of the House of Representatives in 1928. The rooftop and package units we are familiar with today did not start appearing until the 1950s. (Howell et al., 2013). Therefore, it is unlikely that any HVAC units built prior to 1942 are still in operation

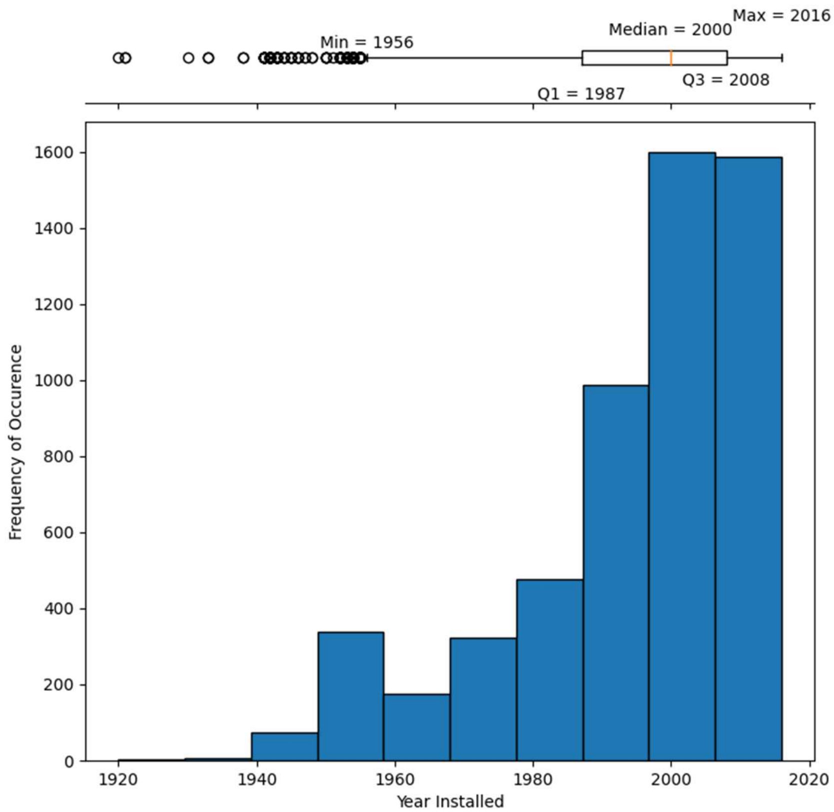


Figure 9. Install Date Histogram and Boxplot

CRV was analyzed but no significant discrepancies were found. The Phase 1 elimination of CRV greater than PRV eliminated the more egregious extremes, such as one component-section with a CRV of $46 million in a building with a PRV of $224,000. In Phase 2, the maximum CRV is $2.9 million. Figure 10 shows the summary data for CRV in a log-log plot to best encapsulate the extend of the data. If not adjusted, the skew for the CRV is 9.7, but by taking the $\log_{10}$ of the CRV values, the skew is reduced to -0.4. The median cost is $29,000 and the average cost is $48,000. Figure 11 shows the probability plot for the logarithmically adjusted data. There are more component-sections with low values than compared to a normal distribution and a kurtosis of -0.48 due to less component-sections existing in general in the tails of the distribution.

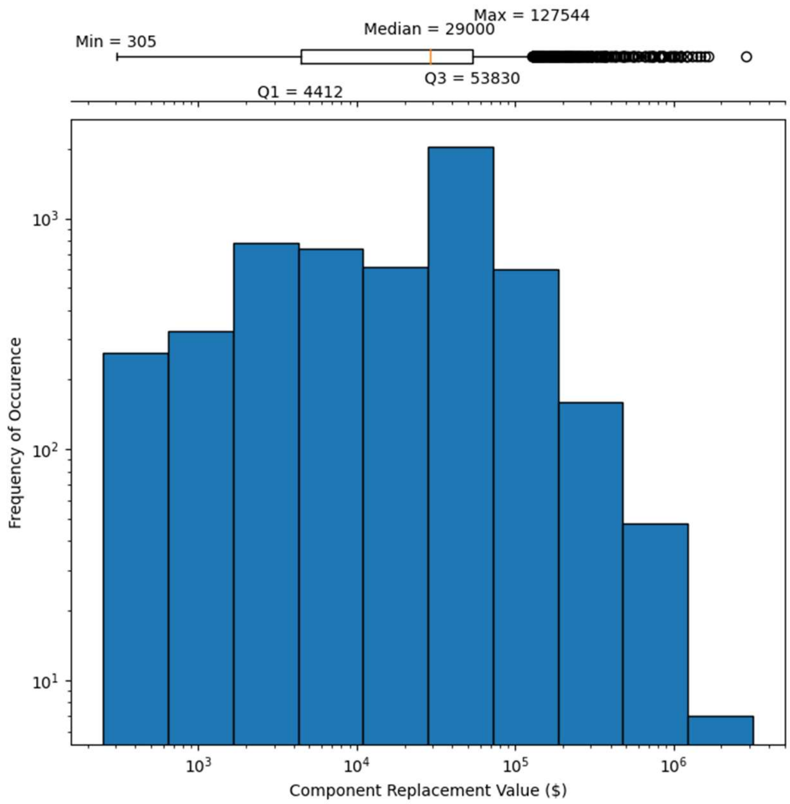


Figure 10. CRV Histogram and Boxplot

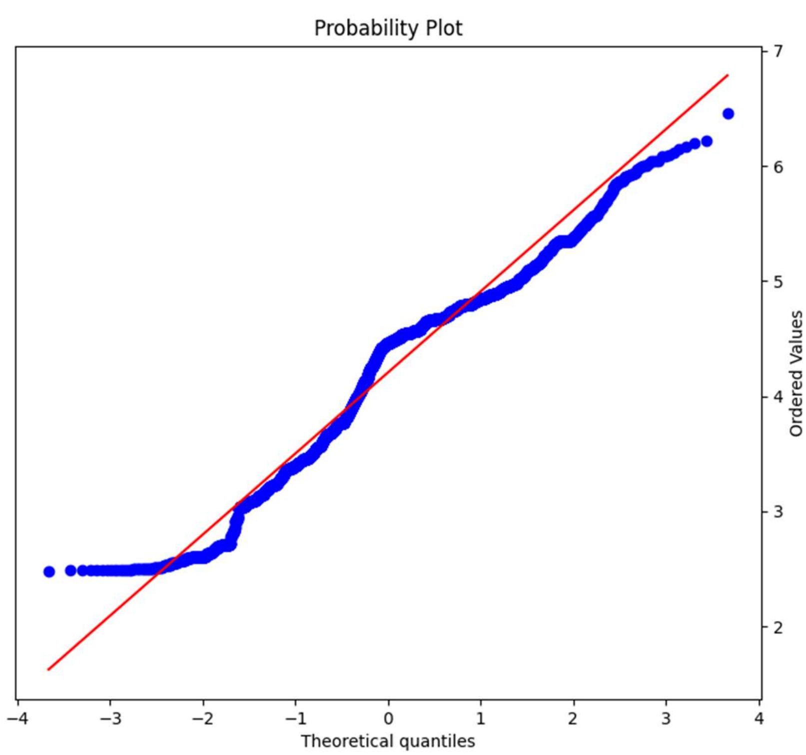


Figure 11. CRV Probability Plot

RSL and Expected Service Life were examined with no component-sections subsequently removed. RSL values are all between 0 and 20, where 20 is the maximum value of Expected Service Life found in D305006 Package Units. When a component-section reaches a CI of 40, its RSL becomes zero. Figure 12 summarizes the remaining service life. The median component-section has seven years of service life remaining and the average is 6.5 years. 844 component-sections represent the mode value of zero years remaining service life. The distribution has low skew of 0.23 but this amount is still statistically significant at a p-value approaching zero.

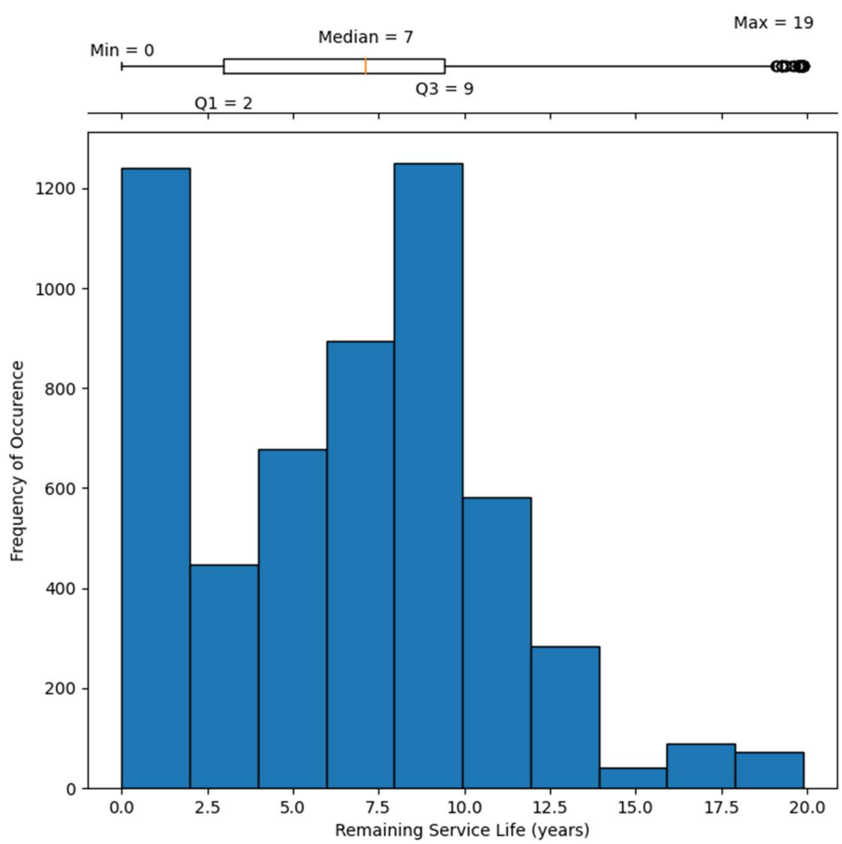


Figure 12. RSL Histogram and Boxplot

The Expected Service Life values are either 10, 15, or 20 years for D305006 Package Units summarized in Table 10. 91% of the 5,574 component-sections have a service life of 20 years. In summary, after Phase 2 of cleaning there are 5,523 component-sections or 99% of D305006 Package Units remaining from the Phase 1 pass of 5,574 component-sections.

Table 10. Expected Service Life Quantities

Expected Service Life	**# Of**
10 years	78
15 years	411
20 years	5,085

Phase 3 Cleaning

Phase 3 of cleaning did not involve the removal of any further component-sections but was used to characterize the types of distributions of data fields that would be used for analysis. The primary field for testing the hypothesis is that of CI in Figure 13. The mean CI for D305006 Package Units is 67.5, and the median is 76. The data has strong negative skew at -1.14 and positive kurtosis of 0.51. To aid in analysis, the skew can be reduced by means of a data transformation. Three different transformations were considered for CI. These are a square root, natural logarithm, and logarithm base 10. The results of these transformations are summarized in Table 11. The log base 10 transformation was chosen for use in assisting further analysis.

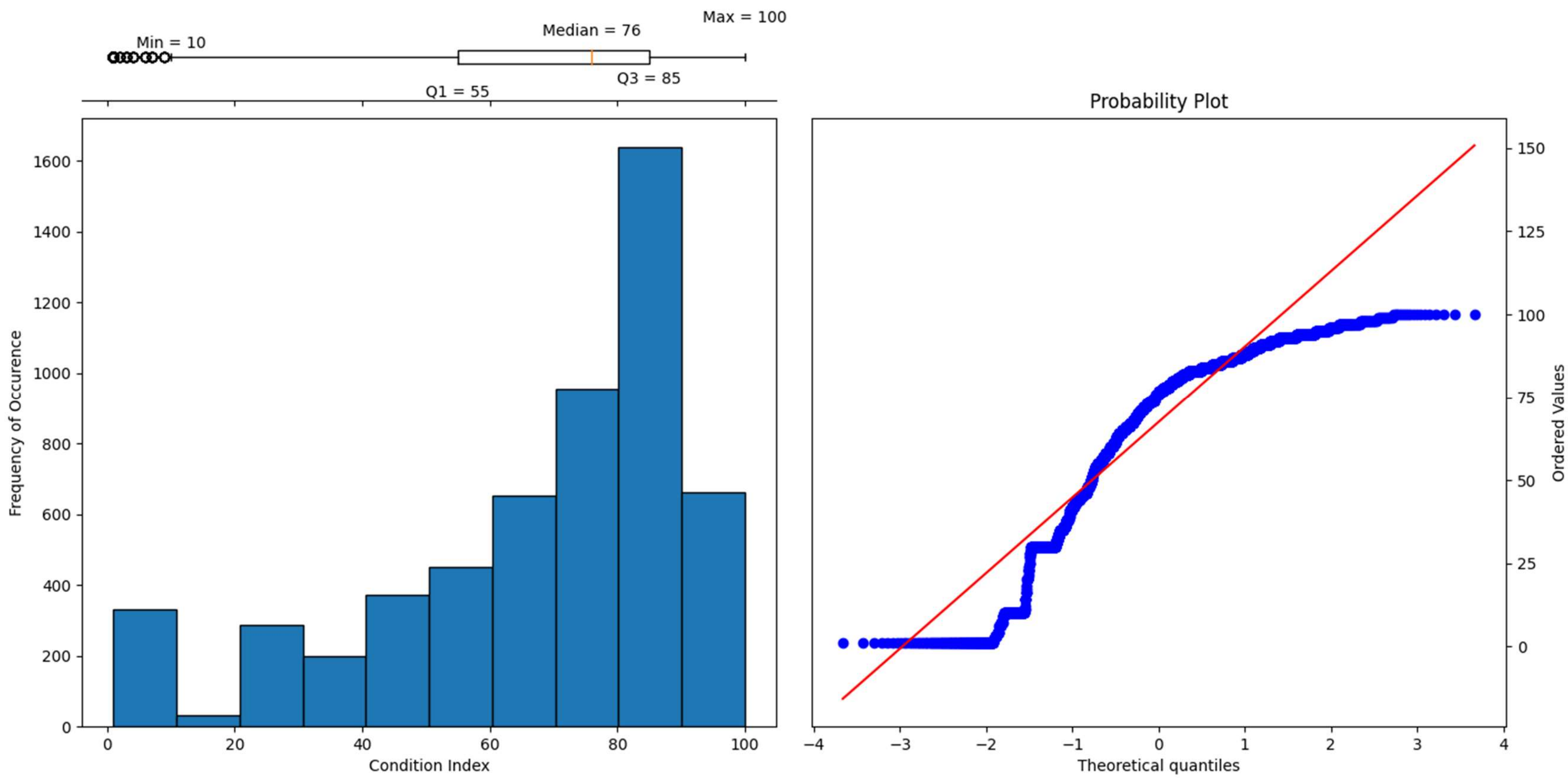


Figure 13. CI Histogram, Boxplot, and Probability Plot

Table 11. CI Transformation Comparison

Transformation	Skew	Kurtosis
Original	-1.1393	0.5110
Square root	0.5328	-0.4481
Natural log	-0.4060	0.4479
Log base 10	-0.4060	0.4479

The improvement in normality for the distribution is shown in Figure 14. The equations for performing the transformations are the following:

$$\text{Square Root} = \sqrt{(101 - \text{CI})} \tag{6}$$

$$\text{Natural log} = \ln(101 - \text{CI}) \tag{7}$$

$$\text{Log base 10} = \log_{10}(101 - \text{CI}) \tag{8}$$

These transformations were conducted as recommended by Osborne (2002). The distribution is negative, so the data must be reflected by making it in negative. Since these functions require positive numbers, the data was subtracted from its maximum value of 100 plus 1. The size of original, non-transformed data is n = 5,574 and per Hazra & Gogtay (2016) sample sizes greater than 100 constitute large distributions and can “nearly always be analyzed with parametric tests” (p. 252).

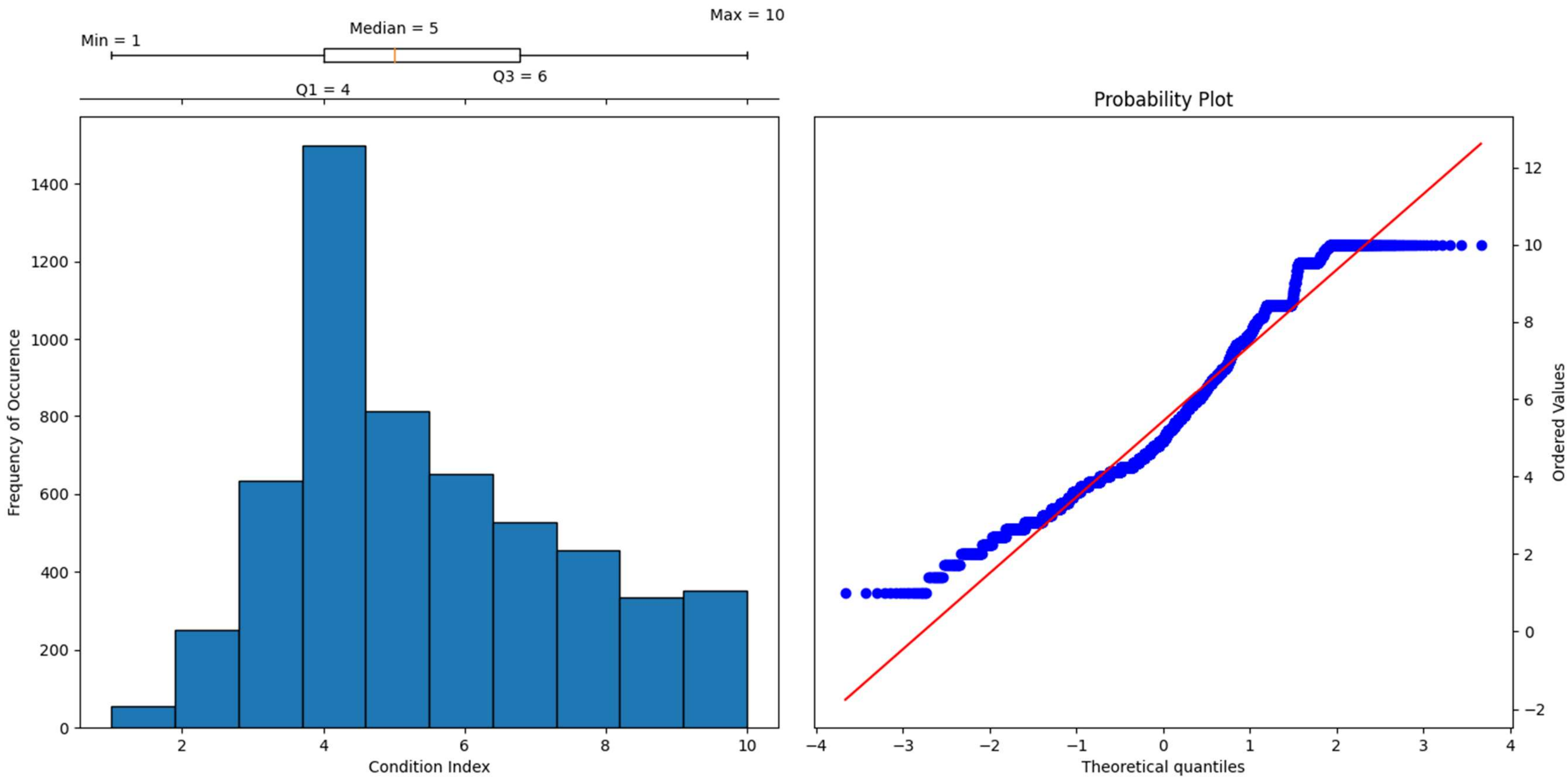


Figure 14. Transformed CI Histogram, Boxplot, and Probability Plot

RSL was examined but does not make a good fit for transformation due to its bimodal nature. As can be seen in Figure 15, 831 component-sections have an RSL of zero. Chapter IV provides further analysis on RSL by way of linear regression with *CI*.

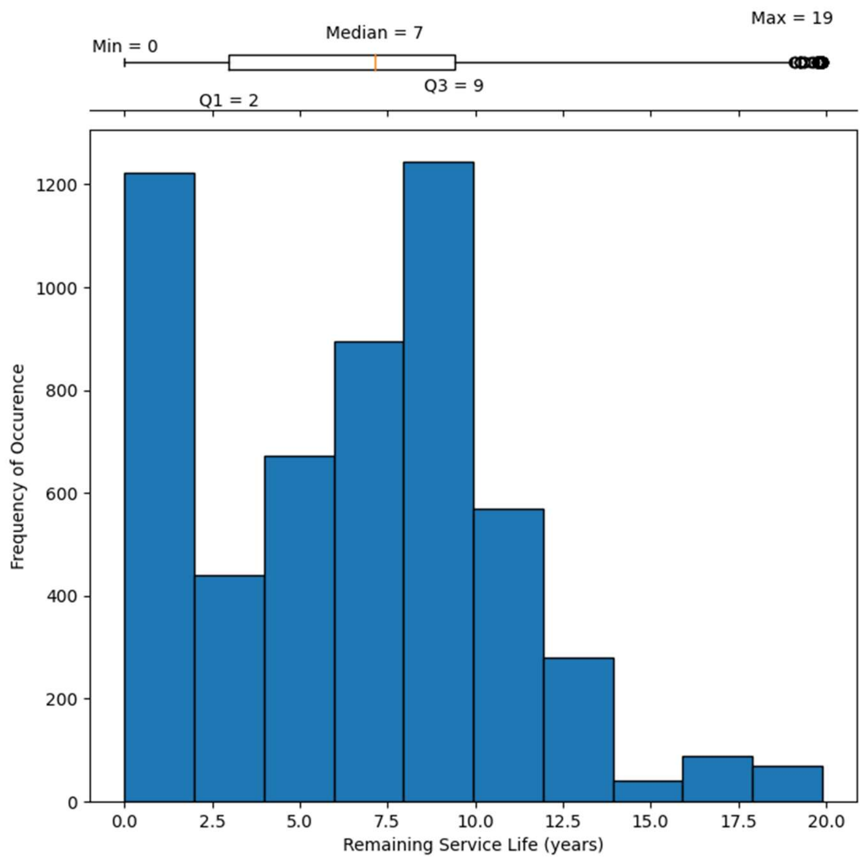


Figure 15. RSL Histogram and Boxplot

Age was transformed to provide a better fit to a normal distribution. The original data is represented in Figure 16 and the chosen log base 10 transform data is in Figure 17.

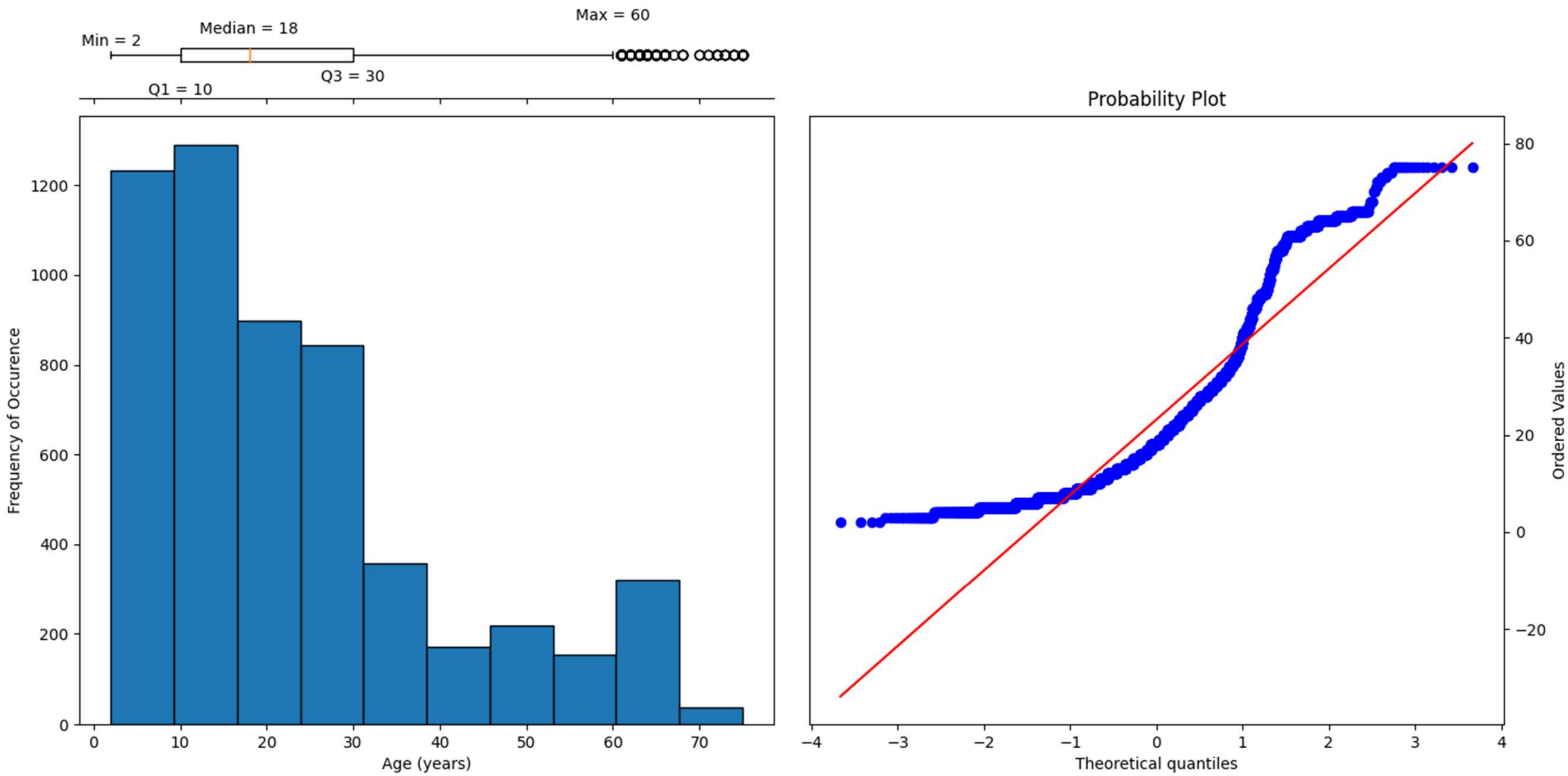


Figure 16. Age Histogram, Boxplot, and Probability Plot

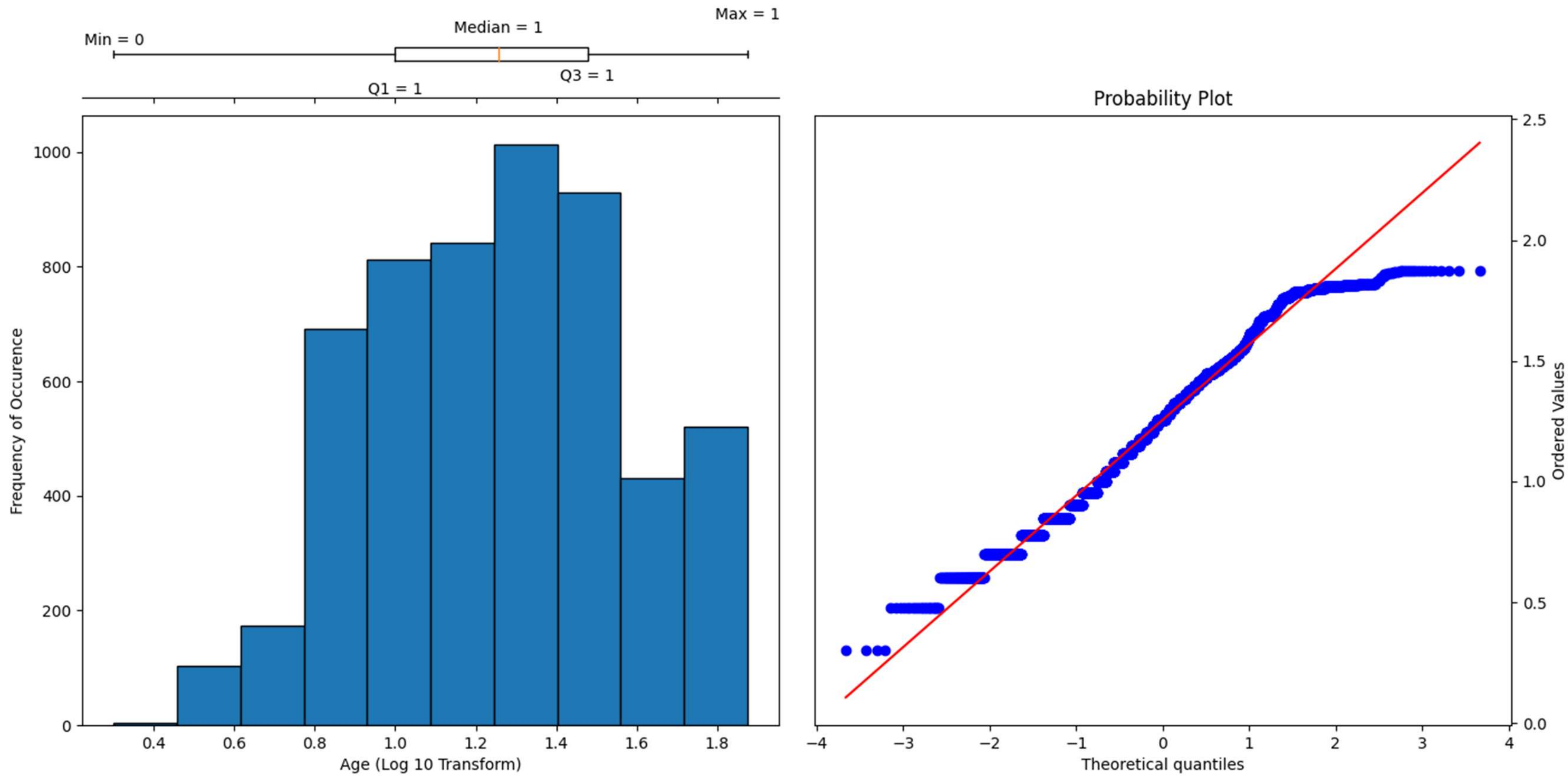


Figure 17. Transformed Age Histogram, Boxplot, and Probability Plot

The equations used for transforming Age are as follows. Age was originally positively skewed so it can be inputted raw into the transformation function without shifting or reflecting it:

$$\text{Square Root} = \sqrt{\text{Age}} \tag{9}$$

$$\text{Natural log} = \ln(\text{Age}) \tag{10}$$

$$\text{Log base 10} = \log_{10}(\text{Age}) \tag{11}$$

A comparison of different transformation methods is summarized in Table 12. Log base 10 performed the best and reduced the skew from 1.16 to -0.06.

Table 12. Age Transformation Comparison

Transformation	Skew	Kurtosis
Original	1.1629	0.5335
Square root	0.5940	-0.4577
Natural log	-0.0581	-0.7312
Log base 10	-0.0581	-0.7312

Methods of Analysis

Three analytic tools were used in Chapter IV for conducting analysis of the data and results are presented there. Analysis of Variance (ANOVA) was used for determining statistical differences between groups. Tukey's Honest Significant Difference (HSD) was used to conduct pairwise comparison of data sets between different categories in those data sets. Linear regression was used to understand the relationships between continuous variables. Here is a summary of the statistical tests and analysis conducted for the "Component-Section Details" data set:

- One-way ANOVA roll-up of D305006 Package Unit CI by location corrosivity category
- Tukey's HSD of D305006 Package Unit *CI* by location corrosivity category
- Linear regression of natural logarithm transformed CI data by RSL

- One-way ANOVA roll-up of D305006 Package Unit Age by location corrosivity category
- Tukey's HSD of D305006 Package Unit *Age* by location corrosivity category

CHAPTER IV

RESULTS

Condition Index

CI was broken divided into the four available categories found in the data, C2, C3, C4, and C5. To conduct an ANOVA, three assumptions must be met:

- Independence of samples. The samples are all independently taken and do not depend on one another, so this condition is met.
- Normality of distributions. The skew and kurtosis of the population are -1.14 and 0.52, respectively. Blanca et al. (2017a) conducted a Monte Carlo simulation study of 22 different distributions and found that for three groups, a skew of 2 and kurtosis of 6 were robust and had Type I error rates less than 5%.
- Equality of variances. Blanca et al. (2017b) conducted a Monte Carlo simulation study of various scenarios for different variance ratios, pairings, and coefficient of sample size variations. The procedure provided on p. 945 was followed to test for this assumption:
    - The variance ratio for the CI data is 1.6, which is greater than the target of 1.5.
    - Calculate correlation of group sample size and variances. Per Figure 18, this correlation of 0.08 is close to 0 and therefore the ANOVA can be conducted.

With the assumptions met, CI was be tested in an ANOVA. Figure 19 depicts the box plots and grand mean for this data set. Corrosivity category C2 has the largest range and C3 the smallest. The grand mean is depicted as the gray line.

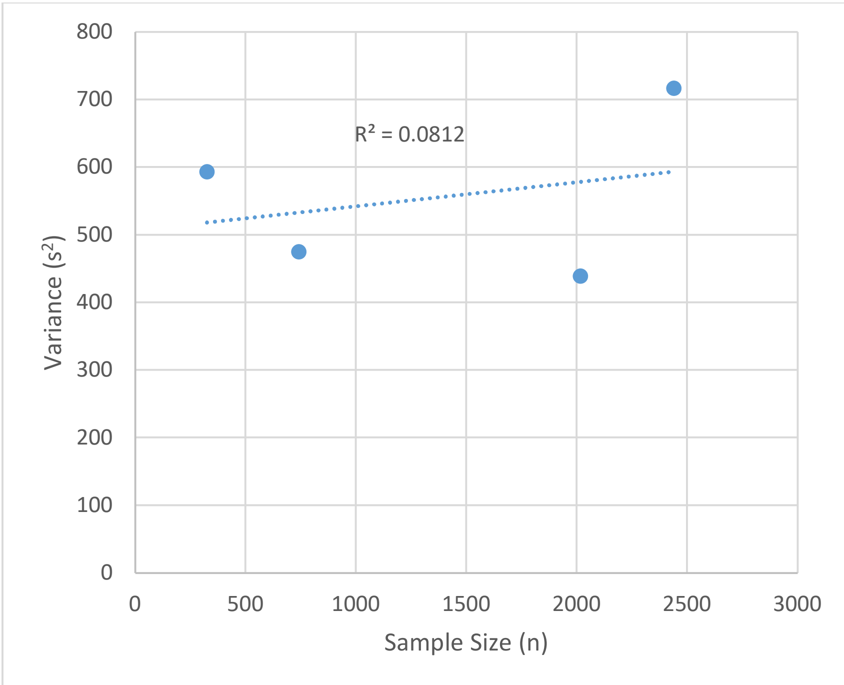


Figure 18. Correlation of Sample Size and Variance for CI Corrosivity Categories

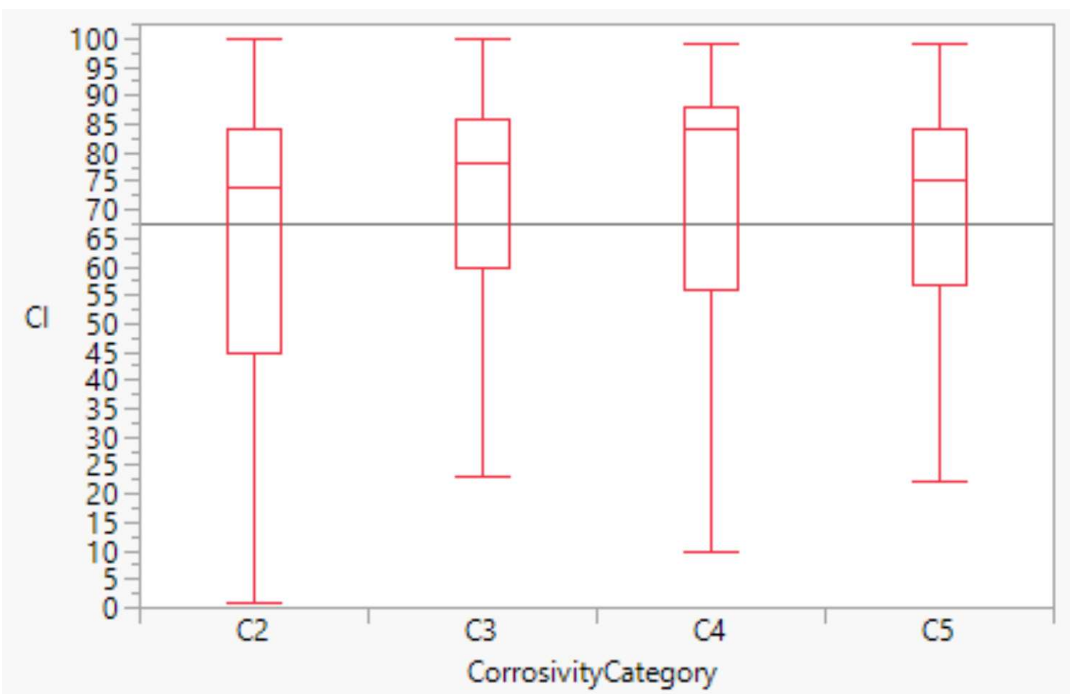


Figure 19. CI Boxplots

The ANOVA results are summarized in Figure 20. The categories are statistically different with a p-value approaching zero. The Tukey HSD is shown in Figure 21.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
CorrosivityCategory	3	55046.6	18348.9	31.8900	<.0001*
Error	5519	3175517.5	575.4		
C. Total	5522	3230564.0			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
C2	2440	64.2344	0.4856	63.282	65.186
C3	2017	71.0550	0.5341	70.008	72.102
C4	325	70.7600	1.3306	68.152	73.368
C5	741	67.9406	0.8812	66.213	69.668

Std Error uses a pooled estimate of error variance

Figure 20. CI ANOVA Results

Connecting Letters Report

Level				Mean
C3	A			71.055032
C4	A	B		70.760000
C5		B		67.940621
C2			C	64.234426

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
C3	C2	6.820606	0.721856	4.96557	8.67564	<.0001*
C4	C2	6.525574	1.416407	2.88567	10.16548	<.0001*
C5	C2	3.706195	1.006132	1.12062	6.29177	0.0013*
C3	C5	3.114411	1.030415	0.46644	5.76239	0.0134*
C4	C5	2.819379	1.595897	-1.28178	6.92054	0.2896
C3	C4	0.295032	1.433758	-3.38946	3.97952	0.9969

Figure 21. CI Tukey HSD Results

On a pairwise basis, C3 and C4 are not statistically different (p-value = 0.9969) and C4 and C5 are not statistically different (p-value = 0.2896). The ranking of mean CI from high to low is C3, C4, C5, and C2. C2 does not match expectations for it having the highest CI and is statistically isolated as the lowest mean CI (highest p-value = 0.0013). C2 also has the lowest standard error of 0.4856. The ratio of largest to smallest sample size is 7.5 from C2 to C4. The unequal sample sizes cannot be corrected for, because most locations in the database are not located in high corrosion zones.

To strengthen the ANOVA result, a test was conducted of log base 10 transformed CI data.

- Independence of samples. No change, samples independent.
- Normality of distributions. The skew and kurtosis of the population are -0.41 and 0.47, respectively. Blanca et al. (2017a) reports that values less than one maintain F-test robustness.
- Equality of variances:
    - The variance ratio for the CI data is 1.4, which is within the target of 1.5.

All assumptions for this ANOVA are met and the boxplot are summarized in Figure 22. Due to the nature of the transformation, lower values represent better condition and higher values worse condition. The ANOVA results are summarized in Figure 23. The F-ratio of 17.5 is lower than the F-ratio of 31.9 in the untransformed data, but the mean CI values across corrosivity categories remain statistically different at p-value approaching zero.

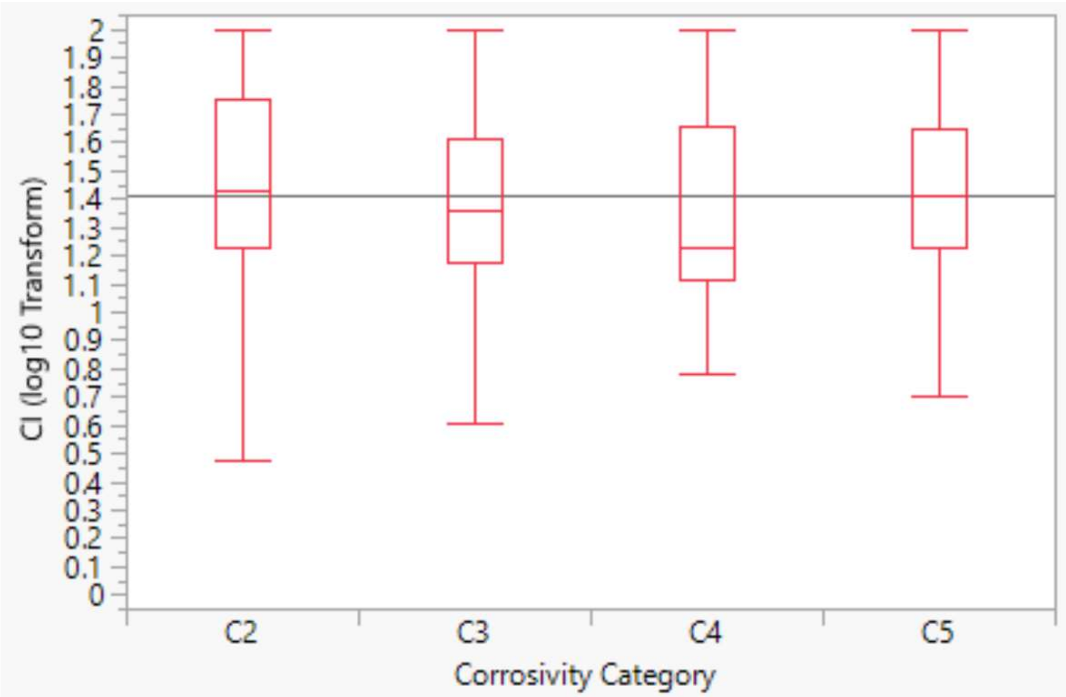


Figure 22. CI log base 10 Boxplots

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Corrosivity Category	3	5.81016	1.93672	17.4890	<.0001*
Error	5519	611.16865	0.11074		
C. Total	5522	616.97880			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
C2	2440	1.43816	0.00674	1.4250	1.4514
C3	2017	1.37389	0.00741	1.3594	1.3884
C4	325	1.34985	0.01846	1.3137	1.3860
C5	741	1.42212	0.01222	1.3981	1.4461

Std Error uses a pooled estimate of error variance

Figure 23. CI log base 10 ANOVA Results

A change in Figure 24 is that the mean CI values are paired in two statistically similar groups, C2 and C5; and C3 and C4. C2 still has the "worst" CI at the transformed value of 1.44. C4 now has the "best" CI at the transformed value of 1.35, but this is not statistically different enough from C3 (p-value = 0.6212).

Connecting Letters Report

Level			Mean
C2	A		1.4381612
C5	A		1.4221152
C3		B	1.3738936
C4		B	1.3498493

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
C2	C4	0.0883119	0.0196499	0.037815	0.1388086	<.0001*
C5	C4	0.0722660	0.0221400	0.015370	0.1291617	0.0061*
C2	C3	0.0642676	0.0100144	0.038532	0.0900027	<.0001*
C5	C3	0.0482216	0.0142950	0.011486	0.0849572	0.0042*
C3	C4	0.0240443	0.0198907	-0.027071	0.0751596	0.6212
C2	C5	0.0160459	0.0139582	-0.019824	0.0519158	0.6587

Figure 24. Log base 10 CI Tukey HSD Results

Remaining Service Life

RSL was correlated with CI to determine if its behavior is different enough to warrant further investigation by ANOVA. The raw correlation is shown in Figure 25. As previously mentioned, RSL cannot be negative so all the values are truncated at a CI of 40 to equal zero, giving the data a discontinuous slope at this point. By taking the log base 10 of CI as in Equation 8, a linear regression can be conducted. This is depicted in Figure 26.

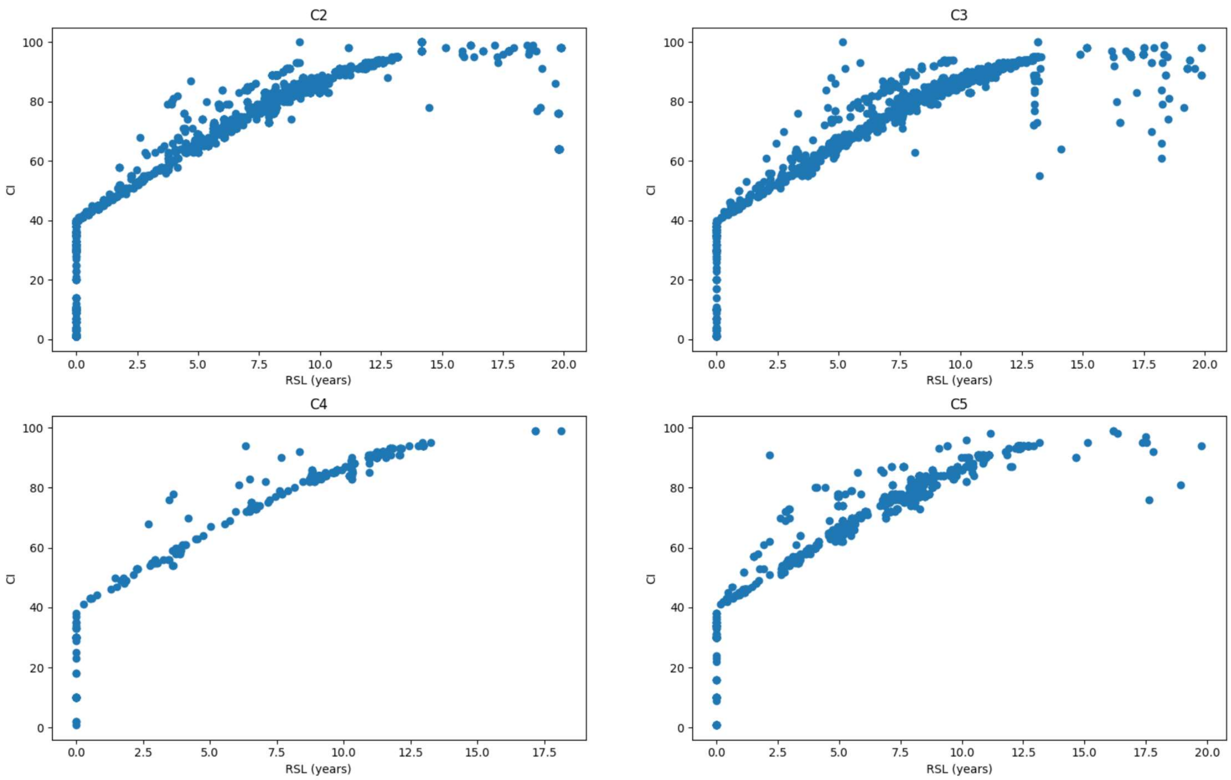


Figure 25. Correlation of RSL and CI Raw Data

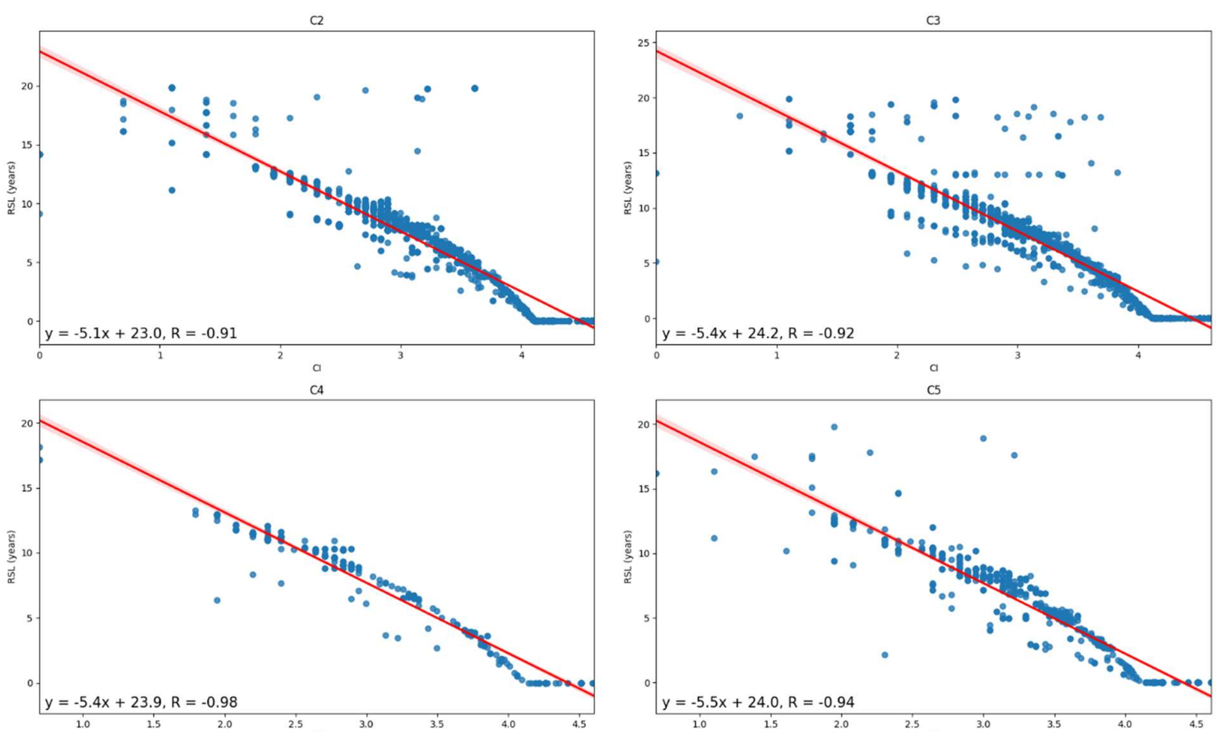


Figure 26. Correlation of RSL and CI Transformed Data

Correlation coefficients are greater than 0.91 for all corrosion categories. This strong fit occurs even with the RSL = 0 values included. The correlation results match the expectation from Equation 2, where RSL can be derived from the expected service life and time in years.

Age

Age was analyzed in a similar fashion to CI for raw and transformed data. Testing the ANOVA assumptions:

- Independence of samples. Sample ages are independent.
- Normality of distributions. The skew and kurtosis of the population are 1.16 and 0.53, respectively. These are lower than the 6 skew and kurtosis 2 values tested by Blanca (2017a).
- Equality of variances:
    - The variance ratio for the CI data is 2.2, which is greater than the target of 1.5.
    - Calculate correlation of group sample size and variances. Per Figure 27, the correlation of 0.15 is close to 0 and therefore the ANOVA can be conducted.

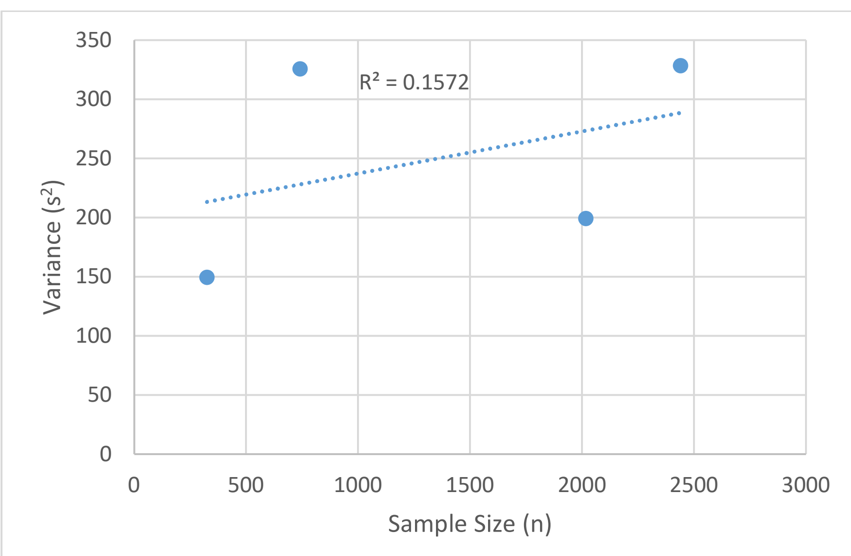


Figure 27. Correlation of Sample Size and Variance for *CI* Corrosivity Categories

Figure 28 shows that C4 has the narrowest range of ages with approximately half younger than 15 years since installation. C5 has the largest range with some component-sections nearly 70 years old. The ages across different corrosivity categories is statistically different at a p-value approaching zero, depicted in Figure 29.

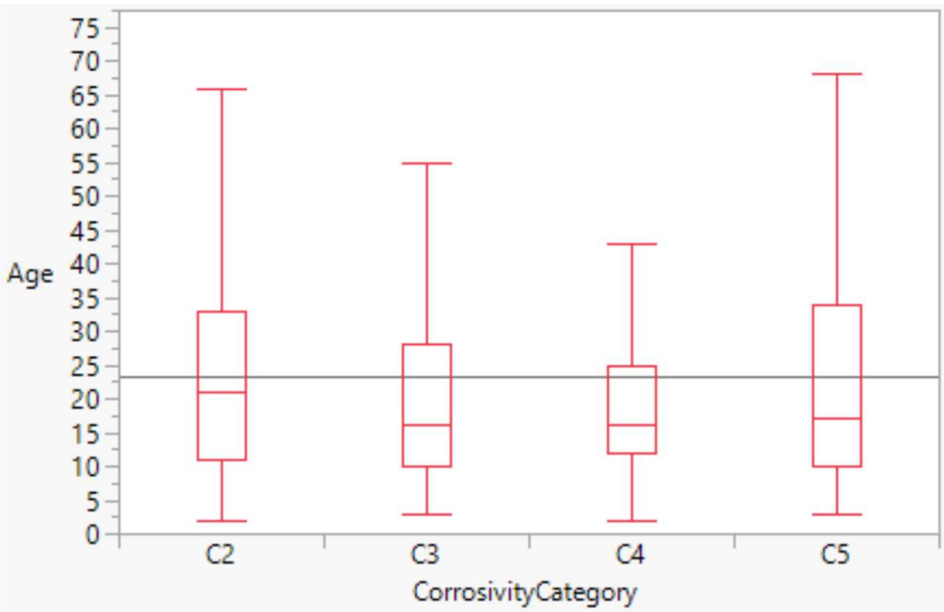


Figure 28. Age Boxplots

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
CorrosivityCategory	3	34953.3	11651.1	43.0657	<.0001*
Error	5519	1493124.8	270.5		
C. Total	5522	1528078.1			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
C2	2440	25.4811	0.33298	24.828	26.134
C3	2017	20.4294	0.36624	19.711	21.147
C4	325	19.1138	0.91238	17.325	20.902
C5	741	24.5587	0.60424	23.374	25.743

Std Error uses a pooled estimate of error variance

Figure 29. Age ANOVA Results

As shown in Figure 30, C2 and C5 are paired statistically significant (p-value = 0.5393) and C3 and C4 are statistically significant pairs (p-value = 0.5386). There is a mean difference of approximately four to five years between the two groups. The order of the groups is the same order as the low to high *CI* found in the log base 10 transformed data in Figure 24.

Connecting Letters Report

Level			Mean
C2	A		25.481148
C5	A		24.558704
C3		B	20.429351
C4		B	19.113846

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
C2	C4	6.367301	0.971245	3.87138	8.863221	<.0001*
C5	C4	5.444858	1.094323	2.63265	8.257067	<.0001*
C2	C3	5.051797	0.494984	3.77978	6.323815	<.0001*
C5	C3	4.129354	0.706566	2.31361	5.945098	<.0001*
C3	C4	1.315504	0.983143	-1.21099	3.842000	0.5386
C2	C5	0.922443	0.689915	-0.85051	2.695397	0.5393

Figure 30. Age Tukey HSD Results

The final test conducted is for log transformed Age data to verify the results. Confirming the assumptions for ANOVA:

- Independence of samples. No change, samples independent.
- Normality of distributions. The skew and kurtosis of the population are -0.06 and -0.73, respectively. These values are within the less than one tolerance for maintaining F-test robustness.
- Equality of variances:

- The variance ratio for the CI data is 1.5, which meets the target of 1.5 needed for no further checking.

Figure 31 summarizes the boxplots for the transformed Age data. Higher values represent younger component-sections and lower values represent older component-sections. The ranges are similar, with C4 as the shortest.

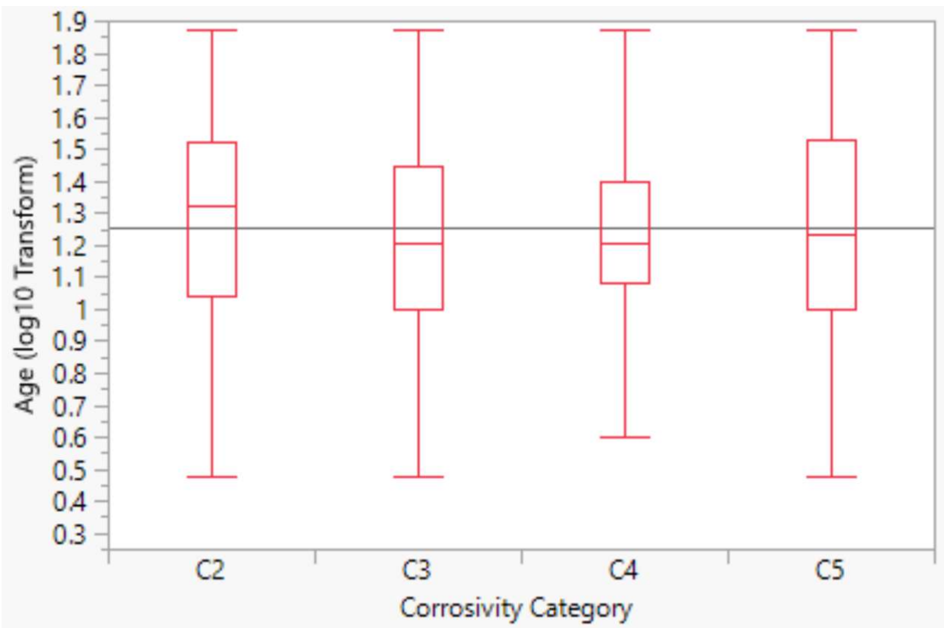


Figure 31. Log base 10 Age Boxplots

In Figure 32, the previous results are repeated for a statisically significant different between categories (p-value approaches zero). The ordering and pairing of values is preserved in Figure 33 as was shown in Figure 30.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Corrosivity Category	3	6.86713	2.28904	23.1893	<.0001*
Error	5519	544.78681	0.09871		
C. Total	5522	551.65395			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
C2	2440	1.28855	0.00636	1.2761	1.3010
C3	2017	1.21635	0.00700	1.2026	1.2301
C4	325	1.20060	0.01743	1.1664	1.2348
C5	741	1.26908	0.01154	1.2465	1.2917

Std Error uses a pooled estimate of error variance

Figure 32. Log base 10 Age ANOVA Results

Connecting Letters Report

Level			Mean
C2	A		1.2885521
C5	A		1.2690808
C3		B	1.2163470
C4		B	1.2005967

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
C2	C4	0.0879555	0.0185521	0.040280	0.1356310	<.0001*
C2	C3	0.0722052	0.0094549	0.047908	0.0965025	<.0001*
C5	C4	0.0684841	0.0209031	0.014767	0.1222012	0.0058*
C5	C3	0.0527338	0.0134964	0.018050	0.0874170	0.0005*
C2	C5	0.0194714	0.0131783	-0.014395	0.0533373	0.4512
C3	C4	0.0157503	0.0187794	-0.032509	0.0640099	0.8360

Figure 33. Log base 10 Age Tukey HSD Results

CHAPTER V

DISCUSSION

The primary objective of finding a statistically significant difference in mean CI between corrosion categories was met. The result of interest is that the Tukey HSD and summary statistics revealed that corrosion category C2 had the lowest mean CI. Outside of C2, the mean CI values in descending order are C3, C4, and then C5. C3 and C4 are statistically similar, so it is difficult to discern if there truly is a decrease from one category to the next. This pattern is close to the one shown in the log base 10 transformed CI data. However, Osborne (2002) notes that transformation of data preserves the *order* of values, but not necessarily the ratios or differences between them. Due to the reversal by subtraction and logarithm taken, the upper CI values are condensed, and the difference in the lower CI values is more greatly exaggerated. This is the intent of the transform, and becomes more apparent in that C4 bridges between C3 and C5 in the original data but does not do so and is independent of C2 and C5 in the transformed data.

As demonstrated in Figures 25 and 26, CI has a strong correlation with RSL. Values "above" the regression lines in Figure 26 have lower Cis than expected for RSL and the opposite is true for values "below" the regression line. It is also apparent that some component-sections are on different "tracks" as seen in the C3 subplot of Figure 26. These likely have similar α values (Equation 2) where α determines the acceleration of deterioration. Since RSL correlates so well with CI, it can be inferred that the C2 and C5 locations also have the lower RSL values, while C3 and C4 have higher RSL values.

The results for Age are similar in both the raw and log base 10 transformed variants. Table 13 shows a comparison of the ranking of Age vs CI by corrosivity categories. They are ranked in order of "worst to best" for each value, so low to high for CI and old to new for Age.

Table 13. Comparison of CI and Age Ranking

CI (low to high)	Age (old to new)
C2	C2
C5	C5
C4	C3
C3	C4

Age is not a primary factor for determining CI since component-sections can undergo multiple repair and part replacements over their lifetime. A repair can bring an aged component-section back to a CI of 95. Nevertheless, it could be that Age and CI exhibit some relationship by corrosion category that results in the category with the oldest inventory also exhibiting the worst CI. Excluding outliers, the C2 category also had the widest range of CI from 1 to 100 whereas the next largest range was C4 at 10 to 100.

There are three major possibilities for the discrepancy with C2. This could be due to (1) the limitations made in Chapter III, (2) errors in the data, or (3) flaws in the premise of the assumptions in Chapter III. The locations in the data provide a good snapshot of active duty (AD) USAF locations at permanent locations across the global and across mission types. Most of the D305006 sections tested are from C2 environments, but the largest category by number of locations was C3.

If there are significant discrepancies in funding or the amount of preventive maintenance or corrective maintenance completed by different corrosivity categories, then these factors may overwhelm the impact of corrosion or amplify to a high degree. Further information and research would be needed to decouple the impact of funding and work completion. The analysis conducted here is also predicated on looking at a single snapshot in time of BUILDER data, and not how condition changes over time. This is a problem that is not easily solved because BUILDER is not fully implemented across the USAF and it will take time to further develop historical data on how CI changes temporally and across repair and replace cycles. There is no equivalent data set or system similar to BUILDER data available for historical research.

Additionally, there could be errors in the data. As part of the cleaning efforts in Chapter III, over 75% of the data for D305006 Package Units was removed due to discrepancies and data that appeared to be unreasonable. Errors are two-fold, good data could have been rejected and bad data could have been kept. Ultimately, these errors cannot be fully corrected or accounted for while one is removed from the physical assets that this data represents. Another issue could be errors due to outdated data. Since condition assessments are conducted at a rate of 20% of total building area per year, some condition information could be five years old by the time it is analyzed. Although BUILDER projects out CI based on its model, faster or slower degradation rates may not be accurately accounted for and the condition may not represent reality if the component-section were assessed today.

Finally, the central premise that corrosion is the primary factor affecting HVAC equipment could be incorrect. Although there is research that supports the notion that corrosion is a major factor in HVAC degradation, Bhatia (2020) notes that material defects and deposition of contaminants are also significant factors. Material defects may not be readily apparent and may be more likely to influence the probability of failure as opposed to long-term and continuous degradation. The probabilistic method proposed by Alley et al. (2017) for incorporating probability of failure into BUILDER may be a necessary step to help account for defects that cause spontaneous failure. Alternatively, existing maintenance ticket data could be used to identify incidences of failure by corrosivity category and see if there is any correlation or if it is decoupled from corrosive effects.

CHAPTER VI

CONCLUSION

Summary

In conclusion, the RQ was only partially answered and yielded an unexpected correlation of average condition by corrosion category for the D305006 Package Units.

RQ. How does HVAC equipment condition correlate by level of environmental corrosivity, as evidenced by BUILDER data on air conditioning equipment?

- A statistically significant difference in HVAC equipment condition by environmental corrosivity was identified.
- The pairwise correlation of categories is inconsistent with the lowest category of environmental corrosivity also exhibiting the worst condition.

The original intent of this paper was to gain a better understanding of the interplay between corrosion and condition data but the results show that answering that question is not straightforward and there are other factors in play for the condition of HVAC systems. Unexpectedly, the age of D305006 Package Units different to a significant degree between corrosion categories. This indicates that those categories have units that are being maintained far beyond their original service life, and this may be why those categories also have the lowest average CI. Additionally, three quarters of the data had to be excluded from analysis. It is apparent from the data that many values are copied and pasted into different fields, resulting in visible, widespread errors such as those described in Chapter III. The data represents a significant investment in labor and resources to collect it, but it is difficult to analyze if not correctly entered into BUILDER.

Contribution

The contribution of this research is an enhanced understanding of how component-section condition data correlates with real physical phenomena and challenges the assumption that corrosion is the primary driver for condition degradation in HVAC equipment. This is also a step in the direction of developing a corrosion cost factor similar to area cost factors for describing the budgetary and condition-based impact of corrosion in large and widespread real property portfolios.

Recommendations

This research raises questions on which factors primarily drive condition degradation of HVAC equipment in FCA data and the assumptions and limitations of this research should be revisited in future works to fully account for their impact. Additional suggestions for researchers include:

- Although there is limited historical BUILDER data due to the recency of its implementation in the USAF, the dataset continues to grow, develop, and receive updates by ongoing condition assessments. A more long-term study would be able to capture the time component of corrosion and degradation. Per AFIMSC, the USAF BUILDER data will not be fully useable for statistical analysis for another 10 years to capture trend data. (Vandeveer, M., personal communication, October 9, 2020).
- Additional statistical analysis of the data can be used. A key limitation in this research is that real world effects and data were analyzed using a normal distribution and transformation was required to validate the analysis. Other distributions may provide better results for the data.

- A more granular corrosivity categorization system could be used. ESI has 20 categories compared to the six using ISO 9223. The ESI system may be a better way to identify discrepancies and possibly used to focus in on locations that strongly deviate from expectations so they can be further analyzed.
- Conduct analysis on the basis of historical MAJCOM data or mission type. Significant historical differences in funding by MAJCOM could contribute to having an inventory that has a statistically significant, lower CI than average. Mission types should be considered as well, because support facilities or non-critical facilities may have historically had less funding than mission critical facilities at a location. For example, installations with active flying missions may spend more maintaining airfield pavements and less on facilities compared to installations that operate space-based equipment and therefore have operational personnel working in space operations facilities.
- Analyze the data weighted based on CRV. This research was conducted on a per item basis, but an analysis based on CRV may yield further insight. It may be that higher corrosivity locations have more expensive equipment than lower corrosion zones and therefore have lower CI when weighted for replacement cost.

Suggestions for practical application by technicians include:

- Improve the data quality in the BUILDER database. As noted in Chapter IV, approximately 75% of the available data was not included for HVAC equipment due to issues identified in data cleaning. Karanja & Mayo (2016) highlight the labor-intensiveness of conducting condition assessments and recommend increased standardization and of tasks and reduction of time spent by inspectors moving from one location to the next. Automation may be a tool for allowing inspectors to spend more time evaluating distresses and less time in transit.

- Conduct rule-based automated maintenance of the database. The errors identified in Table 9 can be used to prevent entry of erroneous data and automatically flag data issues for inspection and correction.

REFERENCES

Abergel, T., Dean, B., & Dulac, J. (2017). *Towards a zero-emission, efficient, and resilient buildings and construction sector. Global Status Report 2017* (Rep.). UN Environment and International Energy Agency.

Ahmed, V., Tezel, A., Aziz, Z., & Sibley, M. (2017). The future of Big Data in facilities management: Opportunities and challenges. *Facilities, 35*(13/14), 725-745. doi:10.1108/f-06-2016-0064

Air Force Civil Engineer Center. (2017). *USAF Built Infrastructure Inventory and Assessments Manual Appendix for HVAC (D30)*. United States Air Force, Air Force Civil Engineer Center. Retrieved from https://buildersummit.com/resources/

AFCEC. (2018). *Sustainment Management System Playbook*. United States Air Force, Air Force Civil Engineer Center. Retrieved from https://buildersummit.com/resources/

Alexander, L., Zhang, X., Peterson, T., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Bruent, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., Vazquez-Aguirre, J. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research, 111*(D5). doi:10.1029/2005JD006290

Allen, J. (Ed.). (2019). *Air Force Instruction 32-1020 Planning and Programming Built Infrastructure Projects* (USA, Department of the Air Force, Headquarters Air Force, Logistics, Engineering, and Force Protection). Washington, DC.

Alley, S. L., Valencia, V. V., Thal, A. E., & White, E. D. (2017). Probabilistic Assessment of Failure for United States Air Force Building Systems. *Journal of Performance of Constructed Facilities, 31*(5), 04017088. doi:10.1061/(asce)cf.1943-5509.000107

Bartels, L. B. (2020). *Work Optimization with Association Rule Mining of Negative Effective Deterioration in Building Components* (Doctoral dissertation, University of Illinois, 2020). Urbana, IL.

Bastidas-Arteaga, E., Chateauneuf, A., Sánchez-Silva, M., Bressolette, P., & Schoefs, F. (2010). Influence of weather and global warming in chloride ingress into concrete: A stochastic approach. *Structural Safety*, 32(4), 238–249. doi: 10.1016/j.strusafe.2010.03.002

Bastidas-Arteaga, E., Schoefs, F., Stewart, M. G., & Wang, X. (2013). Influence of global warming on durability of corroding RC structures: A probabilistic approach. *Engineering Structures*, 51, 259–266. doi: 10.1016/j.engstruct.2013.01.006

Bauer, L. (2020). *BUILDER Condition Assessment Guide Version 3.5.4*. United States Army Corps of Engineers Engineering Research and Development Center - Construction Engineering Research Laboratory. Retrieved from https://support.sms.erdc.dren.mil/downloads/files/builder-condition-assessment-guide-3-5-4

Bastidas-Arteaga, E., & Stewart, M. G. (eds.). (2019). Climate adaptation engineering: risks and economics for infrastructure decision-making. Oxford, England: Butterworth-Heinemann

Bhatia, A. (2020). *HVAC Design Considerations for Corrosive Environments*. Stony Point, NY: Continuing Education and Development. Retrieved August 27, 2020, from https://www.cedengineering.com/userfiles/HVAC Design for Corrosive Env.pdf

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017a). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552-557. doi:10.7334/psicothema2016.383

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017b). Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit? *Behavior Research Methods, 50*(3), 937-962. doi:10.3758/s13428-017-0918-2

Bröchner, J., Haugen, T., & Lindkvist, C. (2019). Shaping tomorrow's facilities management. *Facilities, 37*(7/8), 366-380. doi:10.1108/f-10-2018-0126

Charette, R. P., & Marshall, H. E. (1999). *UNIFORMAT II Elemental Classification for Building Specifications, Cost Estimating, and Cost Analysis* (USA, Department of Commerce, National Institute of Standards and Technology).

Chung, S., Sung, S., Hsien, C., & Shih, H. (2000). Application of EIS to the initial stages of atmospheric zinc corrosion. *Journal of Applied Electrochemistry, 30*, 607-615.

Dept. of Construction Science, Texas A&M University, & ALPHA Facilities Solutions, LLC. (2019). *Life Cycle Management of Air Force Facilities and Assets* [PPTX].

Dillon, J. J., & Herro, H. M. (1997). HVAC Failures. In *Corrosion 97* (Ser. 449). Houston, TX: NACE International.

Downton, M. W., Miller, J. Z. B., & Pielke, R. A. (2005). Reanalysis of U.S. National Weather Service Flood Loss Database. *Natural Hazards Review*, 6(1), 13-22. doi: 10.1061/(asce)1527-6988(2005)6:1(13)

Desrosiers, S. (2020, August 27). BUILDER Data Schema [E-mail to the author].

Francis, R. A., Falconi, S. M., Nateghi, R., & Guikema, S. D. (2010). Probabilistic life cycle analysis model for evaluating electric power infrastructure risk mitigation investments. Climate Change, 106(1), 31-55. doi: 10.1007/s10584-010-0001-9

Fuente, D. D., Díaz, I., Simancas, J., Chico, B., & Morcillo, M. (2011). Long-term atmospheric corrosion of mild steel. *Corrosion Science*, 53(2), 604-617. doi:10.1016/j.corsci.2010.10.007

Fuente, D. D., Castaño, J., & Morcillo, M. (2007). Long-term atmospheric corrosion of zinc. *Corrosion Science, 49*(3), 1420-1436. doi:10.1016/j.corsci.2006.08.003

Gaebel, W. (n.d.). ISO Corrosivity Category Estimation Tool (ICCET). Retrieved September 22, 2020, from https://www.wbdg.org/tools/corrdefense/iso.html

Geusic, S. (2018, August 24). Appendix D - Facilities Environmental Severity Classification Study. Retrieved August 15, 2020, from https://www.wbdg.org/files/pdfs/appendix_d_sup_iccet_mstr_instl_list.pdf

Ginger. J., Henderson, D., Humphreys, M., Konthesinghe, C., & Stewart, M. (2015). Wind loads on the frames of industrial buildings. Australian Journal of Structural Engineering, 16(2). doi: 10.1061/ajrua6.0001062

Grisham, P. R. (2001). *The Effects of Galvanic Corrosion on Air Conditioner Performance* (Master's thesis, Texas A&M University, 2001). College Station.

Grussing, M. N., & Marrano, L. R. (2007). Building Component Lifecycle Repair/Replacement Model for Institutional Facility Management. *Computing in Civil Engineering (2007)*. doi:10.1061/40937(261)65

Grussing, M. N. (2012). *Facility Degradation and Prediction Models for Sustainment, Restoration, and Modernization (SRM) Planning* (USA, US Army Corps of Engineers, Construction Engineering Research Laboratory). Champaign, IL: US Army Engineer Research and Development Center.

Hazra, A., & Gogtay, N. (2016). Biostatistics series module 3: Comparing groups: Numerical variables. *Indian Journal of Dermatology, 61*(3), 251-260. doi:10.4103/0019-5154.182416

Henderson, J. (Ed.). (2019). *Air Force Instruction 32-1001 Civil Engineer Operations* (USA, Department of the Air Force, Energy, Installations & Environment). Washington, DC.

Herrera, G. J., Stokes, C. A., Peña, V., & Howieson, S. V. (2017). *A Review of the BUILDER Application for Assessing Federal Laboratory Facilities* (pp. 1-60, Rep. No. D-8407). Washington, DC: IDA Science & Technology Policy Institute.

Howell, R. H., Coad, W. J., & Sauer, H. J. (2013). *Principles of Heating Ventilating and Air Conditioning* (7th ed.). Atlanta, GA: ASHRAE.

Ilyas, I. F., & Chu, X. (2019). *Data Cleaning*. New York: ACM.

Iowa Environmental Mesonet. (2020). ASOS-AWOS-METAR Data Download (D. Herzmann, Ed.). Retrieved September 02, 2020, from https://mesonet.agron.iastate.edu/request/download.phtml?network=KR__ASOS

IPCC. (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (Rep.) (C. B. Field, V. Barros, T. F. Stocker, Q. Dahe, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G. Plattner, S. K. Allen, M. Tignor, P. M. Midgley, Eds.). New York, NY: Cambridge University Press.

IPCC. (2013). *Climate Change 2013: They Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (T. F. Stocker, D. Qin, G. Plattner, M. M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley, Eds.). New York, NY: Cambridge University Press.

IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Core Writing Team, Pachauri, R. K., Meyer, L. A. (eds.)). IPCC, Geneva, Switzerland.

Japkowicz N., Stefanowski J. (2016) A Machine Learning Perspective on Big Data Analysis. In: Japkowicz N., Stefanowski J. (eds) Big Data Analysis: New Algorithms for a New Society. Studies in Big Data, vol 16. Springer, Cham. https://doi-org.srv-proxy2.library.tamu.edu/10.1007/978-3-319-26989-4_1

Karanja, P., & Mayo, G. K. (2016). State of Practice for Facility Condition Assessment. In *International Facility Management Association*. San Diego, CA.

Kendall, F. (Ed.). (2013). *Facilities and Infrastructure Corrosion Evaluation Study* (U.S.A., Department of Defense, Corrosion Policy and Oversight). Washington, DC.

Khajwal, A. B., & Noshadravan, A. (2020). Probabilistic Hurricane Wind-Induced Loss Model for Risk Assessment on a Regional Scale. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 6(2), 04020020. doi: 10.1061/ajrua6.0001062

Lai, J. H., & Man, C. S. (2018). Performance indicators for facilities operation and maintenance (Part 1). *Facilities, 36*(9/10), 476-494. doi:10.1108/f-08-2017-0075

Lavy, S., Garcia, J. A., & Dixit, M. K. (2010). Establishment of KPIs for facility performance measurement: review of literature. *Facilities*, *28*(9/10), 440–464. doi: 10.1108/02632771011057189

Lavy, S., Garcia, J. A., & Dixit, M. K. (2014). KPIs for facility's performance assessment, Part I: identification and categorization of core indicators. *Facilities*, *32*(5/6), 256–274. doi: 10.1108/f-09-2012-0066

Lavy, S., Garcia, J. A., & Dixit, M. K. (2014a). KPIs for facility's performance assessment, Part II: identification of variables and deriving expressions for core indicators. *Facilities*, *32*(5/6), 275–294. doi: 10.1108/f-09-2012-0067

Lee, J. Y. & Ellingwood, B. R. (2017). A decision model for intergenerational life-cycle risk assessment of civil infrastructure exposed to hurricanes under climate change. Reliability Engineering & System Safety, 159, 100-107. doi: 10.1016/j.ress.2016.10.022

Morales, J., Martín-Krijer, S., Díaz, F., Hernández-Borges, J., & González, S. (2005). Atmospheric corrosion in subtropical areas: Influences of time of wetness and deficiency of the ISO 9223 norm. *Corrosion Science, 47*(8), 2005-2019. doi:10.1016/j.corsci.2004.09.005

Munir, M., Kiviniemi, A., Finnegan, S., & Jones, S. W. (2019). BIM business value for asset owners through effective asset information management. *Facilities, 38*(3/4), 181-200. doi:10.1108/f-03-2019-0036

National Centers for Environmental Information. (2020, April 08). Calculating the Cost of Weather and Climate Disasters. Retrieved August 27, 2020, from https://www.ncei.noaa.gov/news/calculating-cost-weather-and-climate-disasters

Nguyen, M. N., Wang, X., & Leicester, R. H. (2013). An assessment of climate change effects on atmospheric corrosion rates of steel structures. *Corrosion Engineering, Science and Technology*, *48*(5), 359–369. doi: 10.1179/1743278213y.0000000087

Osborne, J. (2002). Notes on the use of data transformations. *Practical Assessment, Research, and Evaluation, 8*(6), 1-7. doi:https://doi.org/10.7275/4vng-5608

Osborne, J. W. (2013). *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Los Angeles: Sage Publications.

Rakanta, E., Daflou, E., & Batis, G. (2007). Evaluation of corrosion problems in a closed air-conditioning system: A case study. *Desalination, 213*(1-3), 9-17. doi:10.1016/j.desal.2006.05.054

Raughton, B. (2018, August 14). Dirt Boys restore taxiway in 23 hrs despite record heat. *51st Fighter Wing Public Affairs.* Retrieved from https://www.osan.af.mil/News/Article-Display/Article/1602575/dirt-boys-restore-taxiway-in-23-hrs-despite-record-heat/

Roper, K. O., & Payant, R. P. (2014). *The facility management handbook* (Fourth ed.). New York, NY: American Management Association.

Santana, J. J., Ramos, A., Rodriguez-Gonzalez, A., Vasconcelos, H. C., Mena, V., Fernández-Pérez, B. M., & Souto, R. M. (2019). Shortcomings of International Standard ISO 9223 for the Classification, Determination, and Estimation of Atmosphere Corrosivities in Subtropical Archipelagic Conditions—The Case of the Canary Islands (Spain). *Metals, 9*(10), 1105. doi:10.3390/met9101105

Silver, N. A., & Gaebel, W. (2017). *Facilities Environmental Severity Classification Study: Final Report* (USA, Department of Defense). Leidos.

Stewart, M. G., & Deng, X. (2015). Climate Impact Risks and Climate Adaptation Engineering for Built Infrastructure. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 1*(1), 04014001. doi:10.1061/ajrua6.0000809

Toliver, S. N. (2019). Supporting Mission Readiness by Mitigating Infrastructure Risk. Maxwell Air Force Base, AL: Air University.

Uzarski, D. R., Grussing, M. N., Mehnert, B. B. (2018). *Knowledge-Based Condition Assessment Reference Manual for Building Component-Sections* (USA, US Army Corps of Engineers, Construction Engineering Research Laboratory). Champaign, IL: US Army Engineer Research and Development Center.

Wang, X., Stewart, M. G., & Nguyen, M. (2011). Impact of climate change on corrosion and damage to concrete infrastructure in Australia. *Climatic Change*, 110(3-4), 941–957. doi: 10.1007/s10584-011-0124-7

Wright, B. A. (Ed.) (2020, June). Air Force and Space Force Almanac 2020. *Air Force Magazine*. https://www.airforcemag.com/app/uploads/2020/06/June2020_Fullissue5.pdf

APPENDIX

Table 14. Component-Section Details, Adapted (Desrosiers, 2020)

ID	Column	Data Type	Description
0	Cost_Escalation_N	Float	Area cost escalation factor
1	CI	Int	Component-section condition index
2	SI	Int	System index
3	MatEquipType	String	Component-section type
4	CRV	Int	Component (component-section) replacement value
5	BldgArea	Int	Building quantity
6	CATCODE	Int	DoD facility category code
7	System	String	Level 2 of UNIFORMAT
8	BldgNum	Int	Building number
9	MDI	Int	Mission dependency index
10	PRV	Float	Building plant replacement value
11	MaximumRSL_Standard	Int	Used for determining replacement (remaining service life)
12	SectionName	String	Assessor chosen name for section
13	RPUID	Int	Real Property Unique ID
14	Component	String	Level 3 of UNIFORMAT
15	CompType	String	Component-section material
16	MinimumCI_Standard	Int	Used for determining repair
17	Age	Int	Age in years since installed or constructed
18	SEC_Beta	Float	β for calculating CI

Table 14. Continued

ID	Column	Data Type	Description
19	FAC	Int	Building type for PRV determination
20	RSL	Float	Remaining service life
21	InstallDate	Int	Install or construct date
22	RPATotalUnitOfMeasureCode	String	Quantity type of measurement for component-section
23	SumOfRPATotalUnitOf MeasureQuantity	Int	Quantity for measurement of component-section
24	SEC_Alpha	Float	α for determining CI
25	Expected_Service_Life	Int	Overall expected service life (*E* in CI calculation)
26	BldgSize_Eng_UM	String	Unit of measure for building
27	Anonymous	String	Anonymized location
28	ClimateZoneCode	String	Climate variable
29	MainClimate	String	Climate variable
30	Precipitation	String	Climate variable
31	Temperature	String	Climate variable

Table 15. Equipment Details

ID	Column	Data Type	Description
0	BldgNum	Int	Building number
1	System	String	Level 2 of UNIFORMAT
2	Component	String	Level 3 of UNIFORMAT
3	MatEquipType	String	Component-section type
4	CompType	String	Component-section material

Table 15. Continued

ID	Column	Data Type	Description
5	SectionName	String	Assessor chosen name for section
6	Install Date	Int	Install or construct date
7	YearSource	String	Qualification of where install date came from
8	Qty	Int	Equipment quantity
9	UoM	String	Quantity measurement type
10	CRV	Int	Component (component-section) replacement value
11	CI	Int	Component-section condition index
12	AssetID	String	Self-explanatory
13	ID_Number	String	Assessor comment of local name
14	Equipment_Type	String	Assessor comment of equipment type
15	Equipment_Make	String	Manufacturing company
16	Model	String	Manufacturer model number
17	Serial_Number	String	Assessor comment on equipment plate serial number
18	Capacity	String	Assessor comment of equipment capacity
19	Manufacturer	String	Manufacturing company
20	Date_Manufactured	String	Self-explanatory
21	Date_Installed	Int	Self-explanatory
22	Control_Type_Make	String	Assessor comment
23	Warranty_Date	Int	Warranty end date
24	Warranty_Company	String	Self-explanatory
25	Location	String	Location in building

Table 15. Continued

ID	Column	Data Type	Description
26	Comment	String	Assessor inspection comment
27	Warranty_Date2	Int	Warranty end ate
28	Warranty_Company2	String	Self-explanatory
29	RPUID	Int	Real Property Unique ID
30	Anonymous	String	Anonymized location
31	ClimateZoneCode	String	Climate variable
32	MainClimate	String	Climate variable
33	Precipitation	String	Climate variable
34	Temperature	String	Climate variable