

Introduction to Forecasting Methods for Engineering Students Using R

David Greenburg and James Righter

The CITADEL School of Engineering

Abstract

At The Citadel, data collection and analysis using statistical models is taught in a statistics course for junior and senior engineering students. The course is designed to teach students how to frame a problem in terms of a hypothesis that can be tested statistically, collect and analyze data using standard statistical models and report analysis results. In one of the course modules students learn to build various forecasting models. This article describes the effort to develop, test and implement three types of forecasting models using a real world data set of traffic flow data obtained from the [State] Department of Transportation Traffic Polling and Analysis System (TPAS). The module introduces the students to the R statistical software and the fundamental aspects of data collection, analysis, interpretation and data visualization. The resulting module is flexible enough to serve beginner and advanced students as part of a semester long statistics course or as a stand-alone introduction to forecasting module. The instruction aims to provide students with the skills to experiment with data using the R software and develop fundamental skills in data collection analysis and forecasting. In the module, various forecasting models were developed and then compared using statistical methods. Students benefit from learning to build forecasting modules using R, which is a free software environment for statistical computing, and graphics that compiles and runs on a wide variety of platforms.

Keywords

Statistics, Forecasting, R software, Data analysis

Introduction

Society is demanding engineering graduates, with strong technical knowledge who can think creatively, communicate, and effectively manage diverse teams to solve challenging problems. The expectation that engineers can make decisions regarding the application of scarce or competing resources makes it clear that we must educate engineering students to be efficient and effective resource managers. Technology and industry needs are influential factors in education and industry relies on distinctive skills to innovate and compete. Engineering students expect that academic institutions will introduce them to appropriate technologies for their successful transformation into industry. Thus, academic institutions are challenged to adopt appropriate strategies to meet the innovative educational demand by both students and industry [1].

Engineers are often involved in the design and management of products or processes which require the accurate forecasting of constrained resources to include labor, equipment and materials and other activities relating to the project. Demand forecasting is a central and integral process for effective technical project management [2]. The success of any technical project is likely to be judged by how well it achieves a defined outcome while meeting performance expectations. Most projects would not be undertaken if management had little confidence in a successful outcome and once started are subject to being cancelled if the forecasted outcome does not meet expectations.

Project planning is an important activity that involves the careful study of historic data to project future demand outcomes. In many situations, engineering managers have access to a good deal of historical data which can be used to develop forecasts for the future. This paper discusses the development of a module to teach forecasting using R software as part of a probability and statistics course for undergraduate engineering students. The module is designed to complement an existing course that covers forecasting techniques. Due to the ease of learning and using the R software it can be used to complement lectures on forecasting and can easily be incorporated for use in homework assignments.

Forecasting Relevance

Because engineers collect and analyze data to make decisions and solve problems knowledge of data analysis and statistical models in the engineering domain is of great value. In particular forecasting is important to engineers as it deals with the collection, presentation, analysis and use of historical data to make future estimates that are used in making decisions, and solving problems [3]. The wide availability of computers allow us to emphasize the interpretation of data. By automating calculations and graphics we can focus student attention on interpreting the fit between statistical models and reality [4]. Today many computer forecasting programs are available and a major goal of this course is to expose students to statistical computing software and integrate its use into the curriculum.

Our approach to teaching forecasting emphasizes sound data collection approaches by introducing the students to readily available primary sources of raw data. One such source is the South Carolina Traffic Polling and Analysis System (TPAS) which provides near real-time

traffic count information at over 160 Continuous Count Stations (CCS) statewide [5]. These stations collect hourly traffic volume data through permanent sensors installed in the roadway. Students are shown how to access the TPAS and download an excel data file containing raw data for a specified period. An example of the data available is shown in Figure 1.

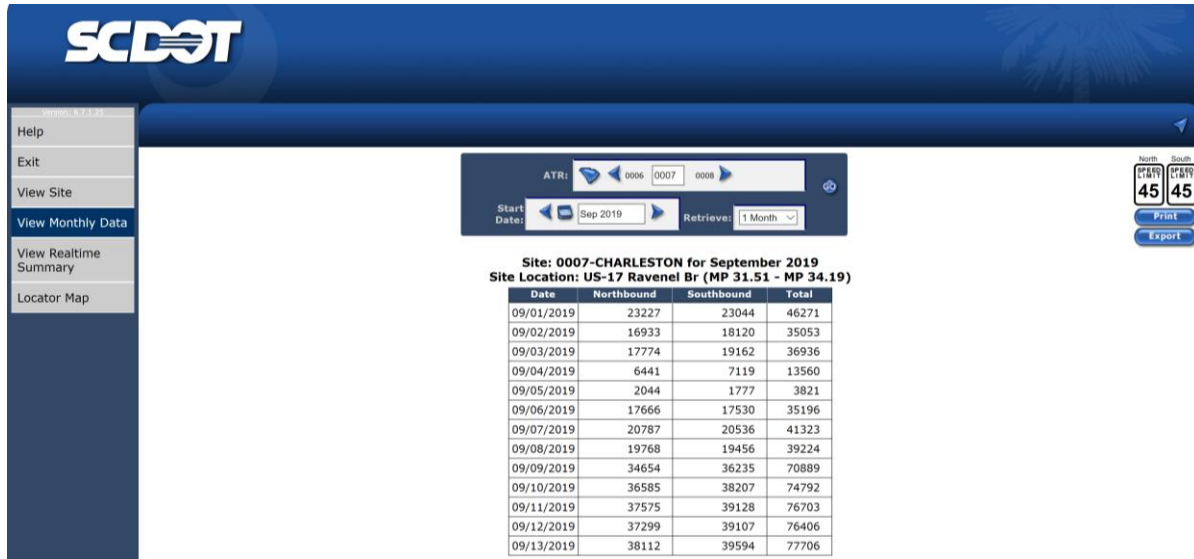


Figure 1. SCDOT TPAS Data for CCS 007

Students are asked to download data for a specific period of time to create time-series data that will be used to develop forecast models to predict future traffic flow. This paper describes several of the common forecasting approaches and how R software is used to perform them. Students are shown how to; create a time-series, plot, and decompose them into various error components. Once the students gain confidence in creating time-series data they are taught how to develop forecast models and evaluate their fit and accuracy.

Developing a Time Series.

TPAS data is easily exported as an excel file and imported into the R software and saved as a data frame which is the standard R data structure. The next step is to place the data frame into a time-series object containing the observations, starting and ending times of the series, and information on the periodicity (monthly data). Once the time-series object is created students can use a variety of R functions to model and plot the data. Students are taught that the first step when investigating a time-series is to plot it in order to identify any patterns, trends or significant error components in the data. The TPAS data is used to create a time-series shown in Figure (2).

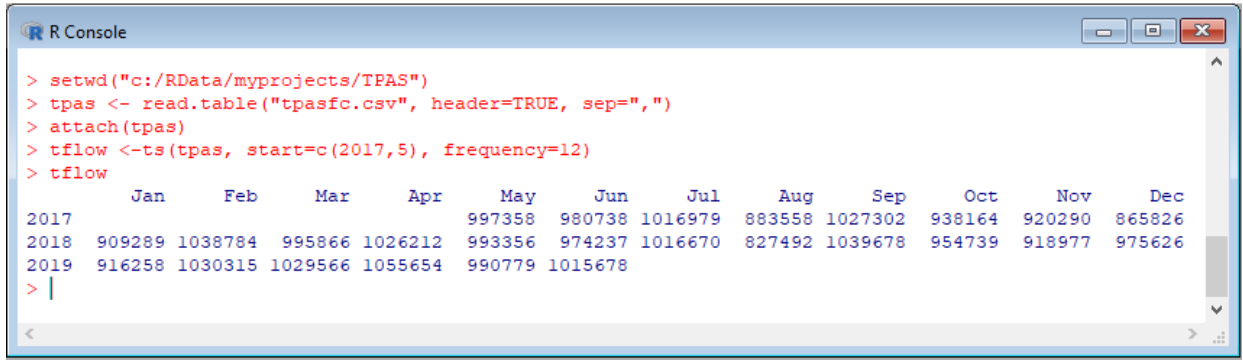


Figure 2. TPAS Time-series

Students are taught that before attempting to model the data it is always a good idea to explore the dataset by plotting it using the robust graphics capabilities in R. Plotting helps identify the irregular or error components. To assist in determining patterns in the data a simple centered moving average can be used which replaces the mean with that of the observation and one observation before and after it.

The equation for the centered moving average is:

$$S_t = (Y_{t-q} + Y_t + \dots + Y_{t+q}) / (2q + 1) \quad (1)$$

Where S_t is the smoothed value at time t and $k = 2q + 1$ [6]. The plot becomes increasingly smoothed as k is increased. Figure (3) shows a 3 period moving average imposed over the time series. The students are asked to experiment with the k value to discover any trend, seasonal or irregular patterns in the data.

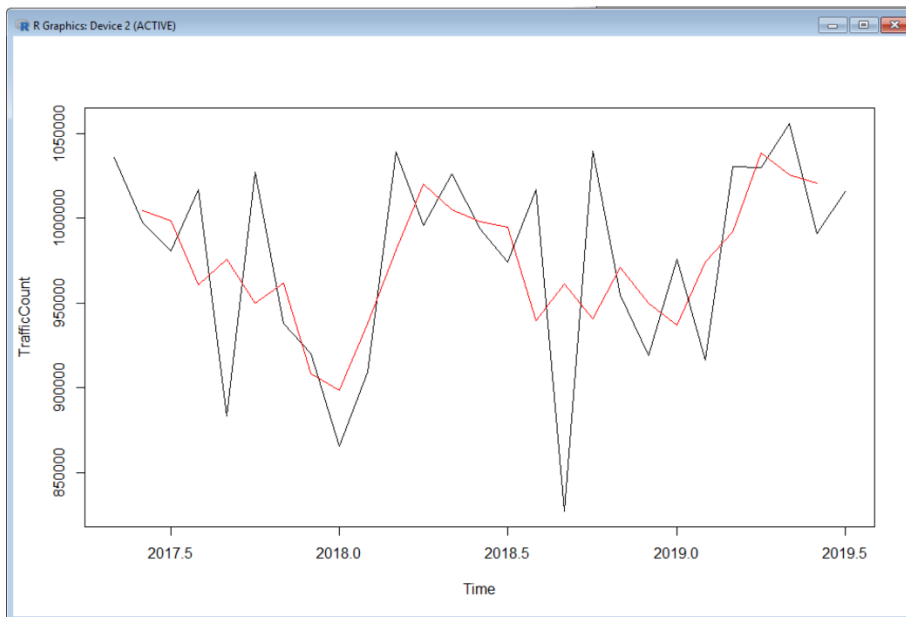


Figure 3. Time-series Plot for Traffic Flow with Superimposed Moving Average

Students are then shown how to decompose time series data into; trend components which captures changes in level over time, seasonal component which captures cyclical effects due to

time of year, and the irregular component whose influences are not described by trend or seasonal effects. The R `stl()` function can be used to decompose and plot the time series, seasonal, trend and irregular components. Standard predictive accuracy measures for the forecast can be displayed using the R `accuracy(fit)` function [6] as shown in Figure 4.

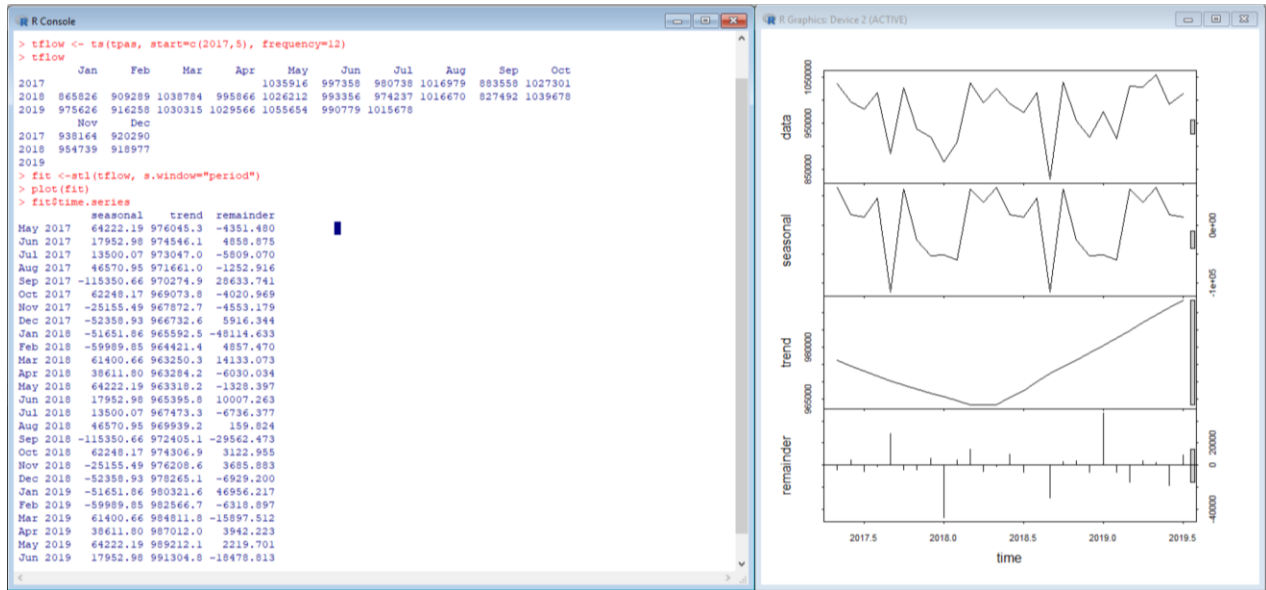


Figure 4. Time-series Decomposition and Plots for Traffic Flow

The students are introduced to various forecasting models to include moving averages, weighted moving averages, and exponential smoothing. The decomposition for the TPAS data shows there are seasonal, trend and irregular components so students are guided towards using the Holt-Winters exponential smoothing method which accounts for all three components of the data [2]. Due to the variety of models and unfamiliarity with how to choose the best model students are shown how to use the R `ets()` function which will automatically select a best-fitting model for the data [6].

$$Y_t = \text{level} + \text{slope} * t + s_t + \text{irregular} \quad (2)$$

Using the follow R commands;

```
tflow <- ts(tpas, start=c(2017,5), frequency=12)
tflow
hw(tflow)
plot(hw(tflow))
```

The holt-Winters forecast for 24 future periods with 80% and 95% confidence intervals is shown in Figure (6) and the forecast plot in Figure (7).

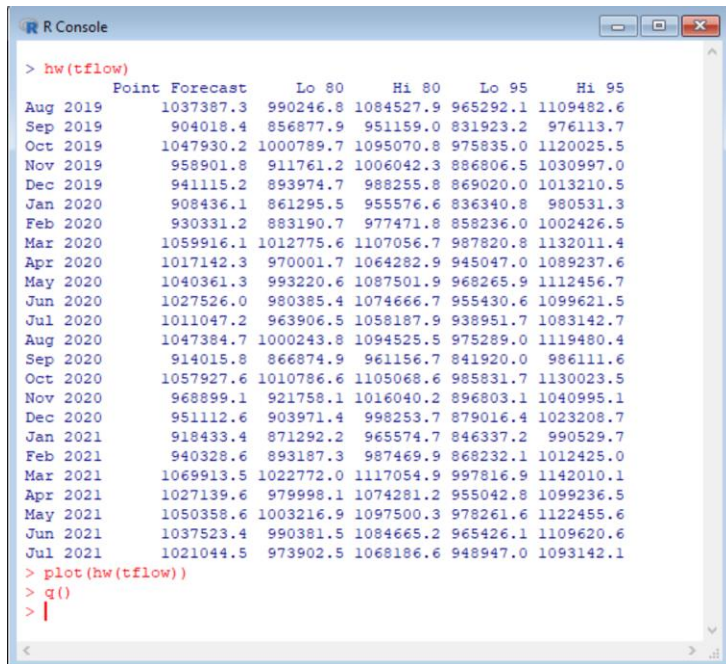


Figure 6. Holt-Winters Forecast With Trend, Seasonal, and Irregular Components.

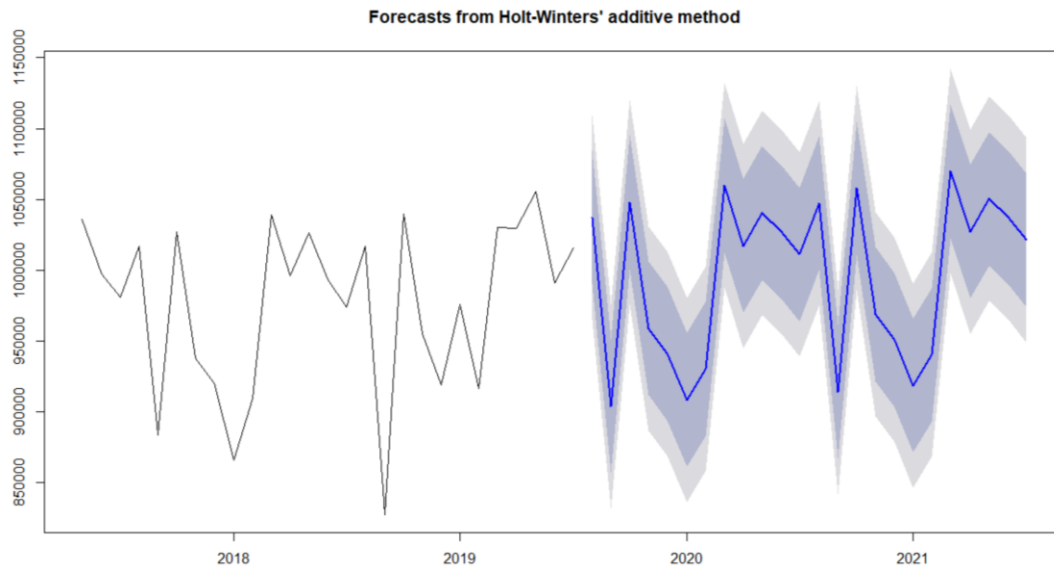


Figure 7. Holt-Winters Forecast Plot.

In Figure 7, the Holt-Winters forecasts are the solid blue line, and the 80% and 95% confidence intervals are provided in light and dark gray, respectively.

Another useful model is the autoregressive integrated moving average (ARIMA) model

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

which predicts each value of the time-series from the past p values and q residuals which can be used for comparison [6].

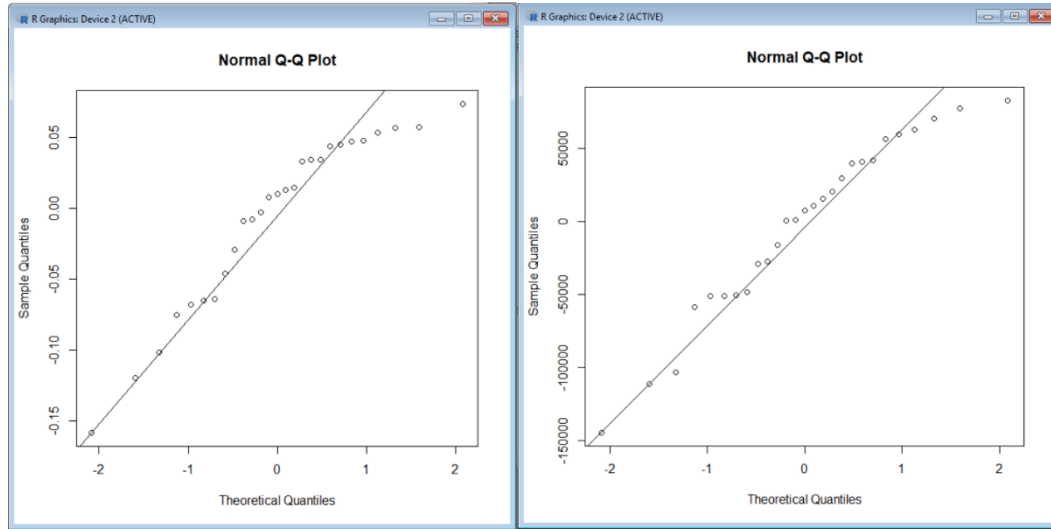


Figure 8. Normal Q-Q Plot for Holt-Winters and ARIMA models.

To test if the residuals are normally and independently distributed students can create and compare the Q-Q plots. Figure 8 shows a comparison using the Holt-Winters model on the left and the ARIMA model on the right. Students can visually assess and compare the inherent error in the model predictions. The R accuracy(fit) function can be used to show the predictive accuracy measures for time-series forecast such as ME, RMSE, MAE, MPE, MAPE, and MASE [6] as shown in Figure 9.

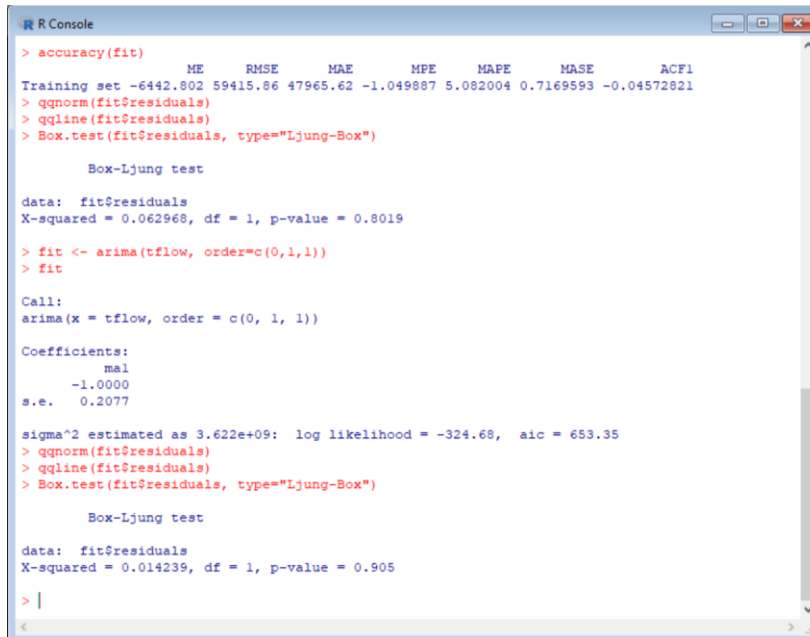


Figure 9. Example of R predictive accuracy measures.

2020 ASEE Southeast Section Conference

Students can then compare these measures for various types of models to assess which is the best one to use.

Conclusions

Demand plays a crucial role in the management of every organization. Understanding demand helps engineers and the organizations they work for to reduce risks and make important decisions. Having a good knowledge of forecasting techniques can play an important role in the management and operations of technical projects. This introduction to forecasting module provides students with the knowledge and practical application experience in building useful forecasting methods using readily available statistical software. By focusing on the concepts of valid data collection, organization, analysis, interpretation, and presentation vice the mechanics of calculation students gain abilities and confidence needed to solve demand based problems using forecasting methods.

REFERENCES

- [1] Anonymous. (2019), "Meeting Industry Needs for Professional and Technical Skills With New Graduate Degrees," Proceedings of the 2019 ASEE Annual Conference on Engineering Education, Tampa, FL, June 16-19.
- [2] Hyndman, R. J. and Athanasopoulos, A. (2012). *Forecasting: principles and practice*. OTexts. <http://otexts.com/fpp>. Accessed July 10, 2018.
- [3] *A Civil Engineering Curriculum for the Future: The Georgia Tech Case*
Michael D. Meyer and Laurence J. Jacobs
<https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%291052-3928%282004%29130%3A3%28160%29>
- [4] Cobb, G. W. (2007). The Introductory Statistics Course: A Ptolemaic Curriculum? *Technology Innovations in Statistics Education*, 1(1). Retrieved from <https://escholarship.org/uc/item/6hb3k0nz>
- [5] SCDOT <http://dbw.scdot.org/Poll5WebAppPublic/wfrm/wfrmHomePage.aspx>. Accessed July 17, 2018.
- [6] <http://127.0.0.1:20660/doc/manual/R-intro.html>

David Greenburg

David Greenburg is an Associate Professor and Program Director for the Department of Engineering Leadership and Program Management (ELPM) in the School of Engineering (SOE) at The Citadel. He served over 20 years of active military service in the United States Marine Corps in a variety of command and staff and leadership positions. Upon completion of active military service, he held executive leadership positions in industry until he joined the faculty at The Citadel. His research interests include modeling project networks, technical decision making and leadership. He is a certified Project Management Professional (PMP).

James Richter

James Richter is an Instructor of Mechanical Engineering in the School of Engineering (SOE) at The Citadel. He served 20 years in the United States Marine Corps in a variety of logistics and other leadership positions. He earned his BS in Mechanical Engineering at the U.S. Naval Academy, his MS in Military Studies from the Marine Corps University Command and Staff College, and his PhD in Mechanical Engineering from Clemson University. His research interests include engineering leadership, design methods, collaborative design, and engineering design education.